Understanding Mechanisms Underlying the Long-term Effects of Scholarships for Secondary School: Evidence from a Large Field Experiment in Colombia

Christian Manuel Posso Suárez

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2015

Approved by:

David Guilkey

Saraswata Chaudhuri

Klara Peter

Helen Tauchen

Tiago Pires

© 2015
Christian Manuel Posso Suárez
ALL RIGHTS RESERVED

ABSTRACT

CHRISTIAN MANUEL POSSO SUÁREZ: Understanding Mechanisms Underlying the Long-term Effects of Scholarships for Secondary School: Evidence from a Large Field Experiment in Colombia.
(Under the direction of David Guilkey)

I examine the long-term effects of educational subsidies programs on adult labor market outcomes through the accumulation of human capital. Combining experimental variation with an appropriate econometric model, I present sufficient conditions to identify the long-term effects on adult outcomes of educational subsidies programs under the presence of unobserved confounders that affect both the causal mechanism and the adult labor market outcomes. Using a large-scale program that randomized scholarships for private secondary school among disadvantaged students in Bogota Colombia in 1994, and Colombias administrative records covering schooling decisions and labor outcomes up to 15 years after the scholarship lottery, I find causal effect of the program on wages and labor force participation through the accumulation of human capital is significant and strong. I also find evidence of dynamic selection on educational attainment and selection on gains. The results provide important implications for the design of educational programs that target young disadvantaged children with the intent to develop human capital in the long run.

Without my wife, Ana Maria Lemos, and my daughter, Maria-Jose Posso this project would have never been started, and would have never been completed. I will always be in debt to you. Thank You.

ACKNOWLEDGMENTS

I am grateful to my advisor Professor David Guilkey for his continuous guidance, encouragement and support at every stage of this dissertation. I am also indebted to my dissertation committee members Saraswata Chaudhuri, Klara Sabirianova Peter, Helen Tauchen, and Tiago Pires for their invaluable advice and support. I would also like to thank Carlos Medina, Arlen Guarin, Donna Gilleskie, Brian McManus and participants of the UNC Applied Microeconomics Workshop for their continue support and helpful comments. It should be noted that this research would not be possible without the continue support of the Central Bank of Colombia, Banco de la República de Colombia. The results and conclusions in this dissertation are my own and do not indicate concurrence by the Banco de la República de Colombia and its Board of Directors.

TABLE OF CONTENTS

LIST OF TABLES . . . . . **viii**

LIST OF FIGURES . . . . . **ix**

1 INTRODUCTION **1**

2 LITERATURE REVIEW **9**

2.1 Literature review on long term effects and causal mechanisms 10

2.2 Literature review on the PACES program 15

3 A SIMPLE ECONOMIC MODEL OF EDUCATIONAL ATTAINMENT WITH AN SCHOLARSHIP **19**

3.1 The scholarship causal mechanism effect 21

4 FRAMEWORK: A SEQUENTIAL MODEL OF EDUCATIONAL ATTAINTMENT 23

4.1 Factor structure 25

4.2 Identification 26

4.3 Defining Treatment Effects 32

5 ESTIMATION STRATEGY **35**

6 DATA **38**

6.1 PACES background 38

6.2 Main sample 39

7 RESULTS **46**

7.1 Estimated Parameters 46

7.2 Average Treatment Effects of Education on Wages 52

7.3 Policy relevant Treatment Effects 56

7.4 Treatment and Policy Effects on Additional Outcomes 57

A COLOMBIA'S EDUCATIONAL SYSTEM AND PACES BACKGROUND 60

B KOTLARSKI'S THEOREM 61

C SUFFICIENT CONDITIONS FOR IDENTIFICATION OF PARAMETERS AND DISTRIBUTIONS 62

D ALTERNATIVE SET OF ASSUMPTIONS 66

E ADDITIONAL TABLES AND FIGURES 67

BIBLIOGRAPHY 85

LIST OF TABLES

6.1 Observable variables in sequential schooling model and outcomes 43

6.2 Sample Statistics: Sequential schooling transitions 44

6.3 Sample Statistics: Outcomes 45

7.1 Estimates for the sequential schooling model 48

7.2 Estimate coefficients for potential outcomes: wages and formal wages 51

7.3 Treatment Effects for final educational levels: High School dropout as a comparison group 53

7.4 Policy Relevant Treatment Effects 57

E.1 Variable definitions 68

E.2 Estimates for the sequential schooling model with an additional instrument 69

E.3 Estimates for the sequential schooling model with different distributional assumption on θ 70

E.4 Estimates for the sequential schooling model with alternative specification 71

E.5 Goodness of Fit for outcomes 71

E.6 Goodness of Fit for outcomes 72

E.7 Treatment Effects for final educational levels using specification with additional exclusion restrictions: High School dropout as a comparison group 73

E.8 Treatment Effects for final educational levels without unobserved heterogeneity: High School dropout as a comparison group 74

E.9 Estimates coefficients for potential outcomes: Days worked formal market and Sisben score 75

E.10 Treatment Effects (other outcomes) for final educational levels: High School dropout as a comparison group 76

E.11 Policy Relevant Treatment Effects for other outcomes 77

LIST OF FIGURES

1.1 Causal Mechanism of a scholarship program 3

4.1 Sequential Schooling Model 24

7.1 Densities of θ by final level of education 50

7.2 Proportion of each decile of $\epsilon_{0,1}$ induced to switch by the PACES scholarship . . . 58

E.1 Densities of $\Pr(D_{0,1} = 1|scholarship = 1)$ and $\Pr(D_{0,1} = 1|scholarship = 0)$ . . 67

E.2 Observed and simulated percentiles of the log of the monthly wage rate in 2010 . . 77

E.3 Observed and simulated percentiles of the log of the monthly formal wage rate in 2010 78

E.4 Observed and simulated percentiles of Days worked in 2010 78

E.5 Colombias Educational System 79

CHAPTER 1

INTRODUCTION

Recent literature establishes that significant investment in child development at an early age is one of the key determinants of future adult outcomes. Huggett, Ventura, and Yaron (2011) shows that for the United States most of the variability in lifetime earnings is explained by personal attributes determined before age 23.[1] Existing studies in labor economics predominantly find that when these investments are focused on young disadvantaged children, the economic returns are even higher.[2] In particular, there is strong evidence of a positive impact of educational subsidies and targeted conditional cash transfers on educational outcomes.[3] Despite the fact that most of these policies have the goal of developing human capital in the long run, there has been little research on long term causal effects on adult outcomes.

This paper estimates the long-term effects of educational subsidies programs on adult labor market outcomes through the accumulation of human capital. Most of the empirical studies that evaluate the impact of educational subsidies focus mainly on establishing whether the policy variable impacts a particular outcome, but usually they fail to explain the causal mechanisms through which the policy affects the outcome.[4] This is particularly important in cases where someone is interested in evaluating the long-term effects on adult outcomes since most of the possible impacts of the program may take place through indirect channels. For instance, a scholarship program for

---

[1] Keane and Wolpin (1997) find evidence in the same direction. Almond and Currie (2011), after an intensive review the literature in this topic, concluded the same.

[2] See for instance Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham (2013), Heckman, Pinto, and Savelyev (2013), Fletcher (2013), and Campbel, Cont, Heckman, Moon, Pinto, Pungello, and Pan (2014).

[3] For example, Angrist, Bettinger, and Kremer (2006), Akee, Copeland, Keeler, Angold, and Costello (2010), Gertler et al. (2013), Lee and Seshadri (2014), Bettinger, Kremer, Kugler, Medina, Posso, and Saavedra (2014), Heckman and Mosso (2014)

[4] An interesting discussion about whether to test directly the causal mechanism versus a standard policy evaluations can be find in Ludwig, Kling, and Mullainathan (2011).

secondary education of teenage students in a developing country may affect positively the accumulation of human capital and then through this channel adult labor market outcomes, but it is unlikely that the scholarship itself affects adult labor market outcomes directly.

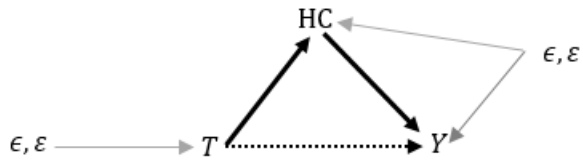
Figure 1.1 graphically illustrates the problem. The policy program T affects adult labor market outcome Y directly and indirectly. The direct effect of the program is the single dotted line going from policy T to outcome Y, while the indirect effect combines two solid lines going from T to the outcome Y through the accumulation of human capital, HC, the causal mechanism of interest. Finally, ϵ and ε represent unobserved confounders associated with the taste for education and abilities that affect labor market outcomes, respectively. The identification of the long-term effects of these type of policies is a challenge for several reasons. First, since the unobserved confounders (e.g. ϵ and ε) affect the policy variable, the human capital accumulation process, and the adult outcomes, identification of the causal mechanism is not possible under standard assumptions such as ignorability of the treatment variable. For instance, to secure exogeneity of the treatment variable, a common practice is to randomize the policy variable.[5] Nonetheless, this strategy alone is insufficient for identification of causal mechanisms (e.g. the human capital accumulation mechanism).[6] Combining experimental variation with an appropriate econometric model, I present sufficient conditions to identify the long-term effects on adult outcomes of educational subsidies programs under the presence of unobserved confounders that affect both the causal mechanism and the adult labor market outcomes.

Second, identification of long-term effects requires access to data where the analyst is able to observe the individuals who were originally affected by the program in several stages of the human capital accumulation process and the adult labor market outcomes. Most of the policy programs follow the agents for a short period of time or only follow a small number of cases, although there are some exceptions. For example, Gertler et al. (2013) studied the long-term effects of a

[5]With non-experimental data, analysts often statistically adjust for the observed differences in the pretreatment observed covariates between the treatment and control groups using regression techniques.

[6]Imai, Keele, , and Yamamoto (2010) and Imai, Keele, Tingley, and Yamamoto (2011) formally show why conventional exogeneity assumptions are insufficient for identifying causal mechanisms.

Figure 1.1: Causal Mechanism of a scholarship program



Notes: The figure illustrates the two paths through which the program T may affect the outcome Y. The causal mechanism of interest is represented by ⟶, where the causal effect of the program on the outcome is transmitted through the human capital accumulation, HC. The direct mechanism of the program is represented by ⇢. Possible confounders ϵ and ε affect both the human capital accumulation and outcome Y is represented by ⟶

program that gave psychosocial stimulation to 129 stunted Jamaican toddlers living in poverty in $1986 - 1987$. The program included children at age $9 - 24$ months. To evaluate the long-term effects of the program, 127 individuals were re-interview again at ages 7, 11, 18 and 22. The adult labor market outcomes were observed in the last interview when the individuals were 22 years old.[7]

Although, in principle, the approach presented here can be applied to any policy that affects human capital accumulation at early stages of life, I illustrate it by applying it to a massive scholarship program in Colombia. In late 1991 Colombia implemented the *Plan de Ampliacin de Cobertura de la Educacin Secundaria* (PACES), a scholarship program for disadvantaged students [8] that provided over $125,000$ vouchers in about $2,000$ private schools located in Colombias largest cities. The program had two main objectives. First, increasing attainment in secondary education, and second reducing the dropout rate of students associated with the low transition rate from public elementary school to secondary school.

Three characteristics made this program unique. First, in municipalities where the demand for the program exceeded the supply, the scholarships were assigned using a lottery. In particular, in

[7]Conti, Heckman, and Pinto (2014) and Heckman et al. (2013) present similar examples associated with the Perry program in United States.

[8]Based on residential location, Colombias largest cities divide households into six strata. The poorest two strata were the focal group of the PACES program.

1994 in Bogota, the selection of the beneficiaries of the program for the next year was awarded randomly, making this program one of the largest field experiments associated with educational subsidies. My baseline sample covers the PACES program applicants in Bogota in 1994. Second, the scholarship was renewed annually contingent upon passing the previous year. By 1995, in Bogota, the scholarship covered most tuition fees for those students who were enrolled in the first year of secondary education.[9] Finally, Colombia has comprehensive individual-level administrative data on secondary education, college enrollment, and labor market experiences that allow me to keep track of human capital accumulation of individuals, and labor market outcomes up to 15 years after the initial intervention.

Similar to most of the existing literature, the research on the PACES program in Bogota exploits the randomization of the program in two ways. First, comparing outcomes between scholarship lottery winners and losers, regardless of the treatment the students actually received. This strategy is called intent-to-treat (ITT) and it tries to avoid two common complications associated with randomized controlled trials: noncompliance and missing outcomes. The standard strategy uses linear regression of the outcome of interest Y_i as a function of some baseline controls, X_i, and an indicator variable T_i for whether the student won the scholarship: $Y_i = X_i'\beta + \gamma T_i + \varepsilon_i$.[10] The second strategy uses the indicator of whether the student won a voucher as an instrument for the actual treatment in order to get the Local Average Treatment Effect (LATE, Angrist and Imbens (1994)).[11] Under perfect compliance, the average treatment effect, ITT and LATE will be identical.

However, in many cases these studies suffer from selection bias (Heckman and Vytlacil (2005))

[9]Nonetheless, the government did not adjust the value of voucher and by 1998 the scholarship covered only about 56 percent of the tuition (Angrist, Bettinger, , Bloom, Kin, and Kremer (2002)).

[10]For the case of the PACES program, the short-run results are robust to whether the regression includes the baseline controls or not.

[11]See Heckman (1996) for a discussion on the identification under randomization. Heckman (1996) discusses these two cases and shows that in general randomized social experiments operate in practice as an instrumental variable. Ravallion (2015) shows the limitation of both approaches when the decision of take up treatment will depend on latent costs and benefits of take up. Notice that the last scenario is a most realistic assumption. In general, if the decision to take up treatment is correlated with the unobservable cost and benefits of choosing the treatment, then the complication of essential heterogeneity is present (see Heckman, Urzua, and Vytlacil (2006)). This also called the correlated random coefficients model when a linear functional form is assumed (see Heckman, Schmierer, and Urzua (2010)).

and essential heterogeneity (see Heckman et al. (2006), Heckman et al. (2010)) and thus they ignore key characteristics of the individual decision making process. For instance, Angrist et al. (2006) evaluate the effect of PACES on academic achievement using Colombia's centralized college entrance examinations. Since the examinations are only for those who graduate from high school, and the PACES program has negative effects on grade repetition (Angrist et al. 2002), and a positive effect on the probability of graduation from high school (Angrist et al. (2006), Bettinger, Kremer, and Saavedra (2010), Bettinger et al. (2014)), direct comparison of test scores between scholarship lottery winners and losers (the ITT strategy) is subject to selection bias. Additionally, if the students who graduate from high school are more likely to experience higher (unobserved) gains associated with the entrance examination, then there is selection on the gain to treatment or essential heterogeneity. Angrist et al. (2006) developed a methodology to deal with the selection bias [12], but they are silent about how to account for the essential heterogeneity problem. In the case of the adult labor markets outcomes 15 years after the intervention, these problems are also important. Since the program has a causal effect on human capital accumulation, and education is a key determinant of labor supply and wages [13], the direct comparison of labor market outcomes between lottery winners and losers is subject to both selection bias and essential heterogeneity.

The present study addresses these two problems by introducing a more flexible set of assumptions. I start by assuming that the only channel through which the PACES program impacts adult outcomes 15 years after the initial intervention is through human capital accumulation. In particular, I assume that the scholarship was a significant (exogenous) cost shifter that affects the decision to graduate from high school but does not affect the decision to enroll in college or graduate from college conditional on graduating from high school directly. Once an individual chooses to graduate or not from high school, the randomization is no longer available, and the selection problem

[12] Angrist et al. (2006) proposed a parametric and a nonparametric approach. The parametric approach is simple modification of the Tobit model where the censored point is modified exogenously. The second approach derivate a series of nonparametric bounds for quantile specific program impacts.

[13] See Oreopoulos and Petronijevic (2013), Harmon, Oosterbeek, and Walker (2003).

will affect future schooling decisions. In order to control for the selection problem, I build a sequential model of educational attainment with three sequential schooling choices[14]: high school graduation, college enrollment, and college graduation. This model allows students with different sets of unobservables to select themselves sequentially over the different schooling choices. This characteristic of the model is called dynamic selection or educational selectivity (Cameron and Heckman (1998), Cameron and Heckman (2001)). The selection is dynamic in the sense that every current schooling choice depends on previous schooling choices.

Once individuals finish their schooling decisions, adult outcomes are realized. My model includes four potential outcomes, each of them associated with one of the final levels of education: High School (HS) Dropout, HS Graduate, College Dropout, and College Graduate. If agents make their schooling choices based on unobservable gains of the final schooling level, then the schooling choice is correlated with return of the level of education even after controlling on observable characteristics. This is the problem of essential heterogeneity in terms of Heckman et al. (2006).[15] In this context essential heterogeneity means that individuals make rational choices about what should be their final level of education based on endowments they know but are unobservable to the analyst, and it is expected that those agents with higher unobserved endowments are more likely to select themselves into higher levels of education and experience higher gains from that choice in the labor market. I also account for the nonlinearity of the effect of schooling on outcomes, which has been shown to be important on outcomes like wages.[16]

The plan of this paper is as follows. Section 2 reviews the literature related to the long-term effect of social programs and causal mechanisms, and the literature associated with the PACES program. Section 3 describes a simple dynamic model of schooling attainment and work decisions

[14]This model also has been called a reduced-form dynamic model of schooling attainment or a Reduced form dynamic treatment effects model or sequential discrete choice model (Belzil (2008); Heckman and Navarro (2007); Cameron and Heckman (2001)).

[15]In a more simple case with only two levels of schooling, this is called a correlated random coefficient model (see Heckman et al. (2010)).

[16]Belzil and Hansen (2007), Belzil (2008) and Heckman, Lochner, and Todd (2006) have found that ignoring the convexity relationship between wages and schooling inflates the cross-sectional variance in the returns to schooling.

to illustrate the possible causal mechanism of scholarship program assigned randomly that lowers the net cost of schooling. This model shows that the scholarship creates strong incentives for the accumulation of human capital, and also demonstrate that the causal mechanism of program on adult outcomes is through schooling.

Section 4 describes my empirical framework and makes explicit the main set of assumptions. In this section I also establish identifiability of my sequential model of educational attainment. The identification exploits the randomization of the PACES program in the first schooling transition. For the second and third transition I require measurable separability between the index function associated with each transition which is achieved by requiring at least one additional exclusion restriction. My model can also be identified without additional exclusion restrictions if the variation in the set of observable variables that affect schooling transitions is sufficiently general. My results are robust for either case. Since I observe only one potential outcome for any person, without additional assumptions I am not able to identify the joint distribution of potential outcomes, which implies that I cannot identify treatment on the treated parameters. I use a factor structure to identify the joint distribution of potential outcomes and the policy relevant parameters of interest.

Section 5 discusses my estimation strategy and in section 6 I present the data that I use to estimate my model. Section 7 includes the main results. In addition to conventional treatment parameters such as the average treatment effect or the average treatment on the treated, I compute the policy-relevant treatment parameters (Heckman and Vytlacil (2001)) defined as the effect on outcomes for those who benefit from the program. I find that that the overall return on wages for those individuals who benefit from the PACES program is 13.8 percent, while when I constrain wages only to those who work in the formal market (formal wages) the return is 12.5 percent. The PACES program also has important effects on the intensive and extensive margin of labor supply. In particular, those who benefit from the PACES program work 45 days more in the formal market per year than those who did not get the scholarship. The effects of the program on poverty are also important. Students for whom PACES is relevant have on average 9.5 points higher on the SISBEN score, a Colombian poverty index, than those who did not participate of the program, which provides evidence of the positive impact of the program on the quality life of the individuals.

The effects of the program are higher when I condition to the final level of education, showing that accounting for the nonlinearity of the effect of schooling on outcomes is important in order to compute policy parameters. I also find that dynamic selection is important and there is substantial sorting at all levels of education. I conclude in the last section.

CHAPTER 2

LITERATURE REVIEW

The plan of this section is to review the literature that is directly related to this paper. The paper contributes to several strands of the literature. First, it contributes to the econometric and applied literature that focuses on the estimation of causal mechanisms and long-term effects of social programs. This paper estimates long-term effects of a high school scholarship program on labor market outcomes through the accumulation of human capital. The existing studies primarily focus on policy evaluation of the program but typically ignore the identification of the causal mechanisms that underlay the policy. In general, if the understanding of the policys mechanisms is weak, either because of the lack of a robust economic theory or the absence of empirical evidence, then a policy evaluation may be the best strategy. Nonetheless, in some cases, it is likely that the analysts have strong beliefs about the mechanisms through which the policy affects outcomes. Such beliefs may be the result of a deep understanding of the design of the policy, the accumulation of empirical evidence or economic theory. In such cases, the objective of the analyst may be to find the appropriate test for the mechanism of interest instead of a simple policy evaluation. Section 2.1 summarizes this literature.

Second, the paper contributes to the literature that evaluates the effects of the PACES program. In the case of the PACES program, the key mechanism through which the policy affects adult labor outcomes is clear. By design, the PACES program was created to increase attainment and reduce the dropout rates in secondary education. Also, economic theory predicts that a tuition scholarship program may increase the current and future value of schooling (see section 3 in this paper, Todd and Wolpin (2006); Attanasio, Meghir, and Santiago (2012); Dubois, Janvry, and Sadoulet (2012)). Finally, the empirical literature related to the PACES program has shown that effectively the PACES program has negative effects on grade repetition (Angrist et al. (2002)), and

a positive effect on the probability of graduation from high school (Angrist et al. (2006); Bettinger et al. (2010); Bettinger et al. (2014)). Overall, there is enough information to believe that the key mechanism of the PACES program on adult labor outcomes is through the accumulation of human capital. Section 2.2 summarizes the results associated with the PACES program.

2.1 Literature review on long term effects and causal mechanisms

Since the seminal papers of Haavelmo (1943) and Koopmans (1947), the cause effect relationships representing economic mechanisms have been central in empirical analysis in economics. Haavelmo (1943) was the first economist to recognize the importance of economic theory to define the mechanism of interest and to guide policy analysis. Recently, Ludwig et al. (2011) reinforce the ideas of Haavelmo (1943), and argue that causal mechanisms should play a more central role in policy analysis, especially in cases where experimental variation is available. Under the presence of randomization of the program, Ludwig et al. (2011) distinguish two types of research: policy evaluations and mechanism experiments. The first type of research compares directly treated and control groups, while the second type test the causal mechanism that underlies the policy. The key differential characteristic between these research areas is that the mechanism experiments incorporate prior knowledge associated with the design of the program, previous findings, and economic theory to identify the effect of interest. Which type of research applies may depend on prior knowledge of the policy in question. Similarly, Hennessy and Strebulaev (2015) argue that simple policy evaluations can be only be understood and interpreted when there is an economic model that clearly identifies the policy underlying the generating process through which the policy would boost the outcome of interest. A clear example that shows how these research areas differentiate is Heckman et al. (2013). They show that previous literature associated with Perry Preschool program focuses on policy evaluations of the program but do not attempt to explain the causal mechanisms that are driving such effects. The authors show that the key mechanism for this program is permanent improvement of psychological skills. To the best of my knowledge, there is no paper that tries to identify the causal mechanism of the PACES program. Most of the literature that empirically evaluates long-term effects of social program has focused on policy evaluation, but there are some important exceptions.

The literature on the long-term effects of policy interventions at early stages of life has expanded in recent years. Almond and Currie (2011) carried out an extensive literature review and found that before 2005 little or nothing about this topic was published in top journals in economics. Nonetheless, there have been at least five articles per year in these journals since 2005. Although most of the empirical literature focuses on short-term policy evaluations of social programs, there are some important recent exceptions. Some cases exploit experimental variation associated with randomization of access to the program in the baseline to implement a policy evaluation. An interesting example is Maluccio, Hoddinott, Behrman, Martorell, Quisumbing, and Stein (2009) who studies the effects of an early childhood nutritional program in Guatemala on adult educational outcomes. Using the ITT approach the authors found important effects of the program 25 years after it ended. In particular, the program increases by 1.2 the number of grades completed for women, and by one quarter of a standard deviation on standardized reading comprehension and non-verbal cognitive ability tests for both women and men.

Another example is Gertler et al. (2013). They studied the long-term effects of a program that gave psychosocial stimulation to 129 stunted Jamaican toddlers living in poverty in $1986 - 1987$. The program included children at ages $9 - 24$ months. To evaluate the long-term effects of the program, 127 individuals were re-interviewed at ages 7, 11, 18 and 22. The adult labor market outcomes were observed in the last interview when the individuals were 22 years old. Similar to Maluccio et al. (2009), Gertler et al. (2013) applied the ITT model, although the authors developed methods to correct for the small sample size for the inference analysis.[1] They found that psychosocial stimulation in early stages of life increased the average earnings of participants by 42 percent.

Two recent papers by Campbel et al. (2014) and Campbell, Pungello, Burchinal, Kainz, Pan, Wasik, Barbarin, Sparling, and Ramey (2012) estimate the long-term health and education effects

[1]The authors address the small sample size problem by using non-parametric permutation tests (see Heckman, Moon, Pinto, Savelyev, and Yavitz (2010) p. 16). This test relies on the idea that under randomization of the program, the joint distribution of outcome and treatment assignments is invariant for certain classes of permutations. The key properties of permutation tests is that they are distribution-free and provide accurate p-values even when the sampling distribution is skewed.

of the Carolina Abecedarian Project (ABC). ABC was designed to investigate whether a stimulating environment during childhood prevents the development of mild mental retardation. The sample includes 111 children that were born between 1972 and 1977 and were living in or near Chapel Hill, North Carolina. As the previous examples, the authors applied the ITT model, and similar to Gertler et al. (2013), they include a correction for the small sample size. Campbel et al. (2014), using new collected biomedical data, found that children treated by the ABC program have significantly lower prevalence of risk factors for cardiovascular and metabolic diseases in their $mid - 30s$. Additionally, Campbell et al. (2012) found that treated individuals attained significantly more years of education, although there are no important effects on income.

In the absence of experimental variation to identify long-term effects, much work has focused on the quasi-experiment variation associated with cross-sectional difference between regions or cities or other dimensions. An interesting example is a series of papers of Akbulut (2014), Akbulut (2015) and Akbulut, Khamis, and Yuksel (Akbulut et al.) that study the long-term effects of the Second World War (WWII) on human capital accumulation, earnings and health in Germany. In all cases, the idea is to exploit region- by- cohort variation in destruction in Germany arising during the WWII as a unique quasi- experiment. The standard strategy is a linear regression of the outcome of interest as a function of an interaction of the Destruction variable that is fixed over time and individuals but varies between regions, and a cohort variable that varies across individuals and time, and includes the specific cohort affected by the war. The regression also includes regions fixed effects, cohorts fixed effects, and other controls. This is basically the Difference- in- Differences (DID) approach where the treatment is associated with those individuals who were living in a region that suffered a significant destruction during WWII.[2] Using this approach, Akbulut (2014) found that exposure to destruction has long-term effects on human capital accumulation, health and earnings even after 40 years of the WWII. Similarly, Akbulut (2015) found that individuals who were exposed to destruction during prenatal and early postnatal period are more likely to be

[2]In particular, the regression is the following: $Y_{irt} = X'_{irt}\beta + \delta_r + \rho_t + \gamma(Destruction_r \times Cohort_{irt}) + \varepsilon_{irt}$, where X_{irt} is a vector of individual characteristics, δ_r are the region fixed effects, ρ_t are the year fixed effects. γ is the parameter of interest. As long as the variables Destruction and Cohort are exogenous, the identification of is possible.

obese and are more susceptible to having chronic health conditions as adults. Kesternich, Siflinger, Smith, and Winter (2014) provides similar evidence for Europe by using variation at country level.

Almond (2006) explores the long-term effects of prenatal adverse shocks. The author estimates the effects of the 1918 influenza pandemic in United Sates on the cohort that was born in that year with respect to the cohort that were born right before and after 1918. Similar to the previous examples, the author exploits the region- by- cohort variation associated with the intensity of exposure to the influence. Using census data, Almond (2006) found that the 1918 influenza pandemic reduced the number of years of schooling, the probability of graduating from high school, and reductions in total income, although the last effect is only significant for males. Finally, the pandemic also increased the rates of physical disability.

Another example is Smith (2009). This paper uses data that collects socio economic status measures of a panel of individuals with their siblings at early stage of life and who are now in adulthood, to measure the effects of poor health as a child on adult outcomes. Similar to previous example, Smith (2009) also use a DID type strategy but in this case the author exploits variation between and within siblings. Although the author did not find effects on education, poor health as a child has large effects on income and wealth.

An important gap in this literature, for both those who wish to exploit experimental variation or those who use quasi-experimental variation, is that little is known about the causal mechanisms producing such impacts. Some papers hypothesized possible causal mechanisms, but rarely do such papers provide an estimate of the economic magnitude and statistical significance of the mechanism of interest. There are some exceptions. For instance, Meghir, Palme, and Simeonova (2012) study the causal effect of education on health using a compulsory education reform in Sweden. They found that the reform has long-term effects on health, mainly because the reform reduced male mortality up to age fifty, the probability of being hospitalized and the average costs of inpatient care. The authors argue that such effects may be explained by three causal mechanism: differences in economic resources, time preferences and knowledge. The paper does not formally discuss the identification of such causal mechanisms. Instead, the author shows that the reform had effects on schooling attainment and earnings but only in individuals that originally belonged

to families with low socio economic status, and conclude that such results provides evidence in favor of the causal mechanism associated with differences in economic resources.

Another interesting example is Chetty, Friedman, Hilger, Saez, Whitmore, and Yagan (2011). This paper evaluates the long-term effects on adult outcomes of Project STAR, a program in Tennessee where both students and teachers were randomly assigned to classrooms with different sizes within their schools from kindergarten to third grade. To evaluate such effects, the authors linked the experimental data with administrative records. The main identification strategy exploits the randomization in the program and estimates the ITT parameter. Nonetheless, in the second part of the paper the authors estimate a correlated random coefficient model that allows them to identify the effect of a bundle of class characteristics on both academic achievement and earnings by exploiting random assignment to classrooms. The effect of interest measures the impact of the classroom-level characteristics that varies exclusively because of exogenous variation associated with the START program. In this way, the authors measure the effect of the program on adult outcomes through the causal mechanism associated with the characteristics of the classroom. The main limitation of the paper is that the identification strategy for the causal mechanism is not clear. The authors conclude that classroom characteristics improve both schooling achievement and adult outcomes, in particular, students who were treated are more likely to attend college, live in better neighborhoods, and save more for retirement.

A more formal definition of the causal mechanism of interest is presented in Heckman et al. (2013). The authors examine some possible channels that explain the long-term effects of the Perry Preschool program on labor market outcomes and health behaviors. The main focus is on the effect of the program on the permanent improvement of psychological skills early in the life cycle that at a later point in time resulted in important effects on adult outcomes. Children began the Perry school program at age three and were enrolled for two years. If an analyst is able to observe all possible skills and the mechanisms through which those skills are formed, the identification of the causal mechanism would be straightforward. As it is the case in most datasets, the Perry Preschool program only includes some imperfect measures of different skills. Heckman et al. (2013) combine

the exogenous source of variation associated with the randomization of the program with econometric methods to identify both the effect of the program on skills and through this channels on adult outcomes.

Heckman et al. (2013) defines $T = 1$ if the individual participated in the Perry Preschool program and $T = 0$ otherwise. The potential outcomes for an individual with treatment given by $T = t$ where $t = 0, 1$ is $Y_t = \tau_t + \sum_{j \in \jmath_p} \alpha^j \theta_t^j + X'\beta + \varepsilon_t$, where the vector X includes baseline controls that are exogenous to T, $theta_t^j$ is a vector of observed skill and $\jmath_p$ is an index set for observable skills, τ_t is the contribution of unmeasured skill or any other variable affected directly by the program, ε_t is a zero-mean error term. Notice that the vector of observed skills and unobserved variables depends of the treatment which in this framework is the causal mechanism of interest. This structure is called in the statistics literature meditation analysis (see Imai et al. (2011)). Using such structure, the causal mechanism effects of the program can be decomposed between two channels, the observable skills channel and the unobservable variables channel, i.e. $Y_1 - Y_0 = \tau_1 - \tau_0 + \sum_{j \in \jmath_p} \alpha^j (\theta_1^j - \theta_0^j) + (\varepsilon_1 - \varepsilon_0)$.

The treatment effect of interest is the expected value of the above equation. Using this framework the authors conclude that the program improved externalizing behaviors such as aggressive and rule-breaking behaviors, academic motivation, and academic achievement which, in turn, reduced the probability of being engaging in criminal activities, and improved labor market outcomes and health behaviors in adulthood. Heckman et al. (2013) is the closest paper in spirit to this paper.

2.2 Literature review on the PACES program

The PACES program was designed with the objective of increasing secondary school attendance and reducing the dropout rate for children who were attending public primary schools. The evaluation of the PACES program has been focused on two levels: municipality-school and individual levels. The first evaluation of the PACES program at the municipality-school level was developed by Ribero and Tenjo (1997). The authors used a random sample of schools that participated in the program, and found that the program effectively targeted poor students, and had a low impact on secondary school attendance. Similarly, King, Rawlings, Gutierrez, Pardo, and Torres (1997) implemented an analysis of the PACES program at the municipality level, and found that

in agreement with the design of the program, schools elected to participate in the program were more likely to offer educational services to students in the lowest economic strata. The authors also found that the PACESs schools offered education that was as good as the educational service offered in public schools.

At the individual level, the focus has been on the policy evaluation of the program on educational outcomes. The common strategy uses linear regression of the outcome of interest Y_i as a function of some baseline controls, X_i, and an indicator variable T_i for whether the student won the scholarship: $Y_i = X_i'\beta +_i +\varepsilon_i$. The main source of information in all cases is the 4044 application forms of the students who applied to the PACES program in 1994 to enter to a private school in sixth grade in 1995 in Bogotá.[3]

The first policy evaluation of the PACES program at the individual level was implemented by Angrist et al. (2002). They interviewed 1176 students from the pool of 4044 applicants of the 1995 cohort in Bogotá. Additionally, they collected standardized test scores for 283 participants in the survey subsample three years after the randomization of the program.[4] For the subsample of 1166 students, they found that three years after the treatment, PACES winners had similar enrollment rates in relation to the losers, but they had completed 0.12 additional years of schooling, and were 6 percent less likely to have repeated a grade with respect to losers. Also, using the test score of 283 students, they found that winners score was 0.2 standard deviations higher than losers.

Angrist et al. (2006) evaluated the effect of PACES on high school graduation and academic achievement using Colombia's centralized college entrance examinations. The authors used all the 4044 applicants[5] and exploited the exogenous variation associated with the lottery applied in the selection of participants in Bogotá. Additionally, Angrist et al. (2006) matched the baseline sample with the administrative records of Colombia's centralized college entrance examinations, ICFES

[3]A detailed description of the PACES program can be find in section 6.1.

[4]Angrist et al. (2002) used *La prueba de Realización*. They invited 473 applicants for testing, but only the 60 percent effectively answer the invitation (283 students).

[5]The final empirical work was done only with those applicants with full information on the baseline characteristics which include 3542 observations.

exam, administered just before graduation from high school in Colombia. These records allow them to observe the students who finished high school during $1999 - 2001$ period. The results suggest that the PACES program had a substantial impact on both high-school graduation rates and achievement; in particular, they found that high-school graduation rates increased from 5 to 7 percent, and also found an increase of 2 to 4 points in language scores and 2 to 3 points in math scores.

Bettinger et al. (2010) extend the previous framework. They consider a model that includes two possible channels through which scholarship could affect academic outcomes: (i) the productive effect and (ii) the redistributive effect (or peer effects of the program).[6] Now, the impacts of the program over educational outcomes of student i are given by $Y_i = X_i'\beta + \bar{X}_s{}'\bar{\beta} +_i +\varepsilon_i$, where $\bar{X}_s$ represents the average level of X_i in school s. In this framework $\bar{\beta}$ is the redistributive effect and γ is a productive effect. Notice that differences in the distributive effect between treated and untreated students after randomization is given by $\gamma(\bar{X}_t - \bar{X}_u)$. Bettinger et al. (2010) considered three hypothesis: (1) $\gamma > 0$ and $\bar{\beta} = 0$, i.e. the productive effect is equal to the total effect. (2) $\bar{\beta} > 0$ and $\gamma = 0$, i.e. the productive effect is zero, but the peer effects are positive. (3) $\bar{\beta} > 0$ and $\gamma > 0$. The total effect of the program, which is given by the difference in the outcome between treated and untreated, is $\bar{\beta}(\bar{X}_t - \bar{X}_u) + \gamma$. To estimate such a model, Bettinger et al. (2010) use the subsample of Angrist et al. (2002) and new data from a survey that they gave to the schools principal. Additionally, the authors split the sample between those who applied to private academic schools and those who applied to private vocational schools. They found that for students who applied to vocational schools, treated applicants were 5 to 6 percent more likely to graduate from high school than untreated. In the cases of non-vocational schools, the effects were 3 to 6 percent. Additionally, the authors showed significant and positive effects of the program in both math and reading among vocational students (although this effect is not significant in non-vocational schools). These results imply that the results of Angrist et al. (2006) may have a slight downward bias.

[6] Bettinger et al. (2010) assume that all schools have the same number of student.

In the same line as previous literature, Bettinger et al. (2014), applying the standard estimation strategy, found that agents that were treated were less likely to repeat grades, and more likely to graduate from secondary school. The authors extended previous literature on the PACES program by evaluating the impact of the program on long-term outcomes, like fertility, college education, and adult labor market outcomes using a novel database. In particular, Bettinger et al. (2014) found that those treated by the program were 6 percentage points less likely to have had a child as a teen, although the effect on overall fertility by age 30 is not significant. Similarly, they found that lottery winners were more likely to obtain a college education, and the total formal sector earnings at age 30 are 8 percent higher, although the effect is not precisely estimated at the 5% level (it is significant at the 7% level). Impacts on estimated future earnings, including imputed values for those currently in tertiary education suggest a 9.3 percent higher for lottery winners.

My empirical application uses Bettinger et al. (2014) data and some additional Colombias administrative records. Similar to the literature on the long-term effects of policy interventions at early stages of life and the PACES program, I exploit the experimental variation associated with the PACES program in order to identify the causal mechanisms effects. Nonetheless my identification strategy is different, and I depart from the common strategy that uses linear regression for the outcome of interest as a function of an indicator of whether the individual got the scholarship. Analogous to Heckman et al. (2013), I combine the experimental variation of the PACES program with an appropriate econometric model that allows me to clearly specify the causal mechanism and the parameter of interest. Unlike Bettinger et al. (2014) that compute the intent-to-treat parameter, my main focus on this dissertation is in the Policy-Relevant treatment parameter which focuses on the causal mechanism and only for those individuals for which the scholarship made them switch their schooling choices.

CHAPTER 3

A SIMPLE ECONOMIC MODEL OF EDUCATIONAL ATTAINMENT WITH AN SCHOLARSHIP

Below I describe a simple dynamic model of schooling attainment and work decisions that are affected by a tuition scholarship assigned randomly and lowers the net cost of schooling. The objective of such model is to help to understand the causal mechanisms through which the scholarship affects the incentives of individuals. The key implication of this model is that individual schooling choices are affected by three possible channels: the scholarship (i) reduces the net cost of schooling in the current period which in turns positively affects the current schooling attainment choice, (ii) reduces the next period cost of schooling conditional on passing the current schooling year, and then it also positively affects the current schooling attainment choice, and (iii) reduces work participation in the current period.

I consider an individual who enters the first year of secondary education ($t = 0$). At the beginning of each period an individual chooses whether to attend school (s_{it}) and/or whether to work (h_{it}) in the labor market. The period utility of each individual is a function of consumption C_{it}, schooling attainment status s_{it}, work status h_{it}, and a time invariant individual-specific taste for schooling/work choice: $u_i(C_{it}, s_{it}, h_{it}, \epsilon_i)$.[1] For simplicity, I assume that u_i is increasing in C_{it} and s_{it}, but decreasing in h_{it}.

The individual maximizes the discounted value of expected utility at time t subject to the period t budget constraint:

$$C_{it} + (p_{it} - \mathbb{S}_{it}) * s_{it} = Y_{it} + w_{it} * h_{it} \tag{3.1}$$

Where p_i is the individual net cost of schooling before scholarship, $\mathbb{S}_{it}$ is the scholarship which

[1] A possible form for the utility function is CRRA: $u_i(C_{it}, s_{it}, h_{it}, \epsilon_i) = \frac{1}{\gamma} C_{it}^{\gamma} + \delta_s s_{it} + \delta_h h_{it} + \epsilon_i$, where $\delta s > 0$ measures the utility of schooling, $_h < 0$ measures the disutility of work, and ϵ_i is the time invariant individual-specific taste for schooling/work choice.

depends on previous period attainment (i.e. conditional on passing the previous grade) and it is assumed to be less than p_{it} but greater or equal than zero, Y_{it} is the households income, and w_{it} is the individual wage for working.[2] The scholarship is given by the following function

$$\mathbb{S}_{it} = \mathbb{S}^*_{it} \quad if \quad s_{it-1} = 1 \quad \text{and} \quad Won \quad scholarship \quad in \quad t = 0, \quad \mathbb{S}_{it} = 0 \quad otherwise \tag{3.2}$$

where $\mathbb{S}^*_{it}$ is the amount of the tuition covered by the scholarship each period. The net cost of schooling is given by $NCS_i = p_i - S_{it}$. Once the secondary education ended, $\mathbb{S}_{it} = 0$ for everyone.

The agent education law of motion is $edu_{it+1} = edu_{it} + s_{it}$, where $edu_{i0} = 5$.[3] The wage w_{it} is a function of previous schooling decisions edu_{it-1}, and an individual-specific ability, ε_i, that affects labor market outcomes.[4] The variables ϵ_i and ε_i are jointly distributed, and represent the individual taste for education and abilities created before the start of the human capital accumulation process. ϵ_i and ε_i are only observable for the individual i. At every period, the individual chooses s_{it} and h_{it} to maximize the value function, which is given by

$$V_{it}(\Omega_{it}) = \max_{S_{it}, h_{it}} u_i(C_{it}, s_{it}, h_{it}, \epsilon_i) + \beta E_t V_{it+1}(\Omega_{it+1}) \quad t < \bar{T} \tag{3.3}$$

where β is the discount factor, $barT$ is the terminal decision period, and Ω_{it} is the state space which includes all the relevant factors affecting current or future utility.[5] In the terminal period, $t = \bar{T}$, the value function is given by $V_{it}(\Omega_{i\bar{T}}) = u_i(\Omega_{i\bar{T}})$. In the empirical framework, the components ϵ_i and εi are assumed to be known by the agent but unobserved by the analyst. This characteristic of

[2] If the individual attended school on period t, but at the end of the year he does not pass, then this model assumes that $s_{it} = 0$ and there is not accumulation of human capital. This implies that the individual will not receive the scholarship the next period.

[3] $edu_{i0} = 5$ is equivalent to number of years accumulated during elementary school.

[4] A possible function is the standard Mincer equation: $\log(w_{it}) = \alpha_0 + \alpha_1 edu_{it} + \varepsilon i$. Alternatively, and since education premiums are increasingly convex, I can use dummy variables to estimate different premiums for different levels of education. The Mincer equation may include other additional controls.

[5] The terminal period $\bar{T}$ is the last year of college education. The value function may also include time-varying unobserved shocks that are independent of the scholarship.

the model creates two common problems: dynamic selection, since the taste for education affect past, current and future schooling choices, and essential heterogeneity in the sense of Heckman et al. (2006) since ε_i may be correlated with ϵ_i, and affects the individual return to education.

3.1 The scholarship causal mechanism effect

Let the value function when $s_{it} = 1$ and for a given work choice be given by the choice-specific value function define as $v_t(1)$,

$$v_t(1) = \max_{h_{it}} u_i(Y_{it} + w_{it} * h_{it} - (p_{it} - \mathbb{S}_{it}), 1, h_{it}, \epsilon_i) + \beta E_t V_{it+1}(\Omega_{it+1}) \tag{3.4}$$

Similarly, let the choice-specific value function for $s_{it} = 0$ be defined by $v_t(0)$,

$$v_t(0) = \max_{h_{it}} u_i(Y_{it} + w_{it} * h_{it}, 0, h_{it}, \epsilon_i) + \beta E_t V_{it+1}(\Omega_{it+1}) \tag{3.5}$$

Since the scholarship renewal is conditional on passing the previous schooling period, then the value of scholarship is zero for the future value of the choice-specific value function $v_t(0)$. Notice that there are individuals insensitive to the scholarship: students such that $v_t(1) > v_t(0)$ for $\mathbb{S}_{it} = 0$ or $v_t(1) < v_t(0)$ for all values of $\mathbb{S}_{it} > 0$. Nonetheless, for some individuals the scholarship is relevant and affects the schooling choices. The scholarship is policy-relevant for individuals such that

$$v_t(1) < v_t(0) \quad \mathbb{S}_{it} = 0 \quad and \quad v_t(1) > v_t(0) \quad for \quad some \quad value \quad \mathbb{S}_{it} > 0 \tag{3.6}$$

The scholarship affects individual choices in three ways. First, it reduces the current net cost of schooling which in this case positively affects the current utility function, and then positively affects schooling attainment in period t. Second, it reduces the next period cost of schooling conditional on passing the current schooling year, and then it also positively affects the current schooling attainment choice. Finally, the scholarship induces an income effect that negatively affects the work choice.[6] Since the program is assigned randomly, the average treatment effect of

[6]Notice that for individuals with relative high cost of schooling, but still willing to attend school, the scholarship should induce individuals to work in order to complement the scholarship and be able to finance the cost of schooling. I

the program T on the level of schooling edu_{it} may be identified.

The scholarship no longer has an effect on human capital accumulation in period t in two cases: (i) if the agent fails to pass previous grade, $s_{it-1} = 0$, and (ii) if the program ends (end of secondary education) or the government does not provide scholarship resources. Nonetheless, once the scholarship program ends, some individuals may choose to accumulate additional years of education. After formal schooling ends in period , the final level of schooling is constant and given by $edu_{i\bar{T}}$. The log earnings for period $t > \bar{T}$ are a function of the final level of education:

$$\log(w_i) = \alpha_0 + \alpha_1 edu_{i\bar{T}} + \varepsilon i \tag{3.7}$$

Notice that the scholarship has a causal effect on wages through the accumulation of human capital.[7] Given the presence of the confounding factors ϵ_i and ε_i, the randomization of the program is not sufficient for the identification of the causal effect of $\mathbb{S}_{it}$ on $\log(w_{it})$ through edu_{it}. The objective of the paper is to test the strength of the causal effect of the scholarship $\mathbb{S}_{it}$ on adult wages for those individuals for whom the program was relevant in terms of the human capital accumulation process.[8] The next section introduces the econometric framework that matches the constraints in the data, and introduces the statistical assumptions that are sufficient for the identification of the effect of interest. The main interest is to evaluate the total program impact on wages. My econometric framework does not allow one to disentangle the three channels mentioned above.

expect that in general the income effect will dominate, and then the scholarship will overall reduce work participation.

[7]Here I focus on the effect of the program through the accumulation of years of schooling. Nonetheless, it is possible that the program negatively affects the accumulation of experience since the scholarship reduces work participation in period t. In the case of the PACES program in Colombia, it is likely that the latter effect is negligible since the program affects students between 11 and 15 years old in secondary education and the options available for this population are usually temporality and manual jobs that do not require any special expertise.

[8]For instance, for those individuals that the receive an amount $\mathbb{S}_t^*$ and the program was relevant, I would like two compare two possible scenarios: wages with and without scholarship, $\log(w_{it})(\mathbb{S}_{ti} = \mathbb{S}_t^*) - \log(w_{it})(\mathbb{S}_{ti} = 0)$

CHAPTER 4

FRAMEWORK: A SEQUENTIAL MODEL OF EDUCATIONAL ATTAINTMENT

Below I outline a sequential discrete choice model [1] for evaluating the effect of a tuition scholarship program that exogenously affects an early schooling decision, and through this channel, other schooling choices and adult labor market outcomes. My sequential model matches the data available, and controls for dynamic selection and accounts for the presence of essential heterogeneity (Heckman et al. (2006)). Interestingly, Cunha, Heckman, and Navarro (2007) provide the specific conditions under which a dynamic discrete choice model of schooling, like Keane and Wolpin (1997), can be represented by a reduced form approximation like the one below.[2]

My sequential schooling model includes three transitions (see figure 4.1). At each transition the individuals choice set includes two alternatives, either transit to the next level of education or dropout from schooling system.[3] I define J as the set of possible schooling states, where j denotes a possible state available to the individual. In my framework there are four possible states: $j = 0$ denotes the state of high school (HS) incomplete, $j = 1$ is the state of being a HS graduate, $j = 2$ is attending college, and $j = 3$ denotes college graduate. Let $D_{j,j'} = 1$ if a person at state j chooses to transit to state j' at the decision node $\{j, j'\}$, where j, j', and $D_{j,j'} \in \mathfrak{D}$, the set of all possible schooling decisions taken by an individual. For instance, if an individual at the decision

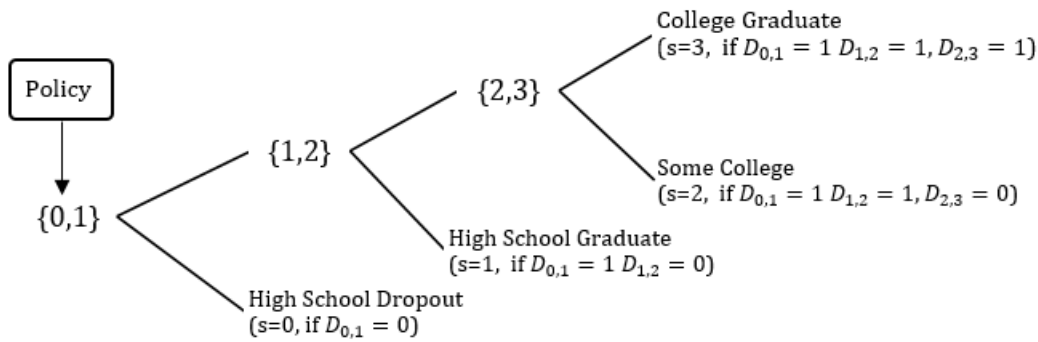
[1]This framework was original developed by Cameron and Heckman (2001). An extension is presented by Heckman and Navarro (2007). Other applications of this framework can be found in Heckman, Humphries, Urzua, and Veramendi (2011), Reyes, Rodrguez, and Urza (2013). Similar ideas has been applied in other context such as Cooley, Navarro, and Takahashi (2014).

[2]My set up satisfies the key condition stated in Cunha et al. (2007) Once an individual in the PACES program leaves school, he never returns.

[3]In my data I only observe individuals who continue in the education system or dropout from it. I do not observe individuals enrolling in graduate school by the last year that the data is available. I excluded from our model schooling decisions beyond undergraduate school and the possibility of return once the individual drops from the educational system. In this model decisions are irreversible. In the data I do not observe transitions year by year, only level by level of education transitions are available.

node $\{0,1\}$ chooses to transit to state $j = 1$, then $D_{0,1} = 1$. If this individual chooses to stay in the state $j = 0$ (and not transit to state $j = 1$), then $D_{0,1} = 0$. If an individual completes the three possible transitions, then $D_{0,1} = 1$, $D_{1,2} = 1$, $D_{2,3} = 1$. I also define s as the realized final level of

Figure 4.1: Sequential Schooling Model



schooling, where s, and $S = \{0, 1, 2, 3\}$. For instance, in the last example $s = 3$ (see figure 4.1), and the individual in the former case, who chooses to stay in the state $j = 0$ ($D_{0,1} = 0$), $s = 0$. There are four possible final schooling levels: HS dropout ($s = 0$), HS graduate ($s = 1$), college dropout ($s = 2$), and college graduate ($s = 3$). The optimal decision at node $\{j, j'\}$ is represented by a threshold-crossing model of the form $D_{j,j'} = 1[V_{j,j'}0]$, where $V_{j,j'}$ is the value perceived by an individual with current state j of attaining schooling level j', and $1[.]$ is the logical indicator function taking the value 1 if its argument is true and the value 0 otherwise. The value function $V_{j,j'}$ is assumed to be a linear (separable) index of Z and ϵ[4]

$$V_{j,j'} = Z_{j,j'}\beta_{j,j'} + \epsilon_{j,j'} \quad if \quad D_{j-1,j'-1} = 1 \tag{4.1}$$

Where $\{j, j'\} = (\{0, 1\}, \{1, 2\}, \{2, 3\})$, and $Z_{j,j'}$ is a vector of variables observed by both the agent and the econometrician that determine the schooling transition decision of an agent at the

[4]Vytlacil (2006) and Vytlacil (2002) demonstrate that assuming a linear or additively separable latent index function is not as restrictive as it appears. Under some conditions, it is possible to represent a non-separable latent index as a latent index with additive separability between observed and unobserved variables. Matzkin (2003) also shows some cases where this assumption can be relaxed.

decision node $\{j, j'\}$, $\beta_{j,j'}$ is a vector of parameters, and $\epsilon_{j,j'}$ is a vector of endowments observed only by the individual but unobserved by the econometrician. $Z_{j,j'}$ may include policy variables that affect the transition from j to j', for instance a scholarship for high school education (see Figure 4.1), or a subsidy to college tuition. Notice that $V_{0,1}$ is a reduced form of the value function given in section 3 that comprise all schooling years during the secondary education.[5] Each final level of schooling s is associated with a set of k potential labor market outcomes[6] (e.g. wages, have a formal job, number of days worked, poverty status, etc.) given by

$$Y_{k,s} = X_{k,s}\gamma_{k,s} + \varepsilon_{k,s} \tag{4.2}$$

where $X_{k,s}$ is a vector of observed variables of the individuals relevant for outcome $Y_{k,s}$, $\gamma_{k,s}$ is a vector of parameters, and $\varepsilon_{k,s}$ is a vector of endowments observed only by the agent. The k observed outcomes Y_k can be written in a switching regression form. Define $H_s = 1$ if s is the final level of schooling reached by the individual and $H_s = 0$ otherwise. Then the observed outcome is

$$Y_k = \sum_{s=0}^{3} H_s Y_{k,s} \tag{4.3}$$

4.1 Factor structure

In general, there are common unobserved endowments associated with both $\varepsilon_{k,s}$ and $\epsilon_{j,j'}$.[7] If the decision to transit to the next decision node of an agent at a particular decision node j, j' is correlated with the unobservable gains from the final schooling level, I have a situation with essential heterogeneity in terms of Heckman et al. (2006).[8] In this context essential heterogeneity

[5]For instance, $V_{0,1} > 0$ means that the choice-specific value function $v_t(1)$ is greater than $v_t(0)$ for $t = 1, \cdots, 6$ which represent the years in the secondary education. If, $V_{0,1} < 0$, then $v_t(1) < v_t(0)$ for some $t = 1, \cdots, 6$. Similarly for $V_{1,2}$ and $V_{2,3}$.

[6]Outcomes may be either continuous or discrete.

[7]$\varepsilon_{k,s}$ and $\epsilon_{j,j'}$ are statistical independent, then the best estimation method is to use linear regressions methods to estimate the parameters $\gamma_{k,s}$ and probit models to estimate the parameters $\beta_{j,j'}$. See Maddala (2010) for versions of switching regression models with exogenous and endogenous switching.

[8]See also Ravallion (2015) for a discussion.

means that individuals make rational choices about whether to transit to the next level of education based on endowments they know but are unobservable to the analyst, and it is expected that those agents with higher unobserved endowments are more likely to select into a higher level of education and experience higher gains from that choice. As a consequence, in general, I cannot identify the joint distribution of outcomes across level of education, and I am not able to identify the treatment on the treated parameters or policy relevant parameters discussed in the following sections (see Carneiro, Hansen, and Heckman (2003), Heckman and Navarro (2007)).

To account for essential heterogeneity I must recover the joint distribution $\varepsilon_{k,s}$ and $\epsilon_{j,j'}$ for all $s = 0, 1, 2, 3$. To simplify this task, I impose a factor structure for $\varepsilon_{k,s}$ and $\epsilon_{j,j'}$ that allows me to recover the joint distribution of (Y_0, Y_1, Y_2, Y_3) (see Aakvik, Heckman, and Vytlacil (1999), Aakvik, Heckman, and Vytlacil (2005) and Carneiro et al. (2003)).

Assumption 1: $\epsilon j, j' = \alpha_{j,j'}\theta + v_{j,j'}$ and $\varepsilon_{k,s} = \alpha_{k,s}\theta + u_{k,s}$, where $\alpha_{k,s}$ and $\alpha_{j,j'}$ are factor loadings, θ is a factor that is independent of all $v_{j,j'}$, $u_{k,s}$, and observable variables. Also, $v_{j,j'}$ is statistical independent $v_{i,i'}$ for all $\{j, j'\} \neq \{i, i'\}$, $u_{k,s}$ is statistical independent of $u_{l,s}$ for all k, and $v_{j,j'}$ is statistical independent of $u_{k,s}$ for all $\{j, j'\}$ and k. I also assume $v_{j,j'}$, $u_{k,s}$ are independent distributed random variables with mean zero and finite variance.[9]

The factor structure also helps me control for dynamic selection, i.e. different types of students might select into specific nodes over the three decisions based on unobservable characteristics. For instance, suppose that θ is known and $\theta = \bar{\theta}$. The value of graduating from high school for a student of type $\bar{\theta}$, at the decision node $\{0, 1\}$ (i.e. $V_{0,1} = Z_{0,1}\beta_{0,1} + \alpha_{0,1}\bar{\theta} + v_{0,1}$), may be correlated with the value of enrollment in college conditional on $D_{0,1} = 1$, at the decision node $\{1, 2\}$ (i.e. $V_{1,2} = Z_{1,2}\beta_{1,2} + \alpha_{1,2}\bar{\theta} + v_{1,2}$). Nonetheless, conditional on , the optimal decision at the node $\{1, 2\}$, given by $D_{2,1}$, is independent of the optimal choice at the node $\{2, 3\}$, given by $D_{0,1}$.

4.2 Identification

I define identification in a standard way. In general, the identification of the parameters $\beta_{j,j'}$, $\gamma_{k,s}$ and the distributions $F_{\varepsilon_{k,s}}$,$F_{\epsilon_{j,j'}}$ is possible if there is sufficient variation in the components

[9] In practice I can relax the statistical independence assumption for a general mean independence assumption. I can also allow a distribution different from normal for $v_{j,j'}$,$u_{k,s}$.

$(Z_{j,j'}, \beta_{j,j'})$ for each value of the index functions, i.e. there is measurable separability among the index value functions. In propositions 1 and 2 in the appendix (see appendix C), I show the sufficient conditions for the identification of a model specified by Equations (1) and (2) and given the characteristics of my data, in particular, I exploit the randomization of the scholarship that affects the first transition.[10]

These results do not allow me to identify the joint distribution of Y_0, Y_1, Y_2, Y_3 because I only observe one of these outcomes for any individual, which implies that I cannot identify treatment on the treated parameters. I use the factor structure in assumption 1 to identify the joint distribution of Y_0, Y_1, Y_2, Y_3 and the policy relevant parameters of interest.

My identification approach exploits the factor structure to control for the possibility of dynamic selection and essential heterogeneity. I first show how the factor structure allows me to obtain the conditional independence assumption of quasi-experimental methods. Then, I show how to identify the joint distribution of $\varepsilon_{k,s}$ and $\epsilon_{j,j'}$ for all $s = 0, 1, 2, 3$.[11] I keep all conditioning on covariates implicit to simplify notation.

Propositions 1 and 2 in the appendix C provide sufficient conditions for the identification of $\beta_{j,j'}$, $\gamma_{k,s}$ and the distributions $\mathrm{F}_{\epsilon_{j,j'}}$,$\mathrm{F}_{\varepsilon_{k,s}}$ for all s, but they are not sufficient to recover the joint distribution of Y_0, Y_1, Y_2, Y_3. Assumption 1 reduces the problem of recovering the joint distribution $\varepsilon_{k,s}$ and $\epsilon_{j,j'}$ to the problem of recovering the loading parameters $\alpha_{k,s}$ and $\alpha_{j,j'}$, and the marginal distribution of θ, $v_{j,j'}$, and $u_{k,s}$.

Now I illustrate the problem using the structure of my schooling model with one outcome which includes three value index functions and four potential outcomes:

---

[10] Sufficient conditions for a other classes of models can be found in Heckman and Navarro (2007), Cunha et al. (2007), Abbring and Van den Berg (2003),Carneiro et al. (2003).

[11] See Cunha, Heckman, and Schennach (2010) and Heckman and Navarro (2007) for more general cases.

$$
\begin{cases}
V_{0,1} = Z_{0,1}\beta_{0,1} + \alpha_{0,1}\theta + v_{0,1} \\
V_{1,2} = Z_{1,2}\beta_{1,2} + \alpha_{1,2}\theta + v_{1,2} \\
V_{2,3} = Z_{2,3}\beta_{2,3} + \alpha_{2,3}\theta + v_{2,3} \\
Y_0 = X_0\gamma_0 + \alpha 0\theta + u_0 \\
Y_1 = X_1\gamma_1 + \alpha 1\theta + u_1 \\
Y_2 = X_2\gamma_2 + \alpha 2\theta + u_2 \\
Y_3 = X_3\gamma_3 + \alpha 3\theta + u_3
\end{cases}
\tag{4.4}
$$

I set the scale of θ by normalizing the factor loading $\alpha_{0,1} = 1$.[12] The unobservable gains of a final schooling level s' with respect to s where $s' > s$ are given by $\varepsilon_{s'} - \varepsilon_s = (\alpha_{s'} - \alpha_s)\theta + (u_{s'} - u_s)$. Notice that $\alpha_{s'} - \alpha_s \neq 0$ implies the presence of essential heterogeneity in my framework because the unobserved gains are correlated with the errors of the index functions, and then with the optimal schooling choices (i.e. $D_{0,1}$,$D_{1,2}$,$D_{2,3}$) since θ affects both.

I also suppose that the first schooling choice $D_{0,1}$ is not correlated with any previous schooling choice, i.e. it is free of dynamic selection,[13] which in turn implies that the parameters of $V_{0,1}$ do not depend on previous schooling choices. For all other choices the selection process is dynamic in the sense that transition choice at the node j, j' depends on the transition choice at the node $j-1, j'-1$. For instance, individual choice at node $\{1, 2\}$ is only possible if at the node $\{0, 1\}$ the optimal choice by the individual was $D_{0,1} = 1$. Since the unobserved endowments θ influence decisions in consecutive transitions, and given that $\alpha_{1,2} \neq 0$, $\alpha_{2,3} \neq 0$, dynamic selection does not allow me to identify the model.

[12] In general, the scale of θ is unknown, since θ is an unobservable. So any other normalization will be as good as $\alpha_{0,1} = 1$.

[13] This assumption is natural in our specific problem since the students who participated in the PACES program were selected by a lottery mechanism in the first year of high school, so everyone in the sample finished elementary school (see Angrist et al. (2002), Angrist et al. (2006)). Also, the dropout rate for elementary education in Bogota by 1995 was low at around 4%.

The assumption for the first schooling choice is relaxed later when I correct for selection using as an exclusion restriction, the exogenous variation associated with the random assignment of PACES scholarships. I further include additional exclusion restrictions to secure measurable separability between $V_{1,2}$ and $V_{2,3}$. I estimate the model under several specifications. The main specification includes the randomization of the scholarship as an exclusion restriction for the index value function $V_{0,1}$. For the indices functions $V_{1,2}$ and $V_{2,3}$, I use the same vector of covariates, i.e. $Z_{1,2} = Z_{2,3}$ where at least two of the variables excluded from $V_{0,1}$ are continuous. The second specification includes specific exclusion restrictions for $V_{2,3}$, which implies that $Z_{1,22,3}$. The proposition 1 and 2 in the appendix C shows sufficient conditions for identification for each case. The results are similar in both cases.

If the factor θ were observed, I can condition my model on θ and then all my educational choices and outcomes will be statistically independent. Given θ, my identification strategy is to assume that conditional independence, i.e. (Y_0, Y_1, Y_2, Y_3) is independent of $D_{0,1}, D_{1,2}, D_{2,3}$ conditional on θ holds.[14]

Now I need to show how the joint distribution of $\varepsilon_{k,s}$ and $\epsilon_{j,j'}$ is identified. Given the normalization that $\alpha_{0,1} = 1$, the parameters $\alpha_{j,j'} = 1$, $\alpha_{k,s} = 1$, and the marginal distributions of θ and $v_{j,j'}$, and $u_{k,s}$ are identified nonparametrically. Given that the first schooling choice is free of selection, and assuming that the loading factors of the schooling model and variance are non-zero, I can form the following covariance

$$\begin{cases} Cov(\epsilon_{0,1}, \epsilon_{1,2}) = \alpha_{1,2}\sigma_\theta^2 \\ Cov(\epsilon_{0,1}, \epsilon_{2,3}) = \alpha_{2,3}\sigma_\theta^2 \\ Cov(\epsilon_{1,2}, \epsilon_{2,3}) = \alpha_{1,2}\alpha_{2,3}\sigma_\theta^2 \end{cases} \tag{4.5}$$

Then, given that $Cov(\epsilon_{0,1}, \epsilon_{1,2}) \neq 0$, $Cov(\epsilon_{0,1}, \epsilon_{2,3}) \neq 0$, and $Cov(\epsilon_{1,2}, \epsilon_{2,3})$, then

[14] I actually integrate out θ and its distributions are identified. Notice that the factor structure extends the idea behind the matching estimator. In this context the idea is to match based not only on variables observable to the econometrician but also on the unobservable factors (similar arguments are applied by Cooley et al. (2014) and Heckman et al. (2011). Heckman and Navarro (2007) and Cunha et al. (2010) generalized the idea of the extended matching.

$$\begin{cases} \alpha_{1,2} = \frac{Cov(\epsilon_{1,2},\epsilon_{2,3})}{Cov(\epsilon_{0,1},\epsilon_{2,3})} \\ \alpha_{2,3} = \frac{Cov(\epsilon_{1,2},\epsilon_{2,3})}{Cov(\epsilon_{0,1},\epsilon_{1,2})} \\ \sigma_\theta^2 = \frac{Cov(\epsilon_{0,1},\epsilon_{1,2})Cov(\epsilon_{0,1},\epsilon_{2,3})}{Cov(\epsilon_{1,2},\epsilon_{2,3})} \end{cases} \tag{4.6}$$

Similarly, assuming that the loading factor of the potential outcomes model are not zero, and using the structure of the potential outcomes I have

$$\begin{cases} Cov(\varepsilon_0, \varepsilon_1) = \alpha_0\alpha_1\sigma_\theta^2 \\ Cov(\varepsilon_0, \varepsilon_2) = \alpha_0\alpha_2\sigma_\theta^2 \\ Cov(\varepsilon_0, \varepsilon_3) = \alpha_0\alpha_3\sigma_\theta^2 \\ Cov(\varepsilon_1, \varepsilon_2) = \alpha_1\alpha_2\sigma_\theta^2 \\ Cov(\varepsilon_1, \varepsilon_3) = \alpha_1\alpha_3\sigma_\theta^2 \\ Cov(\varepsilon_2, \varepsilon_3) = \alpha_2\alpha_3\sigma_\theta^2 \end{cases} \tag{4.7}$$

Then

$$\begin{cases} \frac{Cov(\varepsilon_0,\varepsilon_1)}{\varepsilon_0,\varepsilon_2)} = \frac{\alpha_1}{\alpha_2} \\ \frac{Cov(\varepsilon_0,\varepsilon_1)}{\varepsilon_0,\varepsilon_3)} = \frac{\alpha_1}{\alpha_3} \\ \frac{Cov(\varepsilon_1,\varepsilon_2)\varepsilon_0,\varepsilon_2)}{Cov(\varepsilon_0,\varepsilon_1)} = \alpha_2^2\sigma_\theta^2 \end{cases} \tag{4.8}$$

And

$$\begin{cases} \alpha_2 = \sqrt{\frac{Cov(\varepsilon_1,\varepsilon_2)\varepsilon_0,\varepsilon_2)}{Cov(\varepsilon_0,\varepsilon_1)\sigma_\theta^2}} \\ \alpha_1 = \frac{Cov(\varepsilon_0,\varepsilon_1)}{Cov(\varepsilon_0,\varepsilon_2)}\alpha_2 \\ \alpha_3 = \frac{Cov(\varepsilon_0,\varepsilon_3)}{Cov(\varepsilon_0,\varepsilon_2)}\alpha_2 \end{cases} \tag{4.9}$$

Also, if I assume that the distribution of θ is not symmetric[15] and $\mathbf{E}[\theta^3] \neq 0$,[17] and given that the first schooling transition is free of selection, I can form

$$\begin{cases} \mathbf{E}[\varepsilon_0^2 \epsilon_{0,1}] = \alpha_0^2 \mathbf{E}[\theta^3] \\ \mathbf{E}[\varepsilon_1^2 \epsilon_{0,1}] = \alpha_1^2 \mathbf{E}[\theta^3] \\ \mathbf{E}[\varepsilon_2^2 \epsilon_{0,1}] = \alpha_2^2 \mathbf{E}[\theta^3] \\ \mathbf{E}[\varepsilon_3^2 \epsilon_{0,1}] = \alpha_3^2 \mathbf{E}[\theta^3] \\ \mathbf{E}[\varepsilon_0 \epsilon_{0,1}^2] = \alpha_0 \mathbf{E}[\theta^3] \\ \mathbf{E}[\varepsilon_1 \epsilon_{0,1}^2] = \alpha_1 \mathbf{E}[\theta^3] \\ \mathbf{E}[\varepsilon_2 \epsilon_{0,1}^2] = \alpha_2 \mathbf{E}[\theta^3] \\ \mathbf{E}[\varepsilon_3 \epsilon_{0,1}^2] = \alpha_3 \mathbf{E}[\theta^3] \end{cases} \tag{4.10}$$

Then

$$\begin{cases} \frac{\mathbf{E}[\varepsilon_0^2 \epsilon_{0,1}]}{\mathbf{E}[\varepsilon_0 \epsilon_{0,1}^2]} = \alpha_0^2 \\ \frac{\mathbf{E}[\varepsilon_0^2 \epsilon_{0,1}]}{\mathbf{E}[\varepsilon_1 \epsilon_{0,1}^2]} = \alpha_1^2 \\ \frac{\mathbf{E}[\varepsilon_0^2 \epsilon_{0,1}]}{\mathbf{E}[\varepsilon_2 \epsilon_{0,1}^2]} = \alpha_2^2 \\ \frac{\mathbf{E}[\varepsilon_0^2 \epsilon_{0,1}]}{\mathbf{E}[\varepsilon_3 \epsilon_{0,1}^2]} = \alpha_3^2 \end{cases} \tag{4.11}$$

Finally, I can construct the joint distribution of $\varepsilon_{k,s}$ and $\epsilon_{j,j'}$. From Kotlarskis theorem[18] (see appendix B) and given the identification of the loading factors, it follows that that the distribution of θ, $v_{j,j'}$, and $u_{k,s}$ is nonparametrically identified. From this distribution I can recover the standard treatment effect parameters and the policy relevant parameters discussed in the next section.

---

[15][16] shows that if θ is non-Gaussian, second-to-fourth-order moments may be used in the identification of the full set of loading factors. Under normality this is strategy is not valid.

[17]In particular, I assume that θ is distributed as a mixture of two normals. We also estimate the model assuming a finite mixture distribution with two and three types. The model can be estimated assuming normality for the distribution of θ but further normalizations will be necessary (for instance, σ_θ^2 cannot be estimated. Aakvik et al. (2005) shows a specific set of normalization).

[18] See Kotlarski (1967) in the appendix B

My model is identified without making any distributional assumptions on θ, $v_{j,j'}$, and $u_{k,s}$. For the empirical work I assume that $f(\theta)$ is approximated by a mixture of normals[19], and $v_{j,j'}$, $u_{k,s}$ are independent standard normal distributed. The additional assumption of valid exclusion restrictions aids the identification of the sequential schooling model, although it is not strictly necessary for all transitions (see Heckman and Navarro (2007)). The exclusion restrictions are variables that affect individuals specific schooling transitions but do not affect the adult outcomes directly. For the first transition I exploit the exogenous variation associated with random assignment of PACES scholarships.

4.3 Defining Treatment Effects

The set of assumptions of my model allows me to identify the joint distribution of $\varepsilon_{k,s}$ and $\epsilon_{j,j'}$. Using these distributions I can estimate the distributions of counterfactual outcomes. As before, I keep all conditioning on covariates implicit. The main interest is to compute the causal effect of the scholarship program on wages through the accumulation of human capital. In particular, the scholarship program had a significant effect on the transition from high school (HS) incomplete to HS graduate. The focus is on those individuals who without the scholarship would not finish HS but if they have access to the scholarship they would graduate from HS. In the counterfactual scenery without scholarship, those individual would finish formal schooling as a high school dropouts, while under the counterfactual scenario with the scholarship they would finish as a high school graduate or college dropout or college graduate. These in turn may affect the level of the returns of education on the adult labor market outcomes that the individual may receive given the nonlinear effects of education on wages. The average differences in wages between these two counterfactual sceneries is defined as the Policy Relevant Treatment Effect (PRTE).

Additionally, I am also able to compute the returns between two different levels of education. I focus on two standard average treatment parameters: the average treatment effect of the schooling level s' with respect to the level of education s ($s' > s$) for a randomly chosen individual (ATE), and the average treatment effect on the treated (TT) for the schooling level s'. In the empirical

[19] I also estimate the model under alternative assumptions. The results are similar.

analysis s is set equal to 0, high school dropout, as the reference category, while $s' = 1, 2, 3$, i.e. HS graduate, college dropout and college graduate, respectively.

The difference between two final levels of education is given by

$$Y_{s'} - Y_0 = X(\gamma_{s'} - \gamma_0) + \varepsilon_{s'} - \varepsilon_0 = X(\gamma_{s'} - \gamma_0) + (\alpha_{s'} - \alpha_0)\theta + u_{s'} - u_0 \tag{4.12}$$

The Average Treatment Effect (ATE) is computed over the entire population as follows

$$ATE_{s,0} = \int\int \mathbf{E}[Y_{s'} - Y_0 | X = x, theta = \bar{\theta}] \mathrm{d}F_{X,\theta}(x, \bar{\theta}) \tag{4.13}$$

The Treatment Effect of the Treated (TT) is the average treatment effect associated with the level of schooling s' relative to the high school dropout level and conditional on the individuals whose final level of schooling is s'

$$TT_{s,0} = \int\int \mathbf{E}[Y_{s'} - Y_0 | X = x, theta = \bar{\theta}] \mathrm{d}F_{X,\theta|T_s=1}(x, \bar{\theta}) \tag{4.14}$$

I compute the Policy Relevant Treatment Effect (PRTE) (Heckman and Vytlacil (2001)) which measures the average return for those individuals induced to change their schooling decisions in response to a specific policy, for instance, a high school scholarship.[20] Note that the PACES scholarship only affects the first transition directly. Consider two possible counterfactual scenarios. In the first scenario the individuals receive the PACES scholarship, and the value function is given by $V^p_{0,11}$, while the final outcome is given by Y^p_1. In the second scenario, no students receives the scholarship, and the value function is defined by $V^p_{0,10}$, while the outcome is given by Y^p_0. Then, I want to estimate the average return for those individuals such that under the first scenario $V^p_{0,11} > 0$ (i.e. $D_{0,1}(V^p_{0,11}) = 1$ and $s > 0$), while under the second scenario $V^p_{0,10} < 0$ (i.e. $D_{0,1}(V^p_{0,10}) = 0$ and $s = 0$). Then the PRTE is the average of the difference of outcomes under different policies,

[20]Examples of specific policies evaluated with the PRTE method in the past include a fixed tuition change in the US (Heckman and Vytlacil (2001); Carneiro, Heckman, and Vytlacil (2011)), or a policy that expand the college system (Belskaya, Peter, and Posso (2014)).

$Y_1^p - Y_0^p$, for switchers

$$PRTE_{p_0,p_1} = \int\int \mathbf{E}[Y_1^p - Y_0^p | X = x, theta = \bar{\theta}] \mathrm{d}F_{X,\theta|(D_{0,1}(V_{0,1}^p 1)=1, D_{0,1}(V_{0,1}^p 0)=0)}(x, \bar{\theta}) \quad (4.15)$$

I also compute the PRTE for those induced to switch the schooling transition by the receipt of the PACES scholarship but conditional on the final level of school s for all $s = 1, 2, 3$.

CHAPTER 5

ESTIMATION STRATEGY

In this section I describe the estimation strategy that I follow in this paper. The estimation of the sequential discrete choice model with potential outcomes depends on the distributional assumptions imposed on the random components θ (i.e. $f(\theta)$), $v_{j,j'}$, and $u_{k,s}$, and the normalization on the loading factors $\alpha_{k,s}$ and $\alpha_{j,j'}$. My estimation strategy is implemented in two stages. In the first stage, I estimate the sequential discrete schooling choice model and the empirical distribution of θ (i.e. $f(\theta)$). In the second stage, using the estimations from the first stage, I estimate the model of potential outcomes.[1]

Define the c.d.f and p.d.f of θ as $F(\theta)$ and $f(\theta)$ respectively. Given assumption 1, and conditional on θ, and the set of observable variables, I am able to control for dynamic selection which implies that the schooling transitions are independent of each other. I am also able to account for essential heterogeneity, which in turns implies that the outcome is independent of the sequential choice model. Conditioning on θ, the likelihood has the form

$$\prod_{i=1}^{n} \Pr(\mathfrak{D}_i, Y_i | X_i, Z_i, \theta_i) = \prod_{i=1}^{n} \Pr(\mathfrak{D}_i | Z_i, \theta_i) \Pr(Y_i | X_i, \theta_i) \tag{5.1}$$

The likelihood function with θ integrated out is given by the following expression

$$L = \prod_{i=1}^{n} \int \Pr(\mathfrak{D}_i | Z_i, \theta_i) \Pr(Y_i | X_i, \theta_i) F_\theta \mathrm{d}\theta \tag{5.2}$$

[1]The main gain from estimating the complete model all at once is efficiency. Nonetheless, this comes with a cost since I will use the outcome system to estimate the distribution of θ. Since our exercise includes several outcomes, the estimated distribution of θ will change with each outcome. Under this circumstance, it could be the case that one outcome provides a stronger prediction for the distribution of θ which implies that I am getting tautologically good fits between Y and θ which makes it impossible to interpret of the effects of the unobserved endowments. To estimate the standard errors I use bootstrap methods.

In the first stage, I estimate (jointly) $\Pr(\mathfrak{D}_i|Z_i, \theta_i)$, and the empirical distribution of θ, $\hat{F}_\theta$.

Assumption 2: Assume θ follows a mixture of normal distributions. In particular, suppose

$$\theta i = \sum j = 1^k d_{ij}(\mu_j + \sigma_j \chi_{ij}) \tag{5.3}$$

Where $\chi_{ij} \sim \mathcal{N}(0,1)$ for $j = 1, \cdots, k$, and μ_j, σ_j represents the mean and standard deviation of each of the normal distributions respectively. The random vectors $\Lambda = (d_{i1}, \cdots, d_{ik})$ are i.i.d. with multinomial probability distribution of parameters $p_j = \Pr(d_{ij} = 1)$ for $j = 1, \cdots, k$, where $\sum j = 1^k p_j = 1$. Thus, the c.d.f. of a mixture of k normal random variables is defined by $F_\theta = \sum j = 1^k p_j \Phi(\frac{\chi_j - \mu_j}{\sigma_j})$, and the p.d.f. is $f_\theta = \sum j = 1^k p_j \frac{1}{\sigma_j} \phi(\frac{\chi_j - \mu_j}{\sigma_j})$.[2].

Given Assumption 1, the likelihood function for the first stage is given by

$$\prod_{i=1}^{n} \int \Pr(\mathfrak{D}_i|Z_i, \theta = \theta_m) \mathrm{d}F_\theta(\theta_m) \tag{5.4}$$

Similarly, in the second stage, I use first stage estimates, $\hat{F}_\theta$ and $\hat{\Pr}(\mathfrak{D}_i|Z_i, \theta = \theta_m)$, to express the likelihood function as:

$$\prod_{i=1}^{n} \int \hat{\Pr}(\mathfrak{D}_i|Z_i, \theta = \theta_m) f(Y_i|X_i, \theta = \theta_m) \mathrm{d}\hat{F}_\theta(\theta_m) \tag{5.5}$$

Each stage is estimated using maximum-likelihood. The integral in both likelihood functions cannot be solved analytically. I solve the integrals numerically using adaptive (Bayesian) quadrature methods (see Naylor (1982), Skrondal and Rabe-Hesketh (2004), Rabe-Hesketh, Skrondal, and Pickles (2005)).[3] I also estimate the model with an alternative set of assumptions. In the appendix D, I show the likelihood function for the case where $F(\theta)$ is a discrete distribution with mass points $(P_m(\theta), \theta_m)_{m=1}^{M}$, where $P_m(\theta_m)$ is the probability associate with the mass point θ_m,

[2]where $\mathbf{E}[\theta] = \sum j = 1^k p_j \mu_j$, and $\mathbf{E}[\theta^2] = \sum j = 1^k p_j(\mu_j^2 + \sigma_j^2)$

[3]The integration is iterative. I start with a guess given by $\mu_j^0 = 0$, $\sigma_j^0 = 1$. Using the posterior functions in Naylor (1982), the posterior means ($\hat{\mu}_j$), standard deviations ($\hat{\sigma}_j$), and the conditional likelihood function $L(|\theta)$ are updated in the kth iteration. This sequence is repeated until convergence. I found that this strategy improves upon standard Gauss-Hermite quadrature method. Results with standard Gauss-Hermite quadrature method are also available.

and $\sum mP_m(\theta m) = 1$. In particular, I follow the approach of Mroz and Guilkey (2014) and Mroz (1999). I also estimate the model assuming normality in all random variables of the model, in particular $\theta \sim \mathcal{N}(0, 1)$, following the additional normalization imposed by Aakvik et al. (1999) and Aakvik et al. (2005). The results are robust under Mroz and Guilkey (2014) approach. Nonetheless, my data rejects the normality assumption.

CHAPTER 6

DATA

In this section I describe the PACES program and the various sources of administrative data that I use for the estimation and identification of my sequential model with potential outcomes. In the appendix A, I provide additional details of the Colombias education system.

6.1 PACES background

The PACES program was initially introduced in 1992.[1] This program was created to increase the coverage rate of private secondary school, and reduce the desertion rate of students who were attending public primary schools. The program targeted disadvantage children, in particular, children who by the time of the program were residing in neighborhoods classified in the two lowest socioeconomic strata. The program was renewed annually conditional on passing the previous grade. The PACES program provided over $125,000$ vouchers in about $2,000$ private schools located in Colombias largest cities. Only Bogota, the capital and largest city in Colombia, absorbed thirteen percent of the scholarships. Initially, the program was funded by the Colombias government and the World Bank.[2] The scholarships were worth $US\$190$ in 1998 or about 1.33 times the monthly minimum wage in Colombia in 1998 (see Angrist et al. (2006)).

The key characteristic of this program was that in municipalities where the demand for the program exceeded the supply, the scholarships were assigned using a lottery. In particular, in 1994 in Bogota, the selection of the beneficiaries of the program for the next year was awarded randomly. Once the scholarship was assigned randomly in the first year of secondary education, individuals were able to renew it annually contingent upon passing the previous year.[3] My baseline sample

[1] Acronym in Spanish for *Plan de Ampliación de Cobertura de la Educacin Secundaria* (PACES).

[2] Colombias government ($US\$60$ millions) and World Bank ($US\$90$ million).

[3] Notice that the scholarship is only random in the first period. Renewal depends of individual choices.

covers the PACES program applicants in Bogota in 1994. By 1995, in Bogota, the scholarship covered most tuition cost for those students who were enrolled in the first year of secondary education. Nonetheless, the government did not adjust the real value of the scholarship and by 1998 it only covered about 56 percent of the tuition (see Angrist et al. (2002)).

6.2 Main sample

My baseline information comes from the PACES program in Bogota, which assigned scholarships randomly to those individuals who applied in 1994 to enter the first year of secondary education in a private school in 1995. The scholarship lasted until the end of secondary education for those students who never failed a school year. The PACES data includes names and identification for 4044 individuals. It also contains additional socioeconomic variables like date of birth, and gender. Finally, it includes a variable that indicates whether the individual won the lottery for the scholarship. This data has been previously used by Angrist et al. (2002), Angrist et al. (2006), Bettinger et al. (2010), and Bettinger et al. (2014).

I use the PACES data in two ways. First, as in early papers that evaluate the PACES program, I use the basic individual characteristics from the PACES application form, that includes age, gender and whether the applicant had a phone number at the time of application, as a vector of baseline controls (see Angrist et al. (2002), Angrist et al. (2006), Bettinger et al. (2010), and Bettinger et al. (2014)). [4] Also, because Bettinger et al. (2010) found that applicants to schools with a vocational curriculum, the lottery winners did not attend schools with observably more desirable peers, I include a dummy variable for vocational school.[5]

Second, since the scholarship was assigned randomly between participants, I use the scholarship as an exclusion restriction that provides free variation in the first schooling transition of my

[4] As in other papers that study PACES program, my sample is restricted to applicants with valid adult identification number (3926) that have complete application controls (3903) and have nonmissing information in other controls (3805). Baseline controls include age, female and whether the applicant had a phone number at the time of scholarship application. The definition of all variables is described in Appendix A, Table E.1.

[5] Here I assume that the vocational curriculum is characteristic of the school which may be a good control. The vocational curriculum is an initial condition associated with the schools that participated in the PACES program. Nonetheless, one can assume that the vocational curriculum is a choice that is taking simultaneously with the enrollment in the initial period, and then it may be necessary to model such decision. Such case is out of scope of this paper. In any event, there is no evidence that the scholarship affected the share of individuals in vocational curriculum.

model. Note that the randomization balanced the bias in the sample of winners with the bias in the sample of losers (Heckman (1996)). Also, the lottery allows me to assume independence of the scholarship with respect to other regressors and other important individual and family characteristics not observed at the first transition.

My data also includes administrative records on individual-level education transitions, and labor and social outcomes observed after individuals complete their sequence of schooling decisions.[6] The extent of Colombias administrative records provide a unique opportunity to track PACES applicants across all schooling transitions and adult outcomes with little to no attrition in the data. In order to estimate my sequential schooling model I need data associated with schooling transitions from secondary school until graduation from college for each individual. To identify students who graduated from high school (i.e. the first transition, $D_{0,1}$), PACES data was merged with the administrative records of Colombia's centralized college entrance examinations, ICFES exam[7], administered just before graduation from high school in Colombia. These records allow me to observe the complete population of students in the last grade of secondary education in Colombia.

Secondary education in Colombia lasts six years. Individuals who were promoted on schedule should have registered to take the ICFES exam in the 2000 school year. Because some individuals may also have repeated grades, it is likely that some of them were registered to take the ICFES exam in subsequent years. I merged ICFES data between 2000 until 2008, though most of the individuals in my sample took the exam in 2000 or 2001. An indicator variable for matching with this database is the proxy that I use for high-school graduation (i.e. $D_{0,1} = 1$).[8] The ICFES exam evaluates the students' proficiency in the subjects of math, language (Spanish), physics, biology,

[6]Labor outcomes include earnings, earnings in the formal sector in 2010, number of days worked in formal sector in 2010, and employment status. Also, I have information associated with health insurance and pensions, which is a proxy for work in the informal sector. Social outcomes includes poverty status, and Sisben index (poverty index).

[7]I used applicants names, dates of birth and adult identification numbers from the national registrar to complete the matches with the all administrative datasets.

[8]According with WB (1999), and Angrist et al. (2006) ICFES registration status is an excellent proxy for high-school graduation because 95 percent of graduating seniors take the ICFES exam.

chemistry, geography, philosophy, and history. This data also includes a rich set of socioeconomic variables like family income, parents education, number of siblings and other characteristics of the students, schools, and students peers. I use information at the school level as an additional source of variation.

In the second and third transition, the individuals choose whether to enroll in college or not conditional on being a high school graduate, and conditional on being enrolled in college, whether to graduate from college or not. To identify those individuals who decided to enroll in college and graduated from college, the PACES data was merged with the administrative records from the Colombias National System for Prevention of Desertion from College Institutions, SPADIES.[9] This higher education dataset is an individual-level panel dataset that tracks up to 99 percent of college students from their first year to their degree receipt starting in 1998 up to 2012. An indicator variable for matching with this database is the proxy that I use for college enrollment (i.e. $D_{1,2} = 1$). Similarly, if an individual is reported in SPADIES as having graduated from college, then I use this information as an indicator that the individual completed the third transition (i.e. $D_{2,3} = 1$). This data also includes information on the name and national code of the university, the duration of the program (i.e. 2-year College or 4-year College), the major each individual chose, and if the individual dropped out or graduated from college for every specific year of the panel. Similar to the ICFES data, this data also includes a rich set of socioeconomic variables.

In order to obtain additional information on the colleges that individuals in my sample attended, I matched the codes and names of the college with the National System of Information for College Education (i.e. SNIES).[10] This dataset contains variables like the number of academic programs the colleges have (a proxy for the size of the university), whether the colleges have an accreditation of high quality or not, the cost of the respective program, and a dummy for private/public schools.[11] Finally, using the codes and names of the major each individual chose, I obtain the average earnings

[9]Acronym in Spanish for *Prevention and Analysis System for Desertion in Higher Education Institutions* (SPADIES). This Higher Education Database is similar to the National Student Clearinghouse in the United States.

[10]Spanish acronyms for *Sistema Nacional de Informacion para la Educacion Superior* (SNIES).

[11]Unfortunately, the cost of the program is not available for most of the 2-year college programs.

of individuals who graduate from those majors in 2001. I use information at the college or program level as additional sources of variation.

Information on labor market outcomes comes from Colombias Social Protection Ministrys Integrated Information System of Social Protection (SISPRO) for the years 2008 to 2012.[12] SISPRO is an individual-level panel dataset containing information on health insurance, social security payments, retirement status, days worked, and monthly earnings. Nonetheless, this data does not include information for those workers employed in the informal sector. The informal sector includes all workers that do not contribute to the Colombias social protection system, so informality here refers to an employee who does not receive health and pension benefits through his job.

I supplement the labor market data with information available from the census of the poorest 70 percent of the Colombias population extensively collected in 2010 (Called SISBEN 3). The main objective of this census is to build a quality of life index (i.e. SISBEN score) and classify the Colombian households at different levels of poverty (i.e. SISBEN levels) in order to target social public programs like health insurance, housing, and others conditional cash transfer programs (e.g. Familias en Accion).[13] This data allows me to measure both formal and informal employment and earnings. Also, this data includes information on poverty status, SISBEN score, adolescence pregnancy, family formation, and others. Since SISBEN data is only available in 2010, I use 2010 as my reference year to measure my outcomes.

Table 6.1 categorizes the variables used in the sequential discrete choice model with potential outcomes. Table E.1, in appendix A, details how each variable is constructed. The descriptive statistics in Table 6.2 are reported separately for each schooling transition. Female and younger individuals transitioned to higher levels of education. Interestingly, when in second transition, family human capital and income were low: their mother had mostly a primary or secondary

[12]In Spanish *Sistema Integral de Informacion de la Proteccion Social* (SISPRO). The initial observation was constrained to 2008, since July 2008 SISPRO increased their coverage becoming the national census of social security contributions.

[13]The SISBEN score is an index that goes from 0 to 100, where 0 represents the poorest households. If the score is less than 47.99 then the household is categorized as SISBEN level 1. If the score is between 47.99 and 54.87, then the household is categorized as Sisben level 2. Households categorized as SISBEN 1 and 2 have access to the Colombia subsidized health regime.

degree, 60% owned a house, and the family income for more than half of the sample was less than twice the minimum wage. Those individuals who were able to enroll in college showed higher levels of family human capital and income. Table 6.3 reports the descriptive statistics for the final outcomes for 2010. By this year, 40 percent of the individuals are a high school dropout, while 40 percent are high school graduate. 20 percent of the individuals were able to attend college, although only 20 percent of those who enrolled in college were able to graduate (4 percent of the total sample). Most of the individuals are no longer poor, although 27 percent of the individuals are still in the first two levels of SISBEN.

Table 6.1: Observable variables in sequential schooling model and outcomes

Variable/transition-outcome	**Dropout HS - Graduate HS**	**Graduate HS - Enroll College**	**Enroll College - Graduate College**	**Outcome**
Scholarship (dummy)	x			
Female (age)	x	x	x	x
Age	x	x	x	x
Vocational HS	x	x	x	
missing Vocational HS	x	x	x	
Had Phone (dummy)	x	x	x	
Family Income [1,2) MW		x	x	
Family Income [2,3) MW		x	x	
Family Income [3,+) MW		x	x	
Family Owned a House (dummy)		x	x	
Mother education, Secondary		x	x	
Mother education, College		x	x	
Number of Siblings		x	x	
Had a sibling in College		x	x	
Average Score High School		x	x	
Class Size High School		x	x	
Accredited Program (High Quality)			x	

Notes: The first three columns show the variables included in the three transitions of the sequential schooling model. The only exogenous variables available for the first transition are female, age, vocational school and had phone. The last column shows the variables included in the outcome equations. All equations contain the baseline controls included in previous papers, such that female, age and phone. Phone is not relevant for any outcome. The definition of all variables is described in appendix A, Table E.1.

Table 6.2: Sample Statistics: Sequential schooling transitions

Variable/Transition	**Dropout HS - Graduate HS**		**Graduate HS - Enroll College**		**Enroll College - Graduate College**	
Scholarship (dummy)	0.587	(0.493)		-		-
Female (age)	0.512	(0.500)	0.536	(0.499)	0.544	(0.498)
Age	12.691	(1.299)	12.268	(1.138)	11.914	(1.027)
Vocational HS	0.411	(0.492)	0.446	(0.497)	0.480	(0.500)
missing Vocational HS	0.067	(0.250)	0.048	(0.214)	0.040	(0.197)
Had Phone (dummy)	0.884	(0.321)	0.891	(0.311)	0.910	(0.286)
Family Income [1,2) MW		-	0.562	(0.496)	0.544	(0.498)
Family Income [2,3) MW		-	0.182	(0.386)	0.215	(0.411)
Family Income [3,+) MW		-	0.090	(0.287)	0.151	(0.359)
Family Owned a House (dummy)		-	0.594	(0.491)	0.715	(0.452)
Mother's education, Secondary		-	0.330	(0.470)	0.394	(0.489)
Mother's education, College		-	0.033	(0.178)	0.060	(0.238)
Number of Siblings		-	2.538	(1.762)	2.449	(1.616)
Had a sibling in College		-	0.344	(0.475)	0.393	(0.489)
Average Score High School		-	0.361	(0.072)	0.372	(0.066)
Class Size High School		-	0.163	(0.132)	0.177	(0.138)
Accredited Program (High Quality)		-		-	0.110	(0.313)
Number of observations		3805		2288		766

Notes: Descriptive statistics are provided for the baseline sample. Sample is restricted to applicants with valid adult identification number (3926) that have complete application controls (3903) and have nonmissing information in other controls (3805). Baseline controls include age, female and whether the applicant had a phone number at the time of scholarship application. The definition of all variables is described in appendix A, Table E.1. Standard deviations are in parentheses.

Table 6.3: Sample Statistics: Outcomes

Variable	***N***	**Mean**	**SD**
HS Dropout, %	3805	0.399	(0.490)
HS Graduate, %	3805	0.400	(0.490)
Some College, %	3805	0.161	(0.367)
College Graduate, %	3805	0.040	(0.197)
SISBEN score (Poverty Index)	3805	74.18	27.47
Sisben Nivel 1, %	3805	0.207	(0.405)
Sisben Nivel 2, %	3805	0.060	(0.237)
Unemployed, %	3805	0.287	(0.452)
Employee, %	3805	0.713	(0.452)
Employee in Formal Market, %	3805	0.579	(0.494)
Log Wages	2706	13.309	(0.536)
Log Formal Wages	2202	13.438	(0.371)
Days Worked Formal Market	2706	230.98	(145.13)

Notes: Descriptive statistics are provided for the baseline sample. Sample is restricted to applicants with valid adult identification number (3926) that have complete application controls (3903) and have nonmissing information in other controls (3805). Standard deviations are in parentheses. The definition of all variables is described in appendix A, Table E.1. All variables are measured by 2010. Wages are measured in 2010 Colombian Pesos. The SISBEN score is an index that goes from 0 to 100, where 0 represents the poorest households. If the score is less than 47.99 then the household is categorized as SISBEN level 1. If the score is between 47.99 and 54.87, then the household is categorized as Sisben level 2. Households categorized as SISBEN 1 and 2 have access to the Colombia subsidized health regime. Informal sector includes all workers that do not contribute to the Colombias social protection system.

CHAPTER 7

RESULTS

I now formally present the effects of the scholarship program PACES on adult labor outcomes through the accumulation of human capital by estimating the sequential schooling model with potential outcomes. I start by showing the full set of estimated parameters and the fit of the model (section 7.1). Then I will report the effects of the different levels of education on adult outcomes using the standard average treatment effects parameters, ATE and TT (section 7.2). Finally, I will show the causal effects of the PACES program using the policy relevant treatment parameters, PRTE (section 7.3). I also show below that my results are a significant improvement with respect to a nave model that does not control for dynamic selection and essential heterogeneity. My results are also robust to different assumptions on the structure of the errors and different specifications for identification. The main focus in section 7.1 to 7.3 is on the effects on wages. Section 7.4 present the results on other adult outcomes.

7.1 Estimated Parameters

In this section, I report the maximum likelihood estimates for my model. The likelihood function is given in Equation 4.1. I begin by showing in table 7.1 the estimated coefficients of the sequential schooling model given in Equation (1). Table 7.1 shows the results using the lottery of the program as an exclusion restriction, while the second and third transition share the same set of explanatory variables. Table E.2 to E.4 in the appendix present the results under alternative specifications, including the possibility of additional exclusion restrictions for the last transition.[1]

[1]Table E.2 presents the results when I include an additional instrument for the last transition. Table E.3 shows the results under additional distributional assumptions for θ (see appendix E). In particular, I follow the approach of Mroz and Guilkey (2014), and Mroz (1999) which extends the method of Heckman and Singer (1984). The model is estimated with 3 types. When I include 4 and 5 types the likelihood function does no improve, although the results are similar. Also, Rosenzweig and Wolpin (1980) argues that there exists a negative correlation between child quality related outcomes and family size. Similar evidence has been found by Black, Devereux, and Salvanes (2005) and Juhn and Zuppann (2015). Table E.4 shows the results when the sibling variables are excluded. No significant change has

Overall, these additional specifications are not very different from the baseline specification.[2]

In this exercise, I exploit three characteristics of the data. First, all individuals that participated in PACES program finished elementary school, which implies that the initial condition is defined exogenously, i.e. $D_{-\infty,0} = 1$ for all individuals. Second, the randomization of the program equates the distribution of covariates in the treated and control groups for the first transition. Previous literature on PACES programs has shown that the randomization of the PACES program was effective and the groups are balanced, although it is usually recommended to control for some basic demographic variables. Also, it has been show that PACES programs has a significant effect on graduation from high school. These two characteristics of the data and the additional distributional assumptions allow me two satisfy the condition for the identification of all elements in the first transition (see proposition 1 and 2, condition (iv) in appendix A). Third, I exploit the fact that the data arrives sequentially. For the first transition, the only information available is the baseline characteristics. Nonetheless, for those individuals who arrive to the second and third transition, new information is available. Notice that for these individuals the balance property of the first transition is no longer available, and additional controls are needed. Given that the first transition is already identified, the specifications of the second and third transition benefit from the new information to satisfy the sufficient conditions for the identification (see proposition 1 and 2, condition (iv) in appendix A).

I find that the exclusion restriction associated with the PACES lottery is a strong predictor of the first schooling transition at the 1% level of significance, with a marginal effect of 5.2 percent which is consistent with the literature on the PACES program. Figure E.1, in the appendix, shows the densities of $\Pr(D_{0,1} = 1|scholarship = 1)$ and $\Pr(D_{0,1} = 1|scholarship = 0)$, conditional on observable and unobservable characteristics. Clearly, the PACES scholarship serves as an important cost shifter for secondary schooling decisions. Additionally, the probability of

found. We also include the PACES lottery in the second transition, but it is not statistically significant. Results do not change.

[2]I compare coefficient across models and test if the difference between them is statistically significant. In general I found no statistical difference.

Table 7.1: Estimates for the sequential schooling model

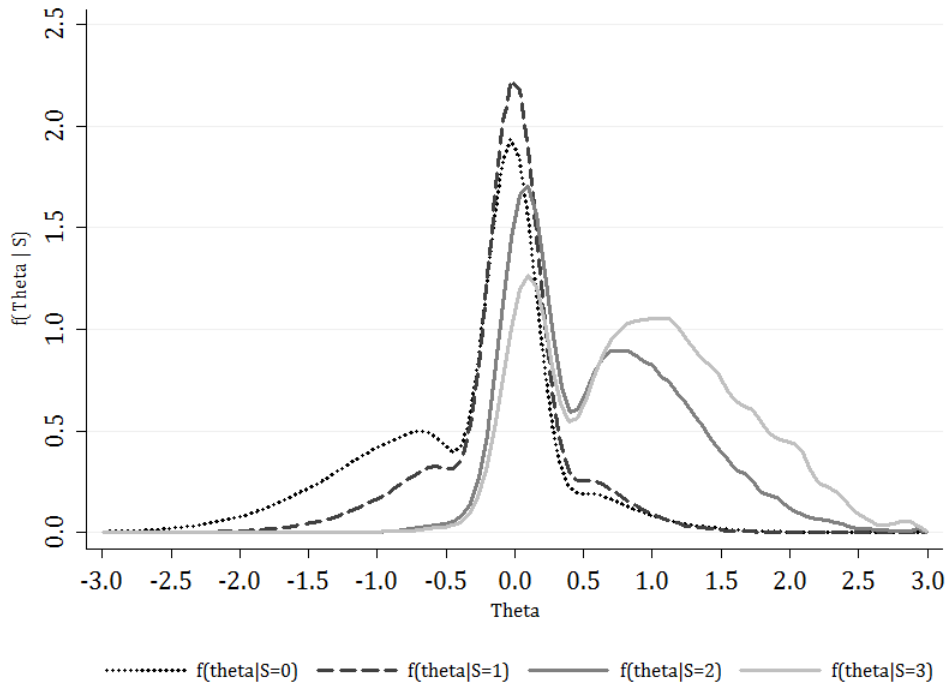
Variable/Transition	Dropout HS - Graduate HS		Graduate HS - Enroll College		Enroll College - Graduate College	
Scholarship (dummy)	0.174***	(0.050)	-	-	-	-
Female (dummy)	0.145***	(0.050)	0.162	(0.100)	0.271**	(0.136)
Age	-0.469***	(0.044)	-0.465**	(0.184)	-0.273*	(0.163)
Vocational HS	0.085*	(0.051)	0.092	(0.089)	0.056	(0.122)
missing Vocational HS	-0.354***	(0.103)	-0.155	(0.224)	0.020	(0.318)
Had Phone (dummy)	0.071	(0.075)	0.198	(0.145)	0.164	(0.220)
Family Income [1,2) MW	-	-	0.639***	(0.236)	0.143	(0.279)
Family Income [2,3) MW	-	-	0.745***	(0.271)	0.451	(0.328)
Family Income [3,+) MW	-	-	1.195***	(0.354)	0.172	(0.348)
Family Owned a House (dummy)	-	-	0.735***	(0.222)	0.608**	(0.277)
Mother education, Secondary	-	-	0.236**	(0.103)	0.155	(0.138)
Mother education, College	-	-	0.674**	(0.271)	0.647**	(0.293)
Number of Siblings	-	-	-0.027	(0.028)	-0.049	(0.042)
Had a sibling in College	-	-	0.179*	(0.097)	0.022	(0.129)
Average Score High School	-	-	2.669***	(0.916)	3.277**	(1.298)
Class Size	-	-	0.748**	(0.358)	-0.116	(0.453)
Constant	6.019**	(0.567)	2.270*	(1.339)	-0.209	(1.037)
Alpha	1.000	-	2.358**	(1.152)	0.772	(0.724)
Pr1				0.67		
Pr2				0.33		
Mean (Theta)				0.06		
SD(Theta)				0.522***		
ll				-3875.17		
LR test (p-value)				0.01		

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Sample is restricted to applicants with valid adult identification number (3926) that have complete application controls (3903) and have nonmissing information in other controls (3805). This table shows the maximum likelihood estimates of the sequential schooling model with potential outcomes for the first stage. Standard errors are in parentheses. The computation of the LR test includes two steps. The first step is to get the likelihood value for the first stage, i.e. the sequential schooling model. In the second step I constraint the loading factor of the sequential schooling model to be zero (i.e. no dynamic selection) and re-estimate the complete set of parameters to get a new likelihood value. Using these two likelihood values the LR test is constructed.

graduating from high school is significantly higher among females, and younger individuals. Vocational school is statistical significant at 10% level. Similar to the first transition, the probability of enrolling in college conditional on graduating from high school and the probability of graduating from college conditional on enrollment in college (second and third transitions respectively) are significantly higher among females, and younger agents. As expected, individuals with higher family income, mothers secondary or college education and those whose family owned a house are more likely to enroll in college. Family income is not relevant for the last transition, although mothers college education and parents home ownership are still strong predictors. The average score for the high school, which is a proxy for the quality of the high school, is a strong predictor of both the probability of enrollment in college and the probability of graduating from college conditional on being enrolled, while the size of the high school is relevant only for the former probability. It is important to note that the unobserved factor θ is a strong predictor of the second transition.

I also estimate the schooling choice model (first stage) with the restriction that all loading factors are equal to zero. This constraint implies that my model is free of dynamic selection bias and there is no selection on gains. A simple likelihood-ratio test rejects the null model (p-value=0.01), which implies that the model with unobserved heterogeneity fits the data better. Table E.5 in the appendix shows the Goodness of Fit for the sequential schooling model. Figure 7.1 shows evidence of selection on the different final levels of education based on the unobserved factor θ. Each function in figure 7.1 represents the density of θ conditional on one of the four final levels of education. I find that individuals in the first two levels of education (i.e. HS dropouts and graduates) have lower levels of θ than those individuals who select on higher levels of education (i.e. college dropouts or college graduates). This result is consistent with the possibility of dynamic selection in school attainment since lower ability students select themselves into the most basic levels of education, while higher ability students continue in the schooling system until college education. In general, this figure supports the necessity of controlling for unobserved factors that may be driving the schooling choices. Table 7.2 shows the estimated coefficients for the potential outcomes given in Equation 4.2, for two outcomes: the log of the monthly wage rate in 2010

Figure 7.1: Densities of θ by final level of education



(from now on wages) which includes both individuals who work in formal and informal labor market and the log of monthly wage rate in 2010 in the formal market only (from now on formal wages). The unobserved factor θ has strong statistically significant effects on both wages and formal wages. This supports the fact that the selection on gains may be important in my setup. Interestingly, although the female variable has a positive effect on all schooling transitions, it has a negative effect on wages and formal wages for all final levels of education except for college graduates. Similarly, age has a negative effect on wages for all final levels of schooling but college graduates. When I estimate the full likelihood function (see Equation 5.5) with the restriction that all factor loadings are equal to zero, the value of the likelihood function is significantly lower than the likelihood value obtained with my model, in particular a likelihood-ratio test rejects the null model at 1 percent level of significance. Table E.6 in the appendix shows the fit of the model for the average and standard deviation of the outcomes, while figures E.2 and E.3 shows the fit of the model for the different percentiles. Overall, the goodness of fit is significantly better when I do not constrain the loading factor to be zero.

Table 7.2: Estimate coefficients for potential outcomes: wages and formal wages

Variable/Transition	HS Dropout	HS Graduate	Some College	College Graduate
		Panel A. Log Wages		
Female (dummy)	-0.157***	-0.120***	-0.072***	-0.101
Age	-0.032***	-0.040***	-0.070***	-0.009
Constant	13.716***	13.877***	14.355***	13.975***
Alpha	0.788***	0.713***	0.735***	0.624***
ll			-5358	
ll0			-5768	
LR-test (pvalue)			0.00	
		Panel B. Log Formal Wages		
Female (dummy)	-0.078***	-0.094***	-0.065**	-0.067
Age	-0.010	-0.027***	-0.059***	-0.030
Constant	13.452***	13.728***	14.244***	14.226***
Alpha	0.459***	0.481***	0.644***	0.589***
ll			-4353	
ll0			-4517	
LR-test (pvalue)			0.00	

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. This table shows the maximum likelihood estimates of the sequential schooling model with potential outcomes for the second stage. To account for the fact that $\hat{F}_\theta$ and $\Pr(\mathfrak{D}_i|Z_i, \theta = \theta_m)$ are estimated in the first stage, the statistical significance of the parameters is obtained using bootstrap methods, in particular, I implement the percentile-t confidence interval method (see Horowitz (2001)). The results are robust to other bootstrapping techniques. The panel A shows the results when the outcome of interest is the log of the monthly wage rate in 2010 (wages) which include both individuals who work in formal and informal labor market. The panel B shows the results when the outcome of interest is the log of monthly wage rate in 2010 in formal market only (formal wages).The computation of the LR test includes two steps. The first step is to get the likelihood value for the likelihood function including both first and second stage. In the second step I constraint the loading factor of the sequential schooling model and the potential outcomes to be zero (i.e. no dynamic selection, no essential heterogeneity) and re-estimate the complete set of parameters to get a new likelihood value. Using these two likelihood values the LR test is constructed.

7.2 Average Treatment Effects of Education on Wages

In this section, I report the simulated treatment effects defined in chapter **??**.[3] Given the identification of the marginal distribution of θ, $v_{j,j'}$ and $u_{k,s}$, I randomly drew values from these distributions to simulate outcomes, and then I computed the difference between two final levels of education, and finally determine the treatment effects given in equations 4.13, and 4.14. Table 7.3 presents estimates of the standard treatment effects parameters such as the average treatment effect in the population and the average treatment effect on the treated, associated with each final level of education s, relative to the high school dropout level ($ATE_{s,0}$, $TT_{s,0}$ respectively). To compute such treatment parameters, I integrate out over the distribution of returns to schooling given by equation 4.12 unconditional on having had the scholarship. The next section focuses on the effects for those individuals that the scholarship was relevant.

In general, the returns to schooling are high, and there is strong evidence in favor of the non-linear relationship between schooling and wages. For instance, the return per additional year of education of an individual who graduates from college is 9.6 percent, while for someone who graduates from high school is only 1.5 percent.[4] Applied literature in economics has shown that when this nonlinearity is ignored, the returns of education are over-estimated (see Belzil and Hansen (2007), Belzil (2008), Belzil (2007)).

The first column of Table 7.3 is the observed difference in the data between two levels of education, i.e. $\mathbf{E}[Y|S = s] - \mathbf{E}[Y|S = 0]$. The other two columns show the $ATE_{s,0}$ and $TT_{s,0}$. This Table shows the results for the log of monthly wage rate in 2010 and the log of the monthly wage rate in 2010 in formal market. For a randomly chosen individual, the average treatment effect on wages of being a high school graduate with respect to being a high school dropout is 7.4 percent (i.e. $ATE_{1,0}$), while the treatment effect on the treated is 8.2 percent (i.e. $TT_{1,0}$). The

[3]Sections 7.2 to 7.4 use the estimated parameters of my baseline model. I also randomly draw regressors (Xs) from the data of each individual, and the estimated distributions of θ, $v_{j,j'}$, and $u_{k,s}$ to simulate the outcomes. Our simulated data has around $1,100,000$ observations.

[4]To compute this numbers I use the treatment effect on the treated (TT) on table 7.3 for wages. To annualize the effect, I divide the TT in the table by 6 in the case of high school and by 5 in the case of college graduate, since in Colombia high school and college usually last 6 and 5 years respectively.

Table 7.3: Treatment Effects for final educational levels: High School dropout as a comparison group

Outcome/Treatment	Observed	ATE	TT
Log wages			
HS Graduate - HS Dropout	0.147	0.074***	0.082***
		(0.021)	(0.024)
Some College - HS Dropout	0.377	0.198***	0.226***
		(0.036)	(0.049)
College Graduate - HS Dropout	0.693	0.578***	0.478***
		(0.072)	(0.242)
Log Formal wages			
HS Graduate - HS Dropout	0.070	0.055***	0.060***
		(0.015)	(0.017)
Some College - HS Dropout	0.256	0.177***	0.284***
		(0.031)	(0.094)
College Graduate - HS Dropout	0.536	0.527***	0.617***
		(0.066)	(0.263)

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Standard errors are bootstrapped (200 replications). Each row compares the outcomes from a particular final level of education s with the high school dropout level. The first panel shows the results when the outcome of interest is the log of the monthly wage rate in 2010 (wages) which include both individuals who work in formal and informal labor market. The second panel shows the results when the outcome of interest is the log of monthly wage rate in 2010 in formal market only (formal wages). The first column is the observed difference in the data between two levels of education, i.e. $\mathbf{E}[Y|S=s]-\mathbf{E}[Y|S=0]$. The other two columns show the $ATE_{s,0}$, the average treatment effect obtained from the comparison of the outcomes associated with the schooling level $S=s$ relative to the schooling level $S=0$ for a random person in the population (see equation 4.13), and $TT_{s,0}$, the average treatment on the treated effect obtained from the comparison of the outcomes associated with the schooling level $S=s$ relative to the schooling level $S=0$ but only for those individuals which the final level of education is $S=s$ (see equation 4.14).

$ATE_{2,0}$ of some college is 19.8 percent, while the $TT_{2,0}$ is 22.6 percent. The effects associated with college graduation are significantly higher ($ATE_{3,0} = 57.8$ and $TT_{3,0} = 47.8$). The causal effect of education explains 50.5 percent of the overall difference in wages between high school graduates with respect to being a high school dropout.[5] Similarly, the causal effect of education explains 52.4 percent and of the overall difference on wages between college dropouts with respect to high school dropouts and 83.3% in the case of college graduates.

The average returns associated with formal wages follow the same trend as wages (see Table 7.3), although the treatment on the treated for some college and college graduate are significantly higher for formal wages with respect to wages (28.4 and 61.7 percent respectively). It is important to note that the causal effect of education explains a higher percentage of the observed differences between levels of education with respect to the model with wages (78.0%, 69.3% and 98.3%, respectively).

The results are very close when I use the model with an additional exclusion restriction for the last transition (see Tables E.2 and E.7 in the appendix). In particular, I include in the last transition, from being enrolled in college to graduate from college, an indicator variable for whether the program the individual is enrolled in has an accreditation of high quality or not. Additionally, Table E.7 in the appendix shows that imposing the restriction that the model is free of dynamic selection bias and there is not selection on gains results in important changes in the ATE and TT and overestimate the causal effect of education on wages. It is important to note that such nave model is equivalent to estimate a simple OLS of the effect of level of education on wages.[6] In columns 5 to 7, Table E.7 also compares the treatment on the treated parameter in three cases: (1) my baseline model (i.e. $TT_{s,0}$), (2) imposing the restriction that the model is free of dynamic selection bias and there is no selection on gains, defined as $TT^2_{s,0}$, and (3) setting the factor loadings of the outcome equation equal to zero, defined as $TT^3_{s,0}$ (i.e. allowing the possibility of selection

[5]In this cases, I compute $(ATE_{1,0}/(\mathbf{E}[Y|S=1] - \mathbf{E}[Y|S=0]) * 100 = 50.5\%$.

[6]To estimate the nave model that assumes both no dynamic selection bias and no selection on gains, I re-estimate the complete set of parameters (of both schooling model and outcome equation) and perform the simulation of the ATE and TT. In particular, the parameters γ_s for $s = 0, 1, 2, 3$ are biased under the omission of the unobserved factor, θ, given the effects of θ on both schooling and wages (see Cameron and Heckman (1998)).

bias, but assuming not selection on gains). Notice that

$$\begin{cases} TT_{s,0} = x * (\gamma_{s'} - \gamma_0) + (\alpha_{s'} - \alpha_0) * \mathbf{E}[\theta | S = s, x] \\ TT^2_{s,0} = x * (\gamma^2_{s'} - \gamma^2_0) \\ TT^3_{s,0} = x * (\gamma_{s'} - \gamma_0) + 0 * \mathbf{E}[\theta | S = s, x] \end{cases} \tag{7.1}$$

To understand the effect of unobservables on the estimation of the TT parameter, I compute the differences in the TT for the cases (1) and (2): $TT_{s,0} - TT^2_{s,0}$. Notice that such difference can be decomposed as follows:

$$TT_{s,0} - TT^2_{s,0} = (TT_{s,0} - TT^3_{s,0}) + (TT^3_{s,0} - TT^2_{s,0}) \tag{7.2}$$

where

$$TT_{s,0} - TT^3_{s,0} = (\alpha_{s'} - \alpha_0) * \mathbf{E}[\theta | S = s, x] \tag{7.3}$$

and

$$TT^3_{s,0} - TT^2_{s,0} = x * [(\gamma_s - \gamma^2_s) + (\gamma_0 - \gamma^2_0)] \tag{7.4}$$

The first component $(TT_{s,0} - TT^3_{s,0})$ shows how much the differences between the TT effects are explained by the unobservable factor θ directly, while the second component $(TT^3_{s,0} - TT^2_{s,0})$ shows how much the difference is associated with the bias in the coefficients of the observable characteristics associated with the omission of the unobservable factor θ (indirect effect). Although both components are relevant, in the case of formal wages, the differences between the TT effects is mainly driven by the direct effect of the unobservable factor θ, while in the case of wages the effect is driven by the indirect effect associated with the bias in the coefficients of the observables in the outcomes equation. This implies that the selection on the gains is stronger in the formal market mainly in higher levels of education (i.e. $TT_{s,0} - TT^3_{s,0} > 0$), whereby individuals expecting to gain more from higher levels of education are more likely to enroll in higher levels of education. On the other hand, when I include informal wages in the analysis, the main effect is dominated by

the selection bias that affects the coefficients, and since the selection bias is negative (i.e. $TT^3_{s,0} - TT^2_{s,0} < 0$) then both dropouts and those in the level education $s > 0$ have a comparative advantage in the labor market.[7]

7.3 Policy relevant Treatment Effects

In this section, I focus on the returns associated with a particular policy: the PACES scholarship. I compute the Policy Relevant Treatment Effect (PRTE) (Heckman and Vytlacil (2001)) which measures the average returns for those individuals induced to change their schooling decisions in response to the high school scholarship, i.e. the PACES program. Notice that the PACES program only affects the first schooling transition directly. I compute the PRTE as the average of the difference in outcomes of those individuals that, under PACES program, would choose to graduate from high school, but without the program they would choose to dropout from high school, i.e. only for switchers. The definition of the PRTE is presented in equation 4.15. Table 7.4 shows the PRTE for the outcomes wages and formal wages.

The PACES scholarship induces 14% of the original participants who under the counterfactual scenario without scholarship would not graduate from high school to graduate from high school. 68.5 percent of those individuals stop at the high school graduate level, while 25.5 percent end with some college and 6 percent graduate from college. Figure 7.2 shows that the individuals who are induced to graduate from high school because the scholarship come from all deciles of the distribution of the unobservable $\epsilon_{0,1} = \theta + v_{0,1}$, but mainly from the middle of the distribution. This result is consistent with the theoretical model which shows that for some individuals $v_t(1) > v_t(0)$ when $\mathbb{S}_{it} = 0$, while for others $v_t(1) < v_t(0)$ for all values of $\mathbb{S}_{it} > 0$. The former case may include high ability individuals (i.e. high values of $\epsilon_{0,1}$), while the latter case may comprise low ability individuals (i.e. low values of $\epsilon_{0,1}$). The scholarship may be relevant for individuals mainly in the middle of the distribution such that $v_t(1) < v_t(0)$ for $\mathbb{S}_{it} = 0$, but when $\mathbb{S}_{it} > 0$ the inequality is reversed and $v_t(1) > v_t(0)$. The causal effects of the PACES program on wages through education are significant and highly non-linear on education. The overall return on wages of those induced

[7] This is consistent with the analysis of Willis and Rose (1979).

Table 7.4: Policy Relevant Treatment Effects

PRTE	Log wages	Log Formal wages
Overall Policy Effect	0.138***	0.125***
	(0.033)	(0.029)
HS Graduate	0.082***	0.058***
	(0.024)	(0.017)
Some College	0.213***	0.177***
	(0.035)	(0.042)
College Graduate	0.584***	0.524***
	(0.080)	(0.076)

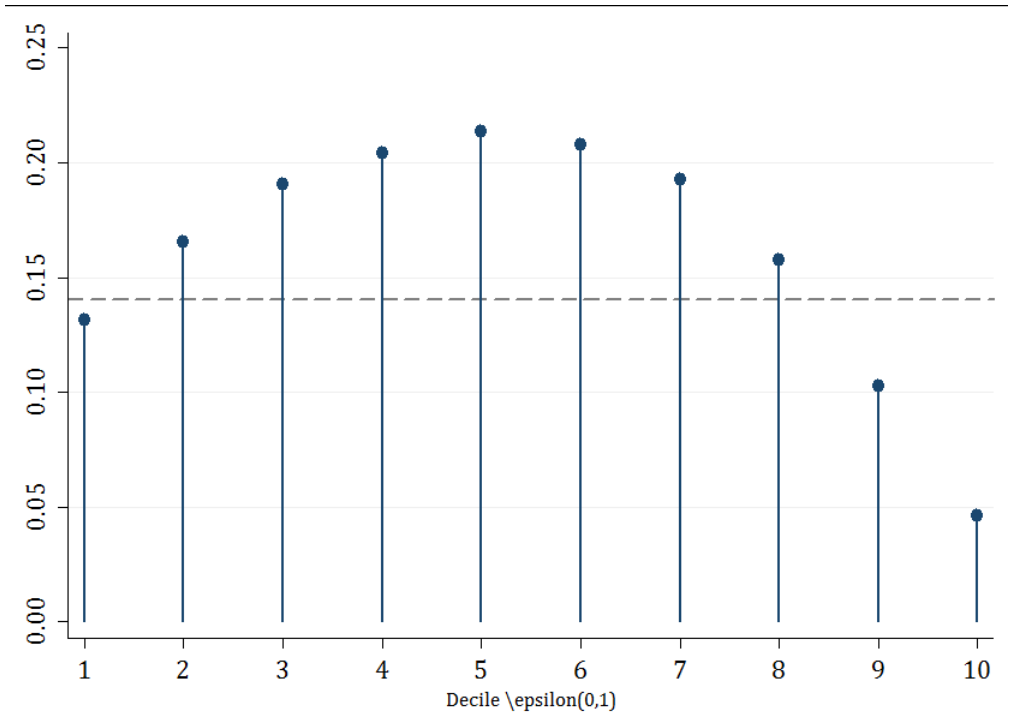
Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Standard errors are bootstrapped (200 replications). Table reports the policy-relevant treatment effect (PRTE), which measures the average returns of those individuals induced to change their schooling decisions in response to the high school scholarship, i.e. PACES program (see equation 4.15). The first column shows the results when the outcome of interest is the log of the monthly wage rate in 2010 (wages) which include both individuals who work in formal and informal labor market, while the second column shows the results when the outcome of interest is the log of monthly wage rate in 2010 in formal market only (formal wages). The first row shows the overall effect of the program. The second row shows the PRTE for those induced to enroll by the policy who then go on to complete high school. Similarly, rows three and four, show the results for individuals who complete some college and graduate from college respectively.

to graduate from high school because of the PACES program is 13.8 percent, while for formal wages the overall return is 12.5 percent. I also compute the PRTE conditional on different levels of schooling reached by those individuals affected by the program. The PRTE of wages for those who decided to stop their education at high school graduation is 8.2 percent, while for those who stopped with some college had a return of 21.3 percent. The PRTE for college graduation is 58.4 percent. In the case of formal wages, the returns are 5.8 percent for high school graduates, 17.7 for individuals with some college, and 52.4 for individuals who graduate from college. In general, the returns for those induced to graduate from high school because the PACES programs are higher than the returns of a randomly chosen individual.

7.4 Treatment and Policy Effects on Additional Outcomes

I also compute the standard treatment effects (ATE and TT) and the PRTE on two additional outcomes: the number of days worked in the formal market in 2010 and a quality life index, SISBEN score. The number of days worked is truncated at 0 and 360, since some individuals

Figure 7.2: Proportion of each decile of $\epsilon_{0,1}$ induced to switch by the PACES scholarship



reported that they only worked in the informal labor market which implies zero hours of work on the formal market, and some agents reported that they worked every day during 2010. To accommodate those cases I allow the part of the likelihood function associated with the outcome to have a Tobit type I form. The SISBEN score is a quality life index that it is used to classify Colombian households at different levels of poverty (i.e. SISBEN levels), which in turn allows the government to target the possible beneficiary of social public programs like health insurance, housing, and others conditional cash transfer programs (e.g. Familias en Accion). The SISBEN score goes from 0 to 100, where 0 represents the poorest households. If the score is less than 47.99 then the household is categorized as SISBEN level 1, while if the score is between 47.99 and 54.87, then the household is categorized as Sisben level 2. Households in Sisben 1 and 2 are considered poor.

Table E.9 shows the estimated coefficients for the potential outcomes given in Equation 4.2, while Table E.10 presents estimates of the standard treatment effects parameters associated with

each final level of education s, relative to the high school dropout level ($ATE_{s,0}$, $TT_{s,0}$ respectively). Finally, Table E.11 shows the PRTE. The causal effects of education on days worked in the formal labor market are substantial for high school graduates.[8] For a randomly chosen individual, the average treatment effect on wages of being a high school graduate with respect to being a high school dropout is 28 days (i.e. $ATE_{1,0}$), while the treatment effect on the treated is 29 days (i.e. $TT_{1,0}$). The causal effect of education explains 47 percent of the overall difference on the outcome between high school graduates with respect to being a high school dropout. For individuals with some college or college graduates the effects are not statistical significant.

The causal effects of the PACES program on days worked in the formal labor market through education are also significant. The overall PRTE for those induced to graduate from high school because of the PACES program is 45 days, while for those who decided to stop their education at high school graduation is 51 days, and 58 days for those who stopped with some college. The effect for college graduates is not statistical significant. In general, the effect on days worked in the formal labor market for those induced to graduate from high school because the PACES programs are higher than the effects for a randomly chosen individual.

Similarly, a randomly chosen individual has 6.8 points higher in the SISBEN score if he is a high school graduate as opposed to being a high school dropout (i.e. $ATE_{1,0}$), while the treatment effect on the treated is 7 points (i.e. $TT_{1,0}$). The ATE and TT are also statistically significant for higher levels of education with respect to the baseline level. Overall, the causal effects of the PACES program on the Sisben score is 9.5. Recall that all students were considered poor at the beginning of secondary school. The causal effect of the PACES program implies that the effect of the PACES program is big enough to move a person from be consider very poor ($Sisbenscore < 47.99$) to be poor ($47.99 < Sisbenscore < 54.87$) or no poor ($54.87 > Sisbenscore$).

[8]Figure E.4 shows the fit of the model for days worked in formal labor market in 2010.

APPENDIX A

COLOMBIA'S EDUCATIONAL SYSTEM AND PACES BACKGROUND

The Colombias educational system includes three levels of education: elementary school, secondary, and university. A child starts elementary school around age 6. The elementary school lasts 5 grade-years (grade 1 to 5). Secondary education includes two levels, Basic (grades 6 to 9) and Mid secondary (grades 10 to 11). By the time that PACES program was in place, 37 percent of the student population attended a private school at the national level. In Bogota, Colombias capital and largest city, 58 percent attended a secondary private school (Angrist et al. (2002)). In practice the transition between these two levels is smooth, which implies that these levels can be treated as one level. The national education system requires compulsory education for grades $1-9$. Students attending the last grade of secondary education are required by law to take the college entrance examinations called the ICFES exam. Over 95 percent of students who take the ICFES exam graduate from high school. Typically, taking the ICFES exam is used as a proxy for high school graduation (Angrist et al. (2006), Bettinger et al. (2014)).

University education is divided into undergraduate degrees and post-graduate. The undergraduate degrees include professional degrees ($4-5$ years), and technical/technological degrees ($2-3$ years), while the post-graduate education includes masters and PhD programs. The total enrollment rate in tertiary education in 2012 was 45 percent (WB (1999)). 75 percent of the students who attend tertiary education choose a professional degrees, while 25 percent choose technical/technological degrees. Figure E.5 summarizes the main characteristics of the Colombias educational system.

APPENDIX B

KOTLARSKI'S THEOREM

$v_{0,1}$, $v_{2,3}$ and θ are three independent variables. Define $\epsilon_{0,1} = \theta + v_{0,1}$ and $\epsilon_{2,3} = \theta + v_{2,3}$. If the characteristic function of ($\epsilon_{0,1}$, $\epsilon_{2,3}$) does not vanish, then the joint distribution of ($\epsilon_{0,1}$, $\epsilon_{2,3}$) determines the distribution of θ,$v_{0,1}$ and $v_{2,3}$ up to location. Proof: Kotlarski (1967)

APPENDIX C

SUFFICIENT CONDITIONS FOR IDENTIFICATION OF PARAMETERS AND DISTRIBUTIONS

Assume a model with one outcome is given by equations 4.1 and 4.2. I focus here on the case with three transitions with four potential outcomes presented in chapter 3.1. I impose an exogenously specified initial condition $D_{\infty,0} = 1$ which means that everyone starts with elementary school completed, and then I observe $V_{0,1} = Z_{0,1}\beta_{0,1} + \epsilon_{0,1}$ conditional on $D_{\infty,0} = 1$. I assume data on $D_{j,j'}$, $Z_{j,j'}$, Y_s, X_s from a random sample across individuals. Like any other discrete choice model, my model needs some normalizations. I implicitly normalize the β parameters associated with the choice of dropout from schooling system in each transition, so each index function is measured relative to the decision of dropout. Let $K_{0,1}$, $K_{2,1}$, $K_{3,2}$ the dimension of the vector parameters $\beta_{0,1}$, $\beta_{2,1}$, $\beta_{3,2}$ respectively. Also define $\epsilon = (\epsilon_{0,1}, \epsilon_{1,2}, \epsilon_{2,3})$, $\varepsilon = (\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3)$ and $\mu_s(X) = X\gamma_s$, $\varphi(Z) = (Z_{0,1}\beta_{1,0}, Z_{1,2}\beta_{1,2}, Z_{2,3}\beta_{2,3})$.

Proposition 1

Parameters $\beta_{j,j'}$,γ_s and distributions $F_{\epsilon_{j,j'}}$, F_{ε_s} for $s = 0, 1, 2, 3$ are exactly identified under the following conditions:

1. All components of (ϵ,ε) are continuous variables with zero mean, finite variance and $Supp(\epsilon, \varepsilon) = Supp(\epsilon)Supp(\varepsilon)$, with defined upper and lower limits for each component.

2. (ϵ,ε) is independent of (X, Z), where it holds for each component in (ϵ,ε).

3. $Supp(\mu_s(X), \varphi(Z)) = Supp(\mu_s(X))Supp(\varphi(Z))$ for $s = 0, 1, 2, 3$.

4. Conditions for the first transition: (a) the initial condition is defined exogenously, i.e. $D_{-\infty,0} = 1$ for all individuals. (b) There exists no proper linear subspace of $\mathbb{R}^K_{0,1}$ with probability 1 under $F_{Z_{0,1}}$. (c) There exists at least one $k \in [2, \cdots, K_{0,1}]$ such that $\beta^k_{0,1} \neq 0$ and such that, for almost every value of $z = (z^1_{0,1}, \cdots, z^{\text{k-1}}_{0,1}, z^{\text{k+1}}_{0,1}, \cdots, z^{\text{K}}_{0,1})$, $\Pr(z^k_{0,1} \in (a, b)|z) > 0$ for all

open intervals $(a, b) \in \mathbb{R}$.

5. Conditions for second and third transitions: (a) There exists no proper linear subspace of $\mathbb{R}^K_{j,j'}$ with probability 1 under $F_{Z_{j,j'}}$. (b) There exists at least one $k \in [2, \cdots, K_{j,j'}]$ such that $\beta^k_{j,j'} \neq 0$ and such that, for almost every value of $z = (z^1_{0,1}, \cdots, z^{\text{k-1}}_{0,1}, z^{\text{k+1}}_{0,1}, \cdots, z^{\text{K}}_{0,1})$, $\Pr(z^k_{j,j'} \in (a, b)|z, D_{j-1,j'-1} = 1) > 0$ for all open intervals $(a, b) \in \mathbb{R}$. (c) There exists a couple (τ_1, τ_2) such that for almost every $(\bar{\epsilon}_{0,1}, \bar{\epsilon}_{1,2}) \in (\epsilon_{1,2}, \epsilon_{2,3})$ with $\bar{\epsilon}_{0,1} \geq \tau_1$, $\bar{\epsilon}_{0,1} \geq \tau_2$, there is no proper linear subspace of $\mathbb{R}^K_{1,2}$ with probability 1 under $F_{Z_{1,2}|Z_{0,1}\beta_{1,0} \geq \bar{\epsilon}_{0,1}}$, and there is no proper linear subspace of $\mathbb{R}^K_{2,3}$ with probability 1 under $F_{Z_{2,3}|Z_{0,1}\beta_{1,0} \geq \bar{\epsilon}_{0,1}, Z_{1,2}\beta_{1,2} \geq \bar{\epsilon}_{1,2}}$.

Proposition 2

Parameters $\beta_{j,j'}$,γ_s and distributions $F_{\epsilon_{j,j'}}$, F_{ε_s} for $s = 0, 1, 2, 3$ are exactly identified under the following conditions:

1. All components of (ϵ,ε) are continuous variables with zero mean, finite variance and $Supp(\epsilon, \varepsilon) = Supp(\epsilon)Supp(\varepsilon)$, with defined upper and lower limits for each component.

2. (ϵ,ε) is independent of (X, Z), where it holds for each component in (ϵ,ε).

3. $Supp(\mu_s(X), \varphi(Z)) = Supp(\mu_s(X))Supp(\varphi(Z))$ for $s = 0, 1, 2, 3$.

4. Conditions for the first transition: (a) the initial condition is defined exogenously, i.e. $D_{-\infty,0} = 1$ for all individuals. (b) There exists no proper linear subspace of $\mathbb{R}^K_{0,1}$ with probability 1 under $F_{Z_{0,1}|D_0=1}$. (c) There exists at least one $k \in [2, \cdots, K_{0,1}]$ such that $\beta^k_{0,1} \neq 0$ and such that, for almost every value of $z = (z^1_{0,1}, \cdots, z^{\text{k-1}}_{0,1}, z^{\text{k+1}}_{0,1}, \cdots, z^{\text{K}}_{0,1})$, $\Pr(z^k_{0,1} \in (a, b)|z, D_0 = 1) > 0$ for all open intervals $(a, b) \in \mathbb{R}$.

5. Conditions for second and third transitions: Let $Z_{1,2} = Z_{2,3} = Z$, where $dim(Z) = K$. (a) $\bar{k} \geq 2$ variables of Z are continuous variables. Without loss of generality this $\bar{k}$ variables are the first components on Z. (b) The coefficients $\beta^1, \cdots, \beta^{\bar{k}}$ associated with the second and third transition are non-zero and linear independent.

Sketch of the proof. In both propositions, the key characteristic is that the initial condition $D_{-\infty,0} = 1$ is exogenously specified, which in my setup means that everyone starts with elementary school complete. This implies that the first schooling transition $D_{0,1}$ is not correlated with any previous schooling choice, i.e. it is free of dynamic selection. I relax this assumption by exploiting the randomization of the program before the first schooling choice. Let the program variable given by $R_{0,10,1}$, which is excluded from $Z_{1,2}$, $Z_{2,3}$. Notice that the randomization balanced or equated the bias in the sample of winners with the bias in the sample of losers (Heckman (1996)). In particular I assume $(Y_s, D_{0,1}, D_{1,2}, D_{2,3}, X)$ is independent of $R_{0,1}$.Then condition (iv) is satisfied in both propositions. The proof may starts by showing that $\beta_{0,1}$ and $F_{\epsilon_{0,1}}$ are identified. This part of the proof is straightforward and follows the strategy of Manski (1988), proposition 2, in both propositions. This strategy cannot be applied for the others transitions since those transitions are only observed for those individuals who choose to continue in previous transitions. Then I use an Identification in the Limit Argument (see Cameron and Heckman (1998)) for the second and third transitions. In general, here I need to generate measurable separability among the index value functions $V_{1,2}$ and $V_{2,3}$. Given that for the first transition $\beta_{0,1}$ and $F_{\epsilon_{0,1}}$ are identified, the condition (iv) in proposition 1 is easily satisfied with an exclusion restriction in the vector $Z_{1,2}$, and then $\beta_{1,2}$ and $F_{\epsilon_{1,2}}$ are identified. Similarly, given that $\beta_{0,1}$, $\beta_{1,2}$ and $F_{\epsilon_{0,1}}$, $F_{\epsilon_{1,2}}$ are identified, and an additional exclusion restriction in the vector $Z_{2,3}$ allow me to identify $\beta_{2,3}$ and $F_{\epsilon_{2,3}}$. The identification of $\beta_{1,2}$, $\beta_{2,3}$, $F_{\epsilon_{1,2}}$ and $F_{\epsilon_{2,3}}$ in proposition 2 requires fewer conditions. In particularly, I just need at least two continues variables which associated coefficients should be different than zero.

Once I identify $\beta_{j,j'}$ and distributions $F_{\epsilon_{j,j'}}$, the distribution of F_{ε_s} can be identified free of selection bias. Then is straightforward to identify γ_s. Finally, using the data Y_s and the construction

of $Y_s - \mu_s(X)$, and $\varphi(Z)$ I can find the joint distribution of $(\epsilon, \varepsilon_s)$ for all $s = 0, 1, 2, 3$. Nonetheless, given the conditions on either proposition 1 or 2, I cannot identify the joint distribution of $(\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3), andthen(\mathrm{Y}_0, Y_1, Y_2, Y_3)$. Assumption 1 allow me to recover the joint distribution of (Y_0, Y_1, Y_2, Y_3) and then the average treatment parameters of interest.

APPENDIX D

ALTERNATIVE SET OF ASSUMPTIONS

Assumptions 3

$F(\theta)$ is a discrete distribution with mass points $(P_m(\theta), \theta_m)_{m=1}^{M}$, where $P_m(\theta_m)$ is the probability associate with the mass point θ_m, and $\sum mP_m(\theta m) = 1$ (Heckman and Singer (1984), Mroz and Guilkey (2014), Mroz (1999)).

Under the assumptions 3, the likelihood function for the first stage is given by

$$\prod_{i=1}^{n} \sum_{m=1}^{M} \mathrm{Pr}(\mathfrak{D}_i | Z_i, \theta = \theta_m) \mathrm{P}_m(\theta_m) \tag{D.1}$$

In the second stage, I use first stage estimates, $\hat{P}_m(\theta_m)$ and $\hat{\mathrm{Pr}}(\mathfrak{D}_i | Z_i, \theta = \theta_m)$, to express the likelihood function as:

$$\prod_{i=1}^{n} \sum_{m=1}^{M} \hat{\mathrm{Pr}}(\mathfrak{D}_i | Z_i, \theta = \theta_m) f(Y_i | X_i, \theta = \theta_m) \mathrm{P}_m(\theta_m) \tag{D.2}$$

Cameron and Heckman (1998) provides the necessary conditions for the identification under assumption 3.

APPENDIX E

ADDITIONAL TABLES AND FIGURES

Figure E.1: Densities of $\Pr(D_{0,1} = 1|scholarship = 1)$ and $\Pr(D_{0,1} = 1|scholarship = 0)$

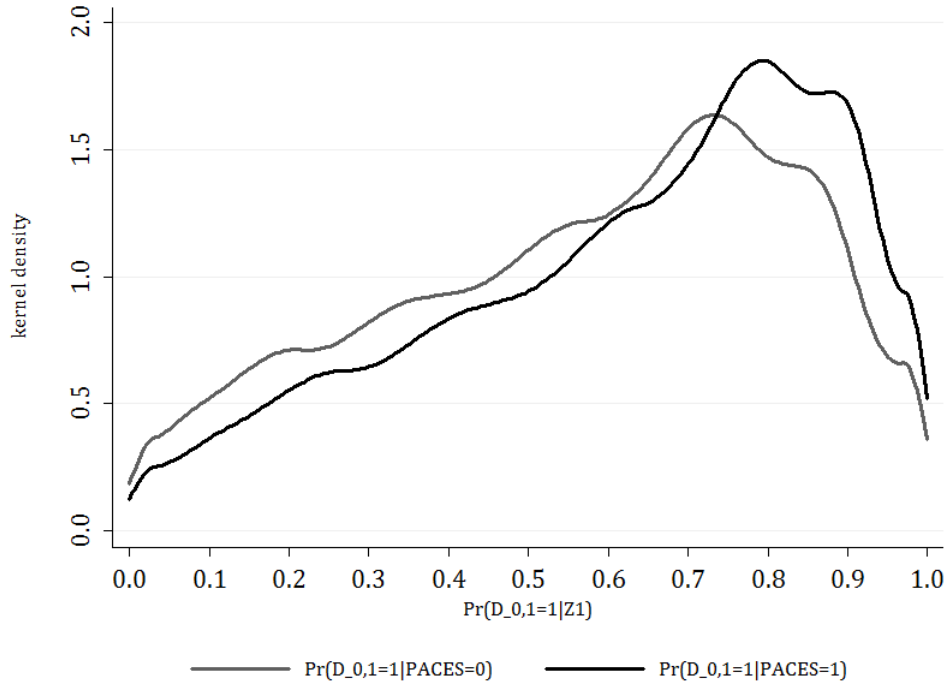


Table E.1: Variable definitions

Variable	Definition
Scholarship (dummy)	= 1 if individual was awarded the scholarship through the lottery.
Female (age)	= 1 if female.
Age	Age from the scholarship application.
Vocational HS	= 1 if High School has a vocational curriculum.
missing Vocational HS	= 1 if variable "Vocational HS" is missing.
Had Phone (dummy)	= 1 if the applicant had a phone number at the time of application.
Family Income [1,2) MW	= 1 if family income in last year of High School belongs to [1 MW,2 MW).
Family Income [2,3) MW	= 1 if family income in last year of High School belongs to [2 MW,3 MW).
Family Income [3,+) MW	= 1 if family income in last year of High School belongs to [3 MW,+).
Family Owned a House (dummy)	= 1 if individual's family owned a house in the last year of high school.
Mother education, Secondary	= 1 if mother has high school diploma in the last year of HS.
Mother education, College	= 1 if mother has college education in the last year of HS.
Number of Siblings	Number of siblings in the last year of high school.
Had a sibling in College	= 1 if individual has at least one sibling in college in the last year of HS.
Average Score High School	Avg score in the ICFES exam of the school in the last year of high school.
Class Size High School	Size of the class of the school in the last year of HS.
Accredited Program (High Quality)	= 1 if college program was certified as a high quality program
Days Worked Year	Number of days worked in the formal market during 2010.
Log Formal wages	Avg monthly labor earnings at primary job in the formal labor market in 2010.
Log wages	Log Formal wages + the self-reported monthly income in 2010.
Days Worked Year	Number of days worked in Colombia's formal labor market during 2010.
Sisben Score	It is a quality life index that goes from 0 to 100 (0 is the poorest).

Note: This table shows the definition of all variables used in the empirical analysis. Outcome variables include Log Formal wages, Log wages, Days Worked Year, Sisben Score. The baseline controls associated with the PACES program and available to all individuals are female, age and phone. Informal sector includes all workers that do not contribute to the Colombias social protection system.

Table E.2: Estimates for the sequential schooling model with an additional instrument

Variable/Transition	Dropout HS - Graduate HS		Graduate HS - Enroll College		Enroll College - Graduate College	
Scholarship (dummy)	0.174***	(0.050)	-	-	-	-
Female (dummy)	0.146***	(0.050)	0.162	(0.101)	0.267**	(0.149)
Age	-0.471***	(0.043)	-0.472**	(0.186)	-0.276*	(0.213)
Vocational HS	0.085*	(0.051)	0.092	(0.090)	0.046	(0.127)
missing Vocational HS	-0.355***	(0.103)	-0.161	(0.227)	0.026	(0.333)
Had Phone (dummy)	0.072	(0.075)	0.201	(0.147)	0.183	(0.234)
Family Income [1,2) MW	-	-	0.647***	(0.241)	0.162	(0.324)
Family Income [2,3) MW	-	-	0.755***	(0.276)	0.480	(0.401)
Family Income [3,+) MW	-	-	1.209***	(0.362)	0.167	(0.412)
Family Owned a House (dummy)	-	-	0.744***	(0.226)	0.687**	(0.384)
Mother education, Secondary	-	-	0.240**	(0.105)	0.186	(0.158)
Mother education, College	-	-	0.679**	(0.274)	0.672**	(0.343)
Number of Siblings	-	-	-0.027	(0.028)	-0.047	(0.044)
Had a sibling in College	-	-	0.181*	(0.098)	0.063	(0.137)
Average Score High School	-	-	2.692***	(0.930)	2.966**	(1.513)
Class Size	-	-	0.749**	(0.361)	-0.101	(0.467)
Accredited Program (High Quality)	-	-	-	-	0.674***	(0.232)
Constant	6.041**	(0.557)	2.307*	(1.344)	-0.341	(1.125)
Alpha	1.000	-	2.350**	(1.079)	0.863	(0.929)
Pr1				0.65		
Pr2				0.35		
Mean (Theta)				0.07		
SD(Theta)				0.532***		
ll				-3867.99		
LR test (p-value)				0.00		

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Sample is restricted to applicants with valid adult identification number (3926) that have complete application controls (3903) and have nonmissing information in other controls (3805). This table shows the maximum likelihood estimates of the sequential schooling model with potential outcomes for the first stage. Standard errors are in parentheses. This estimates includes an additional instrument for the last transition, accredited Program (High Quality). This variables takes the value of 1 if the college program in which the individual is enrolled is considered a program of high quality and has a certification, and 0 otherwise. The computation of the LR test includes two steps. The first step is to get the likelihood value for the first stage, i.e. the sequential schooling model. In the second step I constraint the loading factor of the sequential schooling model to be zero (i.e. no dynamic selection) and re-estimate the complete set of parameters to get a new likelihood value. Using these two likelihood values the LR test is constructed.

Table E.3: Estimates for the sequential schooling model with different distributional assumption on θ

Variable/Transition	Dropout HS - Graduate HS ( )		Graduate HS - Enroll College ( )		Enroll College - Graduate College ( )	
Scholarship (dummy)	0.205***	(0.054)				
Female (dummy)	0.163***	(0.054)	0.136	(0.101)	0.254*	(0.146)
Age	-0.541***	(0.027)	-0.396***	(0.244)	-0.247**	(0.235)
Vocational HS	0.079	(0.055)	0.075	(0.078)	0.049	(0.121)
missing Vocational HS	-0.413***	(0.120)	-0.142	(0.225)	0.031	(0.334)
Had Phone (dummy)	0.088	(0.083)	0.169	(0.135)	0.144	(0.214)
Family Income [1,2) MW			0.513***	(0.180)	0.085	(0.272)
Family Income [2,3) MW			0.600***	(0.208)	0.376	(0.325)
Family Income [3,+) MW			0.978***	(0.300)	0.078	(0.350)
Family Owned a House (dummy)			0.596***	(0.185)	0.537**	(0.273)
Mother education, Secondary			0.196**	(0.088)	0.133	(0.140)
Mother education, College			0.548**	(0.216)	0.598**	(0.294)
Number of Siblings			-0.020	(0.022)	-0.046	(0.041)
Had a sibling in College			0.145*	(0.081)	0.011	(0.125)
Average Score High School			2.184***	(0.766)	2.948**	(1.273)
Class Size			0.600**	(0.281)	-0.156	(0.436)
Constant	12.485***	(2.197)	2.810	(3.830)	0.134	(3.117)
Alpha (type 1)	0.000	-	0.000	-	0.000	-
Alpha (type 2)	6.950	(149.80)	4.412*	(2.524)	1.810	(2.483)
Alpha (type 3)	-5.861***	(1.24)	-0.868	(2.079)	-0.204	(1.994)
Pr1				0.14		
Pr2				0.03		
Pr3				0.83		
ll				3869.30		
LR test (p-value)				0.00		

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Sample is restricted to applicants with valid adult identification number (3926) that have complete application controls (3903) and have nonmissing information in other controls (3805). This table shows the maximum likelihood estimates of the sequential schooling model with potential outcomes for the first stage. Standard errors are in parentheses. This estimates assume that $F(\theta)$ is a discrete distribution with mass points $(P_m(\theta), \theta_m)_{m=1}^{M}$, where $P_m(\theta_m)$ is the probability associate with the mass point θ_m, and $\sum mP_m(\theta m) = 1$ (see Heckman and Singer (1984), Mroz and Guilkey (2014), Mroz (1999). The computation of the LR test includes two steps. The first step is to get the likelihood value for the first stage, i.e. the sequential schooling model. In the second step I constraint the loading factor of the sequential schooling model to be zero (i.e. no dynamic selection) and re-estimate the complete set of parameters to get a new likelihood value. Using these two likelihood values the LR test is constructed.

Table E.4: Estimates for the sequential schooling model with alternative specification

Variable/Transition	Dropout HS - Graduate HS		Graduate HS - Enroll College		Enroll College - Graduate College	
Scholarship (dummy)	0.174***	(0.050)	-	-	-	-
Female (dummy)	0.146***	(0.050)	0.151	(0.101)	0.257*	(0.136)
Age	-0.471***	(0.044)	-0.472**	(0.198)	-0.283*	(0.170)
Vocational HS	0.085*	(0.051)	0.099	(0.090)	0.052	(0.123)
missing Vocational HS	-0.355***	(0.103)	-0.156	(0.227)	-0.008	(0.320)
Had Phone (dummy)	0.071	(0.075)	0.195	(0.146)	0.171	(0.221)
Family Income [1,2) MW	-	-	0.651***	(0.247)	0.156	(0.282)
Family Income [2,3) MW	-	-	0.770***	(0.289)	0.458	(0.333)
Family Income [3,+) MW	-	-	1.234***	(0.378)	0.189	(0.353)
Family Owned a House (dummy)	-	-	0.740***	(0.234)	0.607**	(0.285)
Mother education, Secondary	-	-	0.251**	(0.108)	0.180	(0.140)
Mother education, College	-	-	0.685**	(0.280)	0.694**	(0.302)
Number of Siblings	-	-	-	-	-	-
Had a sibling in College	-	-	-	-	-	-
Average Score High School	-	-	2.746***	(0.959)	3.344**	(1.339)
Class Size	-	-	0.761**	(0.366)	-0.098	(0.456)
Constant	6.039**	(0.572)	2.293	(1.421)	-0.270	(1.042)
Alpha	1.000	-	2.333**	(1.070)	0.795	(0.740)
Pr1				0.65		
Pr2				0.35		
Mean (Theta)				0.00		
SD(Theta)				0.531***		
ll				-3878.05		
LR test (p-value)				0.01		

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Sample is restricted to applicants with valid adult identification number (3926) that have complete application controls (3903) and have nonmissing information in other controls (3805). This table shows the maximum likelihood estimates of the sequential schooling model with potential outcomes for the first stage. Standard errors are in parentheses. The computation of the LR test includes two steps. The first step is to get the likelihood value for the first stage, i.e. the sequential schooling model. In the second step I constraint the loading factor of the sequential schooling model to be zero (i.e. no dynamic selection) and re-estimate the complete set of parameters to get a new likelihood value. Using these two likelihood values the LR test is constructed.

Table E.5: Goodness of Fit for outcomes

Schooling final Level	Data	Model	p-value
HS Dropout	0.3987	0.3977	0.8990
HS Graduate	0.4000	0.4010	0.8991
Some College	0.1608	0.1628	0.7440
College Graduate	0.0405	0.0385	0.5310

Notes: The model data is simulated from the model estimates, and contains 1.1 million observations. The observed data contains 3805 observations from the PACES program. The p-value corresponds to a test of the two proportions are equal, where the null hypothesis is Data=Model.

Table E.6: Goodness of Fit for outcomes

	Mean		SD	
	Data	Model	Data	Model
Log wages	13.31	13.35	0.54	0.46
HS Dropout	13.15	13.23	0.56	0.43
HS Graduate	13.29	13.32	0.46	0.39
Some College	13.52	13.49	0.49	0.40
College Graduate	13.84	13.82	0.56	0.33
Log Formal wages	13.44	13.41	0.37	0.35
HS Dropout	13.33	13.30	0.29	0.25
HS Graduate	13.40	13.36	0.31	0.26
Some College	13.58	13.51	0.42	0.35
College Graduate	13.86	13.84	0.56	0.31
Days Worked Year	234.13	211.54	108.17	92.88
HS Dropout	214.99	206.67	110.19	31.07
HS Graduate	244.28	231.89	105.04	121.29
Some College	246.21	231.96	104.06	120.02
College Graduate	259.22	222.34	114.10	126.43
SISBEN score (Poverty Index)	51.30	53.98	17.57	7.13
HS Dropout	46.01	46.01	16.89	1.83
HS Graduate	53.51	53.51	17.10	2.35
Some College	59.96	59.96	15.66	3.36
College Graduate	63.78	66.43	14.67	14.64

Notes: The model data is simulated from the model estimates, and contains 1.1 million observations. The observed data contains 3805 observations from the PACES program. The p-value corresponds to a test of the two proportions are equal, where the null hypothesis is Data=Model.

Table E.7: Treatment Effects for final educational levels using specification with additional exclusion restrictions: High School dropout as a comparison group

Outcome/Treatment	Observed	ATE	TT
Log wages			
HS Graduate - HS Dropout	0.147	0.075***	0.083***
		(0.020)	(0.022)
Some College - HS Dropout	0.377	0.197***	0.225***
		(0.032)	(0.047)
College Graduate - HS Dropout	0.693	0.579***	0.479***
		(0.066)	(0.225)
Log Formal wages			
HS Graduate - HS Dropout	0.070	0.054***	0.060***
		(0.015)	(0.018)
Some College - HS Dropout	0.256	0.175***	0.281***
		(0.032)	(0.100)
College Graduate - HS Dropout	0.536	0.527***	0.615***
		(0.061)	(0.245)

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Standard errors are bootstrapped (200 replications). Each row compares the outcomes from a particular final level of education s with the high school dropout level. The first panel shows the results when the outcome of interest is the log of the monthly wage rate in 2010 (wages) which include both individuals who work in formal and informal labor market. The second panel shows the results when the outcome of interest is the log of monthly wage rate in 2010 in formal market only (formal wages). The first column is the observed difference in the data between two levels of education, i.e. $\mathbf{E}[Y|S=s]-\mathbf{E}[Y|S=0]$. The other two columns show the $ATE_{s,0}$, the average treatment effect obtained from the comparison of the outcomes associated with the schooling level $S=s$ relative to the schooling level $S=0$ for a random person in the population (see equation 4.13), and $TT_{s,0}$, the average treatment on the treated effect obtained from the comparison of the outcomes associated with the schooling level $S=s$ relative to the schooling level $S=0$ but only for those individuals which the final level of education is $S=s$ (see equation 4.14).

Table E.8: Treatment Effects for final educational levels without unobserved heterogeneity: High School dropout as a comparison group

Outcome	Observed	ATE		TT			Decomposition differences in TT		
		Model	No Selection	(1)	(2)	(3)	(1)-(2)	(1)-(3)	(3)-(2)
Log wages									
HS Graduate - HS Dropout	0.147	0.074	0.131	0.082	0.146	0.079	-0.064	0.003	-0.067
Some College - HS Dropout	0.377	0.198	0.300	0.226	0.387	0.241	-0.161	-0.016	-0.146
College Graduate - HS Dropout	0.693	0.578	0.687	0.478	0.689	0.553	-0.211	-0.075	-0.136
Log Formal wages									
HS Graduate - HS Dropout	0.070	0.055	0.059	0.060	0.066	0.061	-0.005	-0.001	-0.005
Some College - HS Dropout	0.256	0.177	0.203	0.284	0.262	0.230	0.022	0.054	-0.032
College Graduate - HS Dropout	0.536	0.527	0.507	0.617	0.539	0.557	0.078	0.060	0.018

Table E.9: Estimates coefficients for potential outcomes: Days worked formal market and Sisben score

Variable/Transition	HS Dropout	HS Graduate	Some College	College Graduate
		Panel C. Days Worked Formal Market		
Female (dummy)	-44.999***	-2.089	15.722***	-1.860
Age	3.174	-2.196	-4.254***	4.065
Constant	152.763***	369.804***	403.247***	316.624***
Alpha	113.331	237.173***	147.991***	141.365***
ll			-13232	
ll0			-14292	
LR-test (pvalue)			0.00	
		Panel D. Sisben Score		
Female (dummy)	-2.489**	-2.660**	-2.350	-0.429
Age	-1.029**	-1.661***	-2.805**	-3.282**
Constant	60.989***	75.842***	94.781***	104.492***
Alpha	-0.295	-0.575***	0.398	27.099***
ll			-12395	
ll0			-12407	
LR-test (pvalue)			0.00	

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. This table shows the maximum likelihood estimates of the sequential schooling model with potential outcomes for the second stage. To account for the fact that $\hat{F}_\theta$ and $\Pr(\mathfrak{D}_i|Z_i, \theta = \theta_m)$ are estimated in the first stage, the statistical significance of the parameters is obtained using bootstrap methods, in particular, I implement the percentile-t confidence interval method (see Horowitz (2001)). The results are robust to other bootstrapping techniques. The panel A shows the results when the outcome of interest is days worked in the formal market in 2010. The panel B shows the results when the outcome of interest is Sisben score. The computation of the LR test includes two steps.The computation of the LR test includes two steps. The first step is to get the likelihood value for the likelihood function including both first and second stage. In the second step I constraint the loading factor of the sequential schooling model and the potential outcomes to be zero (i.e. no dynamic selection, no essential heterogeneity) and re-estimate the complete set of parameters to get a new likelihood value. Using these two likelihood values the LR test is constructed.

Table E.10: Treatment Effects (other outcomes) for final educational levels: High School dropout as a comparison group

Outcome/Treatment	Observed	ATE	TT
Days Worked Year			
HS Graduate - HS Dropout	61.04	28.52**	28.87**
		(11.17)	(11.29)
Some College - HS Dropout	95.07	34.43	29.83
		(21.83)	(22.66)
College Graduate - HS Dropout	116.06	6.97	22.85
		(26.31)	(21.20)
SISBEN score (Poverty Index)			
HS Graduate - HS Dropout	7.50	6.75***	7.02***
		(1.12)	(1.09)
Some College - HS Dropout	13.96	11.34***	13.48***
		(3.10)	(4.68)
College Graduate - HS Dropout	17.78	16.09***	32.13***
		(3.74)	(3.89)

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Standard errors are bootstrapped (200 replications). Each row compares the outcomes from a particular final level of education s with the high school dropout level. The first panel shows the results when the outcome of interest is days worked in the formal market in 2010. The second panel shows the results when the outcome of interest is Sisben score. The first column is the observed difference in the data between two levels of education, i.e. $\mathbf{E}[Y|S=s]-\mathbf{E}[Y|S=0]$. The other two columns show the $ATE_{s,0}$, the average treatment effect obtained from the comparison of the outcomes associated with the schooling level $S=s$ relative to the schooling level $S=0$ for a random person in the population (see equation 4.13), and $TT_{s,0}$, the average treatment on the treated effect obtained from the comparison of the outcomes associated with the schooling level $S=s$ relative to the schooling level $S=0$ but only for those individuals which the final level of education is $S=s$ (see equation 4.14).

Table E.11: Policy Relevant Treatment Effects for other outcomes

PRTE	Days Worked Year	SISBEN score (Poverty Index)
Overall Policy Effect	45.03**	9.53***
	(20.689)	(1.156)
HS Graduate	50.77***	6.94***
	(19.510)	(1.019)
Some College	57.50**	11.77***
	(28.737)	(2.399)
College Graduate	33.25	14.80***
	(42.815)	(3.269)

Notes: $***$ Significant at 1%. $**$ Significant at 5%. $*$ Significant at 10%. Standard errors are bootstrapped (200 replications). Table reports the policy-relevant treatment effect (PRTE), which measures the average returns of those individuals induced to change their schooling decisions in response to the high school scholarship, i.e. PACES program (see equation 4.15). The first column shows the results when the outcome of interest is days worked in the formal market in 2010, while the second column shows the results when the outcome of interest is the Sisben score. The first row shows the overall effect of the program. The second row shows the PRTE for those induced to enroll by the policy who then go on to complete high school. Similarly, rows three and four, show the results for individuals who complete some college and graduate from college respectively.

Figure E.2: Observed and simulated percentiles of the log of the monthly wage rate in 2010

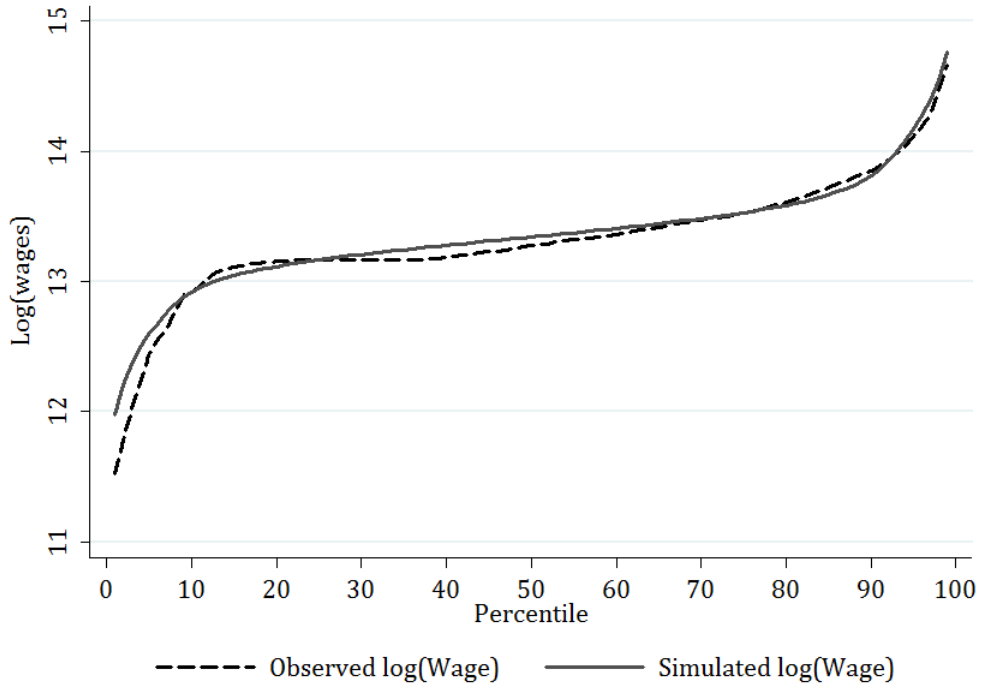


Figure E.3: Observed and simulated percentiles of the log of the monthly formal wage rate in 2010

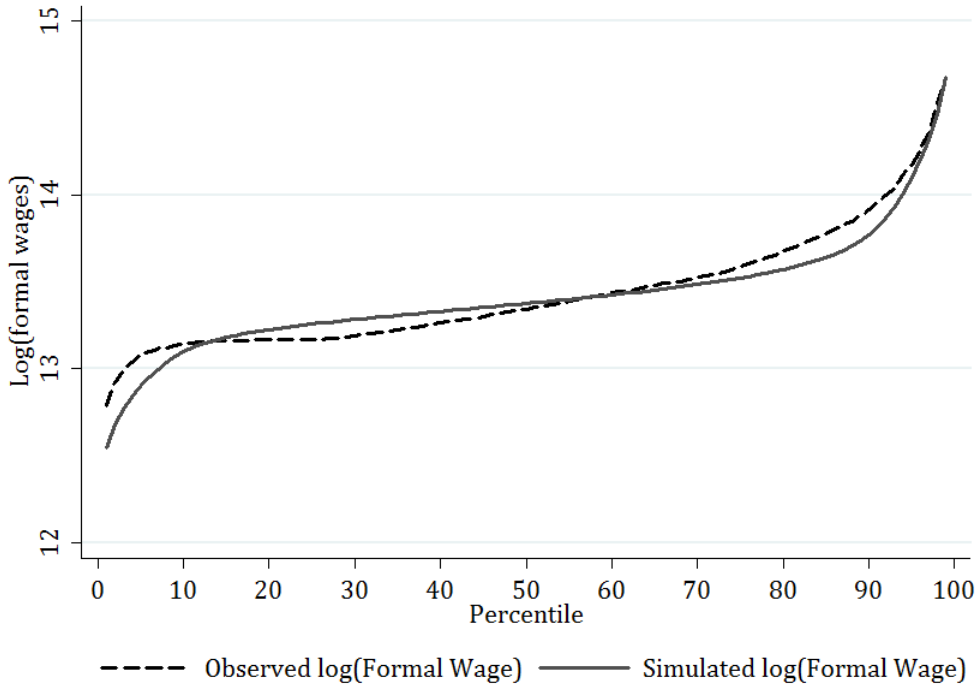


Figure E.4: Observed and simulated percentiles of Days worked in 2010

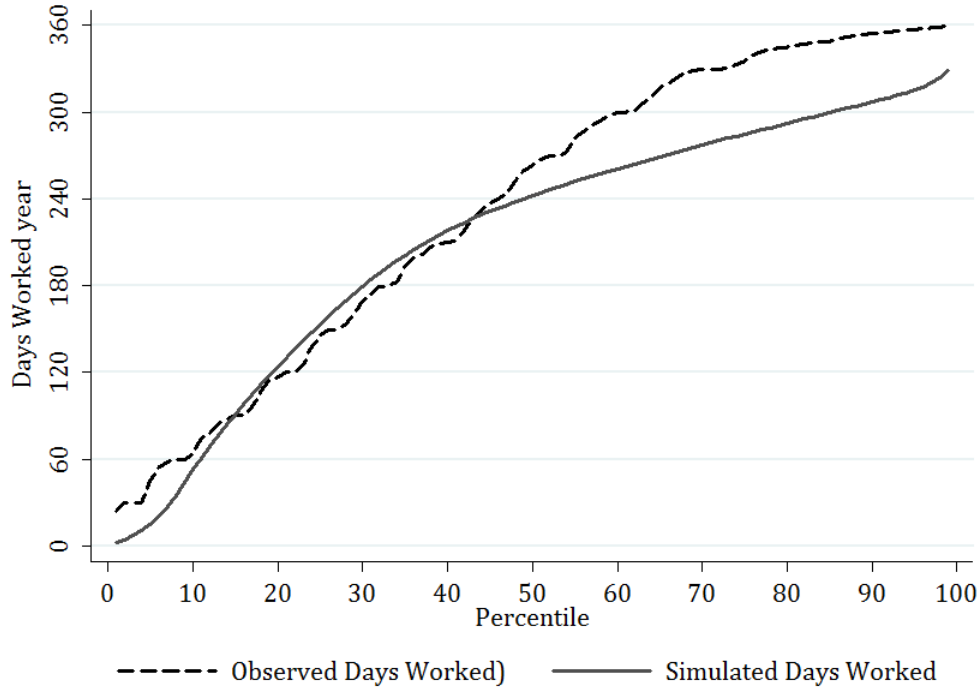
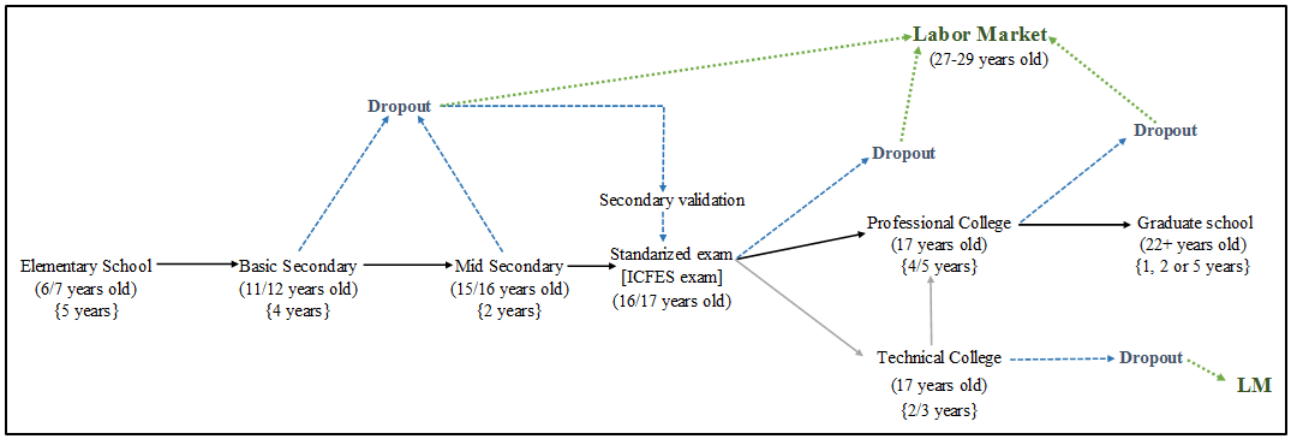


Figure E.5: Colombias Educational System



REFERENCES

Aakvik, A., J. Heckman, and E. Vytlacil (1999). Training Effects on Employment when the Training Effects are Heterogenous : an Application to Norwegian Vocational Rehabilitation Programs. Technical report.

Aakvik, A., J. Heckman, and E. Vytlacil (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs. *Journal of Econometrics 125*(1-2), 15–51.

Abbring, J. and G. Van den Berg (2003). The Nonparametric Identification of Treatment Effects in Duration Models. *Econometrica 71*(5), 1491–1517.

Akbulut, M. (2014). Children of War: The Long-Run Effects of Large-Scale Physical Destruction and Warfare on Children. *Journal of Human Resources 49*(3), 634–662.

Akbulut, M. (2015). War during childhood: The long run effects of warfare on health. *Working Paper*.

Akbulut, M., M. Khamis, and M. Yuksel. Long-term effects of postwar mobilization on female labor market outcomes. *Working Paper*.

Akee, R., W. Copeland, G. Keeler, A. Angold, and J. Costello (2010). Parents' Incomes and Children's Outcomes: A Quasi-experiment Using Transfer Payments from Casino Profits. *American Economic Journal: Applied Economics 2*(1), 86–115.

Almond, A. (2006). Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy 114*(4), 672–712.

Almond, D. and J. Currie (2011). *Human Capital Development before Age Five*, Volume 4 of *Handbook of Labor Economics*, Chapter 15, pp. 1315–1486. Elsevier.

Angrist, J., E. Bettinger, , E. Bloom, E. Kin, and M. Kremer (2002). Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review 92*(5), 1535–1558.

Angrist, J., E. Bettinger, and M. Kremer (2006). Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. *American Economic Review 96*(3), 847–862.

Angrist, J. and G. Imbens (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–75.

Attanasio, O., C. Meghir, and A. Santiago (2012). Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. *Review of Economic Studies 79*(1), 37–66.

Belskaya, O., K. Peter, and C. Posso (2014). College Expansion and the Marginal Returns to

Education: Evidence from Russia. IZA Discussion Papers 8735, Institute for the Study of Labor (IZA).

Belzil, C. (2007). The return to schooling in structural dynamic models: a survey. *European Economic Review 51*(5), 1059–1105.

Belzil, C. (2008). Testing the Specification of the Mincer Wage Equation. *Annales d'Economie et de Statistique* (91-92), 427–451.

Belzil, C. and J. Hansen (2007). A structural analysis of the correlated random coefficient wage regression model. *Journal of Econometrics 140*(2), 827–848.

Bettinger, E., M. Kremer, M. Kugler, C. Medina, C. Posso, and J. Saavedra (2014). Educational, labor market, and fiscal impacts of scholarships for private secondary school: Evidence from colombia. *Working Paper*.

Bettinger, E., M. Kremer, and J. Saavedra (2010). Are Educational Vouchers Only Redistributive? *Economic Journal 120*(546), F204–F228.

Black, S., P. Devereux, and K. Salvanes (2005). The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *The Quarterly Journal of Economics 120*(2), 669–700.

Cameron, S. and J. Heckman (1998). Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy 106*(2), 262–333.

Cameron, S. and J. Heckman (2001). The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy 109*(3), 455–499.

Campbel, F., G. Cont, J. Heckman, S. Moon, R. Pinto, G. Pungello, and Y. Pan (2014). Early childhood investments substantially boost adult health. *Science 343*, 147885.

Campbell, F., G. Pungello, M. Burchinal, K. Kainz, Y. Pan, B. Wasik, O. Barbarin, J. Sparling, and C. Ramey (2012). Adult Outcomes as a Function of an Early Childhood Educational Program: An Abecedarian Project Follow-Up. *Developmental Psychology 48*(4), 10331043.

Carneiro, P., K. Hansen, and J. Heckman (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Carneiro, P., J. Heckman, and E. Vytlacil (2011). Estimating Marginal Returns to Education. *American Economic Review 101*(6), 2754–81.

Chetty, R., J. Friedman, N. Hilger, E. Saez, D. Whitmore, and D. Yagan (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics 126*(4), 1593–1660.

Conti, G., M. S. Heckman, J., and R. Pinto (2014). The long-term health effects of early childhood interventions. *Working Paper*.

Cooley, J., S. Navarro, and Y. Takahashi (2014). A Framework for the Analysis of Time-varying Treatment Effects: How The Timing of Grade Retention Affects Outcomes. *Working Paper*.

Cunha, F., J. Heckman, and S. Navarro (2007). The Identification And Economic Content Of Ordered Choice Models With Stochastic Thresholds. *International Economic Review 48*(4), 1273–1309.

Cunha, F., J. Heckman, and S. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Dubois, P., A. Janvry, and E. Sadoulet (2012). Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico. *Journal of Labor Economics 30*(3), 555 – 589.

Fletcher, J. (2013). Adolescent Depression and Adult Labor Market Outcomes. *Southern Economic Journal 80*(1), 26–49.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, W. Walker, S. Chang, and S. Grantham (2013, June). Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica. NBER Working Papers 19185, National Bureau of Economic Research, Inc.

Haavelmo, T. (1943). The Statistical Implications of a System of Simultaneous Equations. *Southern Economic Journal 11*(1), 1–12.

Harmon, C., H. Oosterbeek, and I. Walker (2003). The Returns to Education: Microeconomics. *Journal of Economic Surveys 17*(2), 115–156.

Heckman, J. (1996). Randomization as an Instrumental Variable: Notes. *The Review of Economics and Statistics 78*(2), 336–41.

Heckman, J., J. Humphries, S. Urzua, and G. Veramendi (2011). The effects of educational choices on labor market, health, and social outcomes. Working papers, Human Capital and Economic Opportunity Working Group.

Heckman, J., L. Lochner, and P. Todd (2006, June). *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond*, Volume 1 of *Handbook of the Economics of Education*, Chapter 7, pp. 307–458. Elsevier.

Heckman, J., S. Moon, R. Pinto, P. Savelyev, and A. Yavitz (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

Heckman, J. and S. Mosso (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics 6*(1), 689–733.

Heckman, J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review 103*(6), 2052–86.

Heckman, J., D. Schmierer, and S. Urzua (2010). Testing the correlated random coefficient model. *Journal of Econometrics 158*(2), 177–203.

Heckman, J. and B. Singer (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica 52*(2), 271–320.

Heckman, J., S. Urzua, and E. Vytlacil (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

Heckman, J. and E. Vytlacil (2001). Policy-Relevant Treatment Effects. *American Economic Review 91*(2), 107–111.

Heckman, J. and E. Vytlacil (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica 73*(3), 669–738.

Hennessy, C. and I. Strebulaev (2015). Natural Experiment Policy Evaluation: A Critique. NBER Working Papers 20978, National Bureau of Economic Research, Inc.

Horowitz, J. (2001). The Bootstrap. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5 of *Handbook of Econometrics*, Chapter 52, pp. 3159–3228. Elsevier.

Huggett, M., G. Ventura, and A. Yaron (2011). Sources of Lifetime Inequality. *The American Economic Review 101*(7), 2923–2954.

Imai, K., L. Keele, , and T. Yamamoto (2010). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science 25*(1), 5171.

Imai, K., L. Keele, D. Tingley, and T. Yamamoto (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review 105*(4), 765–789.

Juhn, C., R. J. and A. Zuppann (2015). The Quantity-Quality Tradeoff and the Formation of Cognitive and Non-cognitive Skills. *Working Paper*.

Keane, M. and K. Wolpin (1997). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Kesternich, I., B. Siflinger, J. Smith, and J. Winter (2014). The Effects of World War II on Economic and Health Outcomes across Europe. *The Review of Economics and Statistics 96*(1), 103–118.

King, E., L. Rawlings, M. Gutierrez, C. Pardo, and C. Torres (1997). Colombia's Targeted Education Voucher Program: Features, Coverage and Participation. *Working Paper*.

Koopmans, T. (1947). Measurement Without Theory. *The Review of Economics and Statistics 29*(3), 161–172.

Kotlarski, I. (1967). On Characterizing the Gamma and Normal Distribution. *Pacific Journal of Mathematics 20*, 69–76.

Lee, S. and A. Seshadri (2014). On the intergenerational transmission of economic status. *Working Paper, University of WisconsinMadison, Department of Economics*.

Ludwig, J., J. Kling, and S. Mullainathan (2011). Mechanism Experiments and Policy Evaluations. *Journal of Economic Perspectives 25*(3), 17–38.

Maddala, G. (2010). *Limited-Dependent and Qualitative Variables in Economics*. Cambridge University Press.

Maluccio, J., J. Hoddinott, J. Behrman, R. Martorell, A. Quisumbing, and A. Stein (2009). The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults. *Economic Journal 119*(537), 734–763.

Manski, C. (1988). Identification of binary response models. *Journal of the American Statistical Association 83*(403), 729738.

Matzkin, R. L. (2003). Nonparametric Estimation of Nonadditive Random Functions. *Econometrica 71*(5), 1339–1375.

Meghir, C., M. Palme, and E. Simeonova (2012). Education, Health and Mortality: Evidence from a Social Experiment. NBER Working Papers 17932, National Bureau of Economic Research, Inc.

Mroz, T. and D. Guilkey (2014). Discrete factor approximation for use in simultaneous equation models with both continuous and discrete endogenous variables. *Working Paper, Carolina Population Center*.

Mroz, T. A. (1999). Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome. *Journal of Econometrics 92*(2), 233–274.

Naylor, J. and, S. A. (1982). Applications of a method for the efficient computation of posterior distribution. *Applied Statistics 31*(2), 214225.

Oreopoulos, P. and U. Petronijevic (2013). Making College Worth It: A Review of Research on the Returns to Higher Education. NBER Working Papers 19053, National Bureau of Economic Research, Inc.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics 128*(2), 301–323.

Ravallion, M. (2015). On the Implications of Essential Heterogeneity for Estimating Causal Impacts Using Social Experiments. *Journal of Econometric Methods 4*(1), 7.

Reyes, L., J. Rodrguez, and S. Urza (2013). Heterogeneous Economic Returns to Postsecondary Degrees: Evidence from Chile. NBER Working Papers 18817, National Bureau of Economic Research, Inc.

Ribero, R. and J. Tenjo (1997). Evaluacion del programa de becas PACES. *Working Paper*.

Rosenzweig, M. and K. Wolpin (1980). Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment. *Econometrica 48*(1), 227–40.

Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Chapman Hall/CRC.

Smith, J. (2009). The Impact of Childhood Health on Adult Labor Market Outcomes. *The Review of Economics and Statistics 91*(3), 478–489.

Todd, P. and K. Wolpin (2006). Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility. *American Economic Review 96*(5), 1384–1417.

Vytlacil, E. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica 70*(1), 331–341.

Vytlacil, E. (2006). A Note on Additive Separability and Latent Index Models of Binary Choice: Representation Results. *Oxford Bulletin of Economics and Statistics 68*(4), 0305–9049.

WB (1999). *World development report 1998/99*. Oxford University.

Willis, R. and S. Rose (1979). Education and Self-Selection. *Journal of Political Economy 87*(5), S7–S36.