

Wandering Astray: Teenagers' Choices of Schooling and Crime*

Chao Fu^a, Nicolás Grau^b, and Jorge Rivera^{b†}

August 30, 2021

Abstract

We build and estimate a dynamic model of teenagers' choices of schooling and crime, incorporating four factors that may contribute to the different paths taken by different teenagers: heterogeneous endowments, unequal opportunities, uncertainties about one's own ability, and contemporaneous shocks. We estimate the model using administrative panel data from Chile that link school records with juvenile criminal records. Counterfactual policy experiments suggest that, for teenagers with disadvantaged backgrounds, interventions that combine mild improvement in their schooling opportunities with free tuition (by adding 157 USD per teenager-year to the existing high school voucher) would lead to an 11% decrease in the fraction of those ever arrested by age 18 and a 13% increase in the fraction of those consistently enrolled throughout primary and secondary education.

*We thank Tatiana Reyes and Damián Vergara for excellent research assistance. We thank seminar and conference participants at Duke, PUC-Rio, Universidad de los Andes (Colombia), Universidad de Chile, and Econometric Society Meetings for helpful comments. Nicolás Grau thanks the Centre for Social Conflict and Cohesion Studies (ANID/ FONDAP/15130009) for financial support. Jorge Rivera thanks the support from the ANID PIA/APOYO AFB180003. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). All errors are ours.

†*a* : University of Wisconsin and NBER, cfu@ssc.wisc.edu, *b* : Faculty of Economics and Business, University of Chile, ngrau@fen.uchile.cl, jrivera@fen.uchile.cl.

1 Introduction

Teenage years are a critical period in life, featuring major physical, psychological and attitudinal transitions. Faced with all these complications, some teenagers may experience a particularly difficult transition to adulthood and wander astray, dropping out of school and/or engaging in criminal activities. Juvenile delinquency is a serious problem worldwide. For example, in the U.S., over 725,000 teenagers were in detention centers in 2011, 40% of whom were black. Understanding why some teenagers wander astray is key to many social policies.

In this paper, we build and estimate a dynamic model of teenagers' decisions of schooling and crime. We consider, in a coherent framework, four potentially important factors underlying different paths chosen by teenagers. The first factor is the heterogeneous endowments received by teenage years: Teenagers come from different family backgrounds, attend primary schools of different quality, have different ability levels and different preferences.¹

The second factor consists of frictions that lead to unequal opportunities. First, schools can select students based on their backgrounds. Such selection is prevalent worldwide, although it may be more explicit in some countries than it is in other countries.² Second, a teenager's payoff from schooling may depend directly on one's family background.³ Given these institutional frictions, teenagers of the same caliber but from different backgrounds may make different choices, which endogenously exacerbates initial inequality.

The third factor is the uncertainty faced by teenagers about themselves and their future prospects. In our model, a forward-looking teenager is uncertain about his own ability or productivity in school, but has a belief about it. Given his belief, a teenager chooses, in each period (if not in jail), whether or not to attend school and whether or not to participate in crime and thereby face the risk of being arrested and punished. If choosing to attend school, the teenager's test score (GPA) is realized at the end of the period, which depends on one's school and family characteristics, one's ability and a test score shock. A teenager uses his test score to update his belief about his ability, which will in turn affect his future decisions.

¹Our model focuses on teenagers' decisions and takes the endowment by teenage years as pre-determined. The word ability throughout this paper refers to one's ability pre-determined by one's innate ability and the investment received by teenage years. See Heckman and Mosso (2014) for a comprehensive review of the literature on early human development and social mobility.

²For example, in many countries, the allocation of students to public schools often uses neighborhood-based priority rules. As a result, children living in poor neighborhoods have little chance of getting into good public schools.

³For example, a high school degree may not provide low-family-income teenagers the same option value of going to college as it does for teenagers with richer parents, if colleges are not affordable for the poor.

The last factor is luck, modeled as contemporaneous shocks that may affect one’s choices. Although contemporaneous, these shocks can have long-term impacts because of the dynamic nature of one’s choices, i.e., current choices may affect choice-specific payoffs in the future. Moreover, the impact of “luck” can be particularly strong when it interacts with other factors. For example, given uncertainties about oneself, bad test shocks can lower a teenager’s belief about his prospect of schooling and discourage him from following this path. Bad test scores can also limit a teenager’s choice set given institutional frictions such as selective high school admissions.

We apply our model to administrative data from Chile, which offer a great opportunity to study the paths chosen by different teenagers. In particular, we have linked administrative primary school and high school records of teenagers from the 34 largest Chilean municipalities with their juvenile criminal records. The linked data sets provide information on these teenagers’ family backgrounds, school characteristics, annual academic records of enrollment and performance, and annual criminal records of arrests and sentences. We estimate the model via the maximum likelihood method to recover parameters governing the distribution of preferences and abilities across teenagers from different backgrounds, the degree of uncertainties they face, the outcome test score production function, and how primary-to-secondary school transfer opportunities may vary with one’s observable and unobservable characteristics. The estimated model fits the data well.

We use the estimated model to first understand the importance of the key factors in our model in explaining choices made by teenagers from different backgrounds. In particular, we compare the outcomes of three groups of teenagers in the baseline and in counterfactual scenarios with respect to initial endowment, institutional friction, and information friction. The three groups of teenagers are representative of those with disadvantaged, middle, and advantaged backgrounds, respectively. We find that family background is by far the most important factor underlying the cross-group difference in teenagers’ choices, followed by primary-secondary school transfer opportunities and unobservable initial endowments.

Then, we conduct a series of counterfactual policy interventions, targeted at teenagers with disadvantaged family and primary school backgrounds, who are more likely to wander astray. We find that policies that make schools free alone and policies that improve schooling opportunities (primary school environment and primary-to-secondary school transfer opportunities) would have limited effects. A combination of these two types of measures would be twice as effective in keeping the targeted teenagers on the right track as either type of measures alone. For

example, a free-tuition treatment combined with a decent schooling opportunity would increase the fraction of those consistently enrolled throughout primary and secondary education from 69.3% to 78.5%, and reduce the fraction of those ever arrested by age 18 from 5.38% to 4.79%.⁴ Specifically, this policy requires adding about 157 USD to the existing voucher per targeted teenager per high school year, enrolling a targeted teenager in a median-quality primary school and giving him the same primary-to-secondary school transfer opportunities as faced by his counterpart from middle family backgrounds. At a higher cost of 492 USD per teenager-year, enhancing the previous intervention with a guaranteed free seat in a decent private high school would almost double the effect.

Our paper contributes to the broad literature on youth crime, which has investigated a wide range of potential factors affecting criminal activities.⁵ Closer to our paper is a strand of studies on the relationship between schooling and crime. For example, Lochner and Moretti (2004) find that school attainment significantly reduces participation in criminal activity. Freeman (1996) and Lochner (2004) study the incapacitation effect of schooling and emphasize that education increases opportunity costs of crime.⁶ Lochner (2011) explores the effect of education on crime reduction via social and emotional development. Using a sequential model of schooling choice and a logistic model of arrests, Aucejo and James (2019) examine the importance of family backgrounds, early skills and behaviors in explaining gender gaps in education attainment. We contribute to this literature by developing and estimating a dynamic model of teenagers' choices of schooling and crime, and using the estimated model to examine how counterfactual interventions may affect education attainment as well as criminal behavior.

Another related set of studies focus on the persistence of youth crime.⁷ For example, Nagin and Paternoster (1991), Nagin and Land (1993), Nagin et al. (1995), and Broidy et al. (2003)

⁴As is true in any individual decision model, results from our experiments abstract from potential equilibrium impacts and should be interpreted at the individual level. See, for example, Billings et al. (2014), who study the effect of a large-scale school re-segregation policy on education and crime outcomes.

⁵For example, Levitt and Lochner (2000) examine biological, social, criminal justice, and economic determinants of crimes. Factors that have been examined in more detail include the effect of police, incarceration and conditions in prison (Levitt 1996, 1997, 1998; Katz et al. 2003; and Di Tella and Schargrodsky 2004), social capital, social interactions and peer effects (Case and Katz 1991; Glaeser et al. 1996; Gaviria and Raphael 2001; Jacob and Lefgren 2003; Kling et al. 2005; and Sickles and Williams 2008), and family circumstances and structure (Glaeser and Sacerdote 1999; Donohue and Levitt 2001).

⁶Jacob and Lefgren (2003) find that youth property (violent) crime rates are lower (higher) on days when school is in session. Other studies have explored the extensions of the mandatory schooling age or the cutoff birth date for enrollment to study the effect of education on crime (Machin et al. 2011; Clay et al. 2012; Anderson 2014; Hjalmarsson et al. 2015; and Cook and Kang 2016).

⁷See Blumstein et al. (1986), Gottfredson and Hirschi (1990), Sampson and Laub (1995), and Wilson and Herrnstein (1985) for examples in criminology.

find that both unobserved individual types and state dependence are important. Merlo and Wolpin (2009), via a VAR approach, find important roles for heterogeneity in initial conditions and stochastic events in one’s youth in determining outcomes as young adults. Mancino et al. (2015) find small effects of experience (the accumulation of education and crime) and stronger evidence of state dependence. Building on these studies, we incorporate unobserved heterogeneity, information friction, institutional friction and stochastic events into a coherent framework to study teenagers’ schooling-crime paths.

Most studies in economics on crime follow the framework of Becker (1968), where individuals make rational choices.⁸ For example, Imai and Krishna (2004) model an individual’s dynamic decisions on crime participation, where being caught can affect one’s high school graduation and labor market outcomes.⁹ They find that one’s concerns over future labor market outcomes is important in deterring crime. Our paper well complements Imai and Krishna (2004). We explicitly model one’s choices of both crime and schooling, the latter being treated as exogenous in their paper; however, our panel is not long enough for us to explicitly study the link between arrests and job market outcomes. As in theirs, our paper maintains the assumption of forward-looking and rationality. However, we allow for the possibility that teenagers may be ill-informed about their own abilities.

In the rest of the paper, Section 2 describes the background, Section 3 describes our data, Section 4 introduces the model, Section 5 describes our estimation strategy, followed by estimation results, Section 7 conducts policy simulations, Section 8 concludes the paper. The appendix contains additional details.

2 Background

2.1 Teenage Crime in Chile

As is true in many other countries, teenage crime is a serious social issue in Chile. For example, in 2013, the number of arrests per 100 teens was 2.93 in Chile, as compared to 3 in the U.S.

⁸Some studies have questioned the assumption of rationality and forward-looking. For example, Wilson and Herrnstein (1985), Katz et al. (2003) and Lee and McCrary (2005) suggest that potential offenders may have very high discount rates or be myopic. On the other hand, consistent with individuals being forward-looking, Levitt (1998) shows that arrests decline faster after age 18 in states where punishments are relatively mild for juveniles than for adults than states where juvenile punishments are relatively harsh.

⁹Munyo (2015) calibrates a dynamic model of youths decisions between working and crime, with a special focus on different punishments for youths than for adults. Hjalmarsson (2008) and Cortés et al. (2020) find that arrest and incarceration prior to age 16 reduces the probability of high school graduation.

in 2014 (Department of Justice). In 2007, the Chilean Parliament passed the new Adolescent Criminal Responsibility Law. Under this law, offenders aged 14 to 18 may be sentenced to participate in social reinsertion programs in closed or semi-closed detention centers; the sentence is up to 5 or 10 years depending on whether one is aged below or above 16. Other penalties include parole, community service and damage reparation. Between 2007 and 2014, the average number of teens prosecuted under this law was about 20,160 per year, 83% of whom were male. Given the rare incidence of crimes committed by girls, we focus our study on boys.

2.2 Primary and Secondary Education in Chile

Three Types of Schools: In 1980, the Chilean government introduced a major reform of its education system, one component of which was the introduction of vouchers such that schools were funded with a flat voucher per attendee.¹⁰ Both public and private schools can receive vouchers from the government. Since 1994, private schools have been allowed to charge additional fees up to a cap while still being eligible to receive government vouchers, subject to progressive crowd-outs.¹¹ Partly because of the tuition cap, some private schools opted out of the voucher system. As a result, there are three types of schools: municipal (public), private with voucher (voucher-private), and private without voucher charging much higher tuition (non-voucher-private). In 2014, 54.2% of the schools were public, 41% were voucher-private, and 4.8% were non-voucher-private. Of all students between Grade 1 and Grade 12, 39.1% were enrolled in public schools, 53.2% were enrolled in voucher-private schools, and the remaining 7.7%, those with “elite” family backgrounds, were enrolled in non-voucher-private schools. We focus on students enrolled in public and voucher-private schools.

Primary-Secondary School Transfers: Among all public schools, 9.9% offer both primary education (Grades 1 to 8) and secondary education (Grades 9 to 12), 80.6% offer only primary education and 9.5% offer only secondary education. These figures are 38.5%, 48.5% and 13% among voucher-private schools. Therefore, most students have to change schools between pri-

¹⁰The other major components of the reform were 1) decentralizing public schools from the central government to municipalities, 2) making teachers’ contracts more flexible. For a summary of these reforms, see Gauri (1999) and Mizala and Romaguera (2000). In 2008, additional targeted vouchers (Grades 1 to 4) were introduced to transfer more resources to schools catering the poorest 40% of the population. Flat vouchers were used for our sample cohort (Grade 1 in 2002).

¹¹In 2014, the cap was 84,232 pesos (\$1 is about 790 pesos) per month (there are 10 academic months in Chile). Attendance fees crowd out government vouchers progressively, where the crowd-out rates are 0 for the first 0.5USE (1USE is about 22,321 pesos), 10% for the part between 0.5USE and 1USE, 20% for the part between 1USE and 2USE, and 35% for the part above 2USE.

mary and secondary education. During our sample period (2006-2013), school admissions were decentralized and conducted by individual schools. Some high schools selected students not only on primary school GPAs but may also on family backgrounds.¹² The selective admission procedure, together with differential fees, contributed to the segregation of students of different socioeconomic statuses (SES) across schools. An OECD review noted that “social segmentation has deepened such that increasingly, students from the same or similar socioeconomic backgrounds are schooled together”.¹³

SIMCE Test and GPA: As a way to measure student learning and school performance, the government introduced in 1988 a system of national standardized tests (SIMCE), in which all students in proper grades participate (Meckes and Carrasco 2010). The government uses SIMCE results to allocate resources and to inform the public about the quality of schools by publicizing school-level results. However, individual test results are not released to students or schools. We have obtained individual test results, which will facilitate our identification.

At the individual level, the most important performance measure is one’s GPA, which directly affects one’s grade progression. In particular, the Ministry of Education provides guidelines for grade retention, where a student is to be retained if his GPA or attendance falls below certain cutoffs. These rules are not always respected and schools have certain flexibility in implementing them (Díaz et al. 2021).

3 Data

Our data cover a cohort of teenagers from the 34 largest Chilean municipalities, whom we follow from the year they entered primary schools (2002) until 2013. These 34 municipalities include *all* the big cities in Chile, accommodating about 50% of the Chilean population. Of all juvenile crimes recorded annually in Chile, over 60% occur in these 34 municipalities.¹⁴ We construct our sample by linking four data sets at the individual level and one at the school level, supplemented with aggregate information at the municipality level.

¹²Many schools, especially non-public schools, required parents to report their income and present their marriage certificates; some schools even conduct parental interviews. See Contreras et al. (2010) for an analysis of the selection process. In 2015, the Chilean Parliament passed a law to centralize the admission procedure for all public and voucher-private schools, which eliminated selection based on family background. This was gradually implemented from 2016 (after our sample period).

¹³*Reviews of National Policies for Education: Chile 2004*, OECD, Page 57. For studies on the education market and SES segregation in Chile, see, for example, Valenzuela et al. (2013).

¹⁴Table B1 in Online Appendix compares the characteristics of these 34 municipalities with the rest of Chile.

Three of our data sets are administrative records from the Ministry of Education of Chile. The first data set contains each student’s Grades 1-12 annual records of attendance, GPA, and grade retention, the ID of one’s school, and some basic demographic information (e.g., gender). The second data set contains individual-level records of standardized test (SIMCE) scores in Grade 4 and Grade 10. We use Grade 4 (Grade 10) SIMCE scores to standardize GPAs across primary schools (high schools), and use the standardized GPAs throughout our analysis. The third data set contains school-level information of tuition and school characteristics, including the school’s social economic status (SES) as classified by the Ministry of Education.

The fourth data set is from the Defensoria Penal Publica (DPP), which contains administrative criminal records of almost all arrested youths between 2006 and 2013.¹⁵ For each arrest, we observe the time of the accusation and the sentence received. We link schooling records and criminal records by individual ID.

As explained in Section 2, we focus on boys whose primary schools were either public or voucher-private (47,665 teenage boys). We supplement administrative data with rich information on these teenagers’ family backgrounds from a parent survey conducted at the time of the SIMCE tests. Of the 47,665 boys, the parents of 45,130 boys filled out the survey. Excluding observations missing critical information, such as parental education and family income, our final sample contains 34,783 teenage boys, whose trajectories between age 11 and age 18 are the focus of our study.

For municipality characteristics, we obtain information on the average household income from the National Socioeconomic Characterization Survey (CASEN), and information on aggregate crime rates and arrest rates from the Ministry of the Interior and Public Security.

3.1 Summary Statistics

In this section, we present data patterns that inform us of the important factors to include in our model. The first key factor in our model is teenagers’ heterogeneous observed and unobserved endowments. Table 1 provides a look into the observable part, i.e., family and school backgrounds. The upper panel shows school-level statistics by school type, which take all students (boys and girls) into account. The first row shows that, relative to voucher-private schools, public schools have lower average student test scores (Grade 4 SIMCE). The second row shows that public schools are basically free to attend, while the average annual tuition

¹⁵Fewer than 2% of cases were handled by private attorneys, about which we have no information.

is over 105,800 pesos in voucher-private schools (throughout this paper, tuition refers to the out-of-pocket cost for households net of government vouchers). The next two rows show that, based on the classification by the Ministry of Education, over 54% of public schools are low-SES and only 5% are high-SES; in contrast, only 19% of voucher-private schools are low-SES and 44% are high-SES.

Table 1: School and Individual Characteristics

Primary School	Public		Voucher-Private		
School Ave. SIMCE Score	-0.32 (0.69)		0.32 (0.91)		
Annual Tuition (1,000 pesos)	0.6 (7.4)		105.8 (123.1)		
School SES: Low	54.5%		18.9%		
School SES: High	5.4%		44.1%		
Number of Schools	964		1,221		
	Primary School Type		Primary School Ave. SIMCE		
Family Background	Public	Voucher-Private	1st quar	2nd quar	Above median
Monthly Income/person (1,000 pesos)	43.4 (41.5)	78.1 (71.7)	35.3 (31.7)	47.7 (42.5)	84.5 (74.4)
Parental Education: Low	28.8%	13.1%	36.7%	23.2%	9.9%
Parental Education: High	13.4%	32.2%	7.8%	15.1%	36.7%
Social Welfare Enrollee	22.1%	17.6%	24.4%	29.1%	16.8%
Family Size	5.1 (1.8)	4.8 (1.5)	5.3 (1.9)	5.0 (1.7)	4.7 (1.5)
Number of Teenagers	15,009	19,775	8,627	8,708	17,448

Notes: Cross-school (cross-student) standard deviations are in parentheses in the upper (lower) panel.

The lower panel of Table 1 provides some evidence of the inequality across students at the primary school stage. We summarize student-level statistics for boys enrolled in different primary schools, where schools are grouped by public versus voucher-private, and by the ranking of school average Grade 4 SIMCE scores (1st quartile, 2nd quartile and above median). Relative to their counterparts, students attending voucher-private and/or higher-score schools have much higher family income and better educated parents.¹⁶

¹⁶For parental education, we use the mother's education whenever possible, if the mother is not present, we use the father's education. Parental education is defined as high for those with some college education or more, as low for those without any secondary education, as middle otherwise.

The inequality at the primary-school stage persists into the secondary-school stage. To see this, Table 2 shows the primary-to-secondary school transition matrix among secondary-school enrollees, where we have divided primary (secondary) schools into different groups based on school average SIMCE scores in Grade 4 (Grade 10). There exists high persistence in the quality of one’s enrolled school between the two education stages. For example, if the transition were random, which suggests perfectly equal transition opportunities, one would expect 50% of students in each row to attend a high school with above-median school-level SIMCE, however, this fraction is only 25.6% for those attending bottom-quartile primary schools.

Table 2: Primary-to-Secondary School Transition (%)

Primary School-Average SIMCE	Secondary School-Average SIMCE		
	1st quartile	2nd quartile	Above median
1st quartile	45.1	29.3	25.6
2nd quartile	29.5	32.9	37.6
Above median	12.7	18.9	68.3

The difference in endowments is associated with different outcomes across teenagers. Table 3 shows outcomes among teenagers grouped by their parental education, and by the quality of their primary schools as proxied by school average SIMCE. We define a student as being enrolled in year t if his total attendance days in t were at least 50% of school days, and as being unenrolled in t otherwise.¹⁷ Using this definition, Rows 1-3 show the percentages of teenagers belonging to each of the following three enrollment-arrest categories between 2002 and 2013: One was ever arrested; one was not enrolled in at least one year but was never arrested (“not always enrolled, 0 arrest”); and one was enrolled every year and was never arrested (“always enrolled, 0 arrest”). A clear correlation exists between teenagers’ backgrounds and their enrollment-arrest statuses: The fraction of teenagers ever arrested is 1.3% among those with highly-educated parents, versus 5.6% among those whose parents have low education; similar disparity exists between teenagers grouped by the quality of their primary schools. Furthermore, as shown in Rows 4-6, there is a close link between backgrounds and academic achievement. For example, the gap in GPAs is over 0.8 standard deviations between an average student with low-education parents and one with high-education parents.

¹⁷The fraction of teenagers with $\frac{\text{attendance days}}{\text{school days}} \in (0, 0.5)$ is very small, e.g., 0.1% at age 11 and 0.7% at age 18 (see Table A1 in the appendix).

Table 3: Outcomes By Background

	Parental Education			Primary School-Average SIMCE		
	Low	Middle	High	1st quartile	2nd quartile	Above median
Ever arrested %	5.6	3.2	1.3	5.6	3.4	1.9
Not always enrolled, 0 arrest %	22.2	14.9	11.6	21.9	17.2	11.7
Always enrolled, 0 arrest %	72.2	81.9	87.1	72.5	79.4	86.4
GPA (standardized)	-0.39	-0.01	0.43	-0.45	-0.16	0.34
Grade Retention %	7.3	5.2	3.8	6.9	5.8	4.2
Grade Completed by age 18	11.1	11.4	11.5	11.3	11.5	11.6
Number of Teenagers	6,911	19,494	8,379	8,627	8,708	17,448

Finally, relative to the standard Becker (1968) framework, one additional feature of our model is information friction, which is partly motivated by findings from previous literature that students face non-trivial uncertainties about their learning abilities (Altonji 1993, Arcidiacono 2004, Cunha et al. 2005, Stange 2012, Bordon and Fu 2015, and Arcidiacono et al. 2016). While these studies are about college-age students, we expect uncertainties to be relevant for younger individuals as well. To see whether this hypothesis may hold in our data, we conduct the following exercise. First, we regress GPA on individual dummies (DI_{i1}) and grade dummies ($DG_{G_{it}}$):

$$GPA_{it} = DI_{i1} + DG_{G_{it}} + \varepsilon_{it}, \quad (1)$$

where GPA_{it} and G_{it} are i 's standardized GPA and grade level in year t . Net of individual-specific and grade-specific factors, the residual ε_{it} arguably captures the deviation of one's GPAs from the expected level. Then, we regress enrollment status (e_{it}) on individual dummies (DI_{i2}) and the lagged GPA residuals $\{\widehat{\varepsilon}_{it'}\}_{t' < t}$.

$$e_{it} = DI_{i2} + \sum_{n>0} b_n \widehat{\varepsilon}_{it-n} + \iota_{it}, \quad (2)$$

where ι_{it} is an error term. Table 4 shows the results from (2) in three specifications with increasing numbers of lagged GPA residuals. The estimated coefficients on residual scores are significantly positive, which is consistent with information friction and learning: Better GPA shocks in the past improve one's belief about his prospect of schooling and hence increase his enrollment probabilities. Readers should be aware that information friction is just one

possible explanation and that we are abstracting from other potential explanations, which are indistinguishable from information friction using our data.¹⁸

Table 4: Regression of Current Enrollment on Lagged GPA Residuals

	Specification 1		Specification 2		Specification 3	
ε_{it-1}	0.018	(0.001)	0.017	(0.001)	0.020	(0.001)
ε_{it-2}			0.004	(0.001)	0.003	(0.001)
ε_{it-3}					0.008	(0.001)

Notes: Enrollment indicator regressed on individual and grade dummies and lagged GPA residuals. Standard errors are in parentheses.

4 Model

4.1 Primitives

There are L locations (i.e., municipalities in our data). Locations differ in average household income, aggregate crime rates in the population and arrest probabilities. Across L locations, there are in total K primary schools and K' high schools, each characterized by W_k ($k \in K$ or $k \in K'$).¹⁹

Endowment: A teenager i is endowed with a vector of family background X_i (one component of X_i is home location l_i), a primary school k_{i0} , a type $\chi_i \in \{1, 2\}$ and ability a_i . One knows his type χ_i but not a_i . Ability is normally distributed with type-specific means:

$$a_i \sim N(\bar{a}_{\chi_i}, \sigma_a^2). \quad (3)$$

One has initial belief given by (3), and updates it over time as new information (GPA) comes in. The researcher observes neither χ_i nor a_i , both of which capture teenagers' unobservable heterogeneity but in different ways. The variable a_i captures heterogeneous schooling abilities and embodies a major information friction faced by teenagers. The variable χ_i captures heterogeneity known to teenagers. Different types of teenagers differ in the distributions of their abilities and preferences.

¹⁸For example, one alternative explanation is that a teenager who plans to drop out of school in period t starts to reduce study effort in periods $t' < t$.

¹⁹If both levels of education is offered in the same school, the school is counted both as a primary school and as a high school.

Choices: In each period $t = 1, \dots, T$, except when one is in jail, a teenager decides whether or not to enroll in school (e_{it}) and whether or not to participate in criminal activities (d_{it}): $(e_{it}, d_{it}) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

If $e_{it} = 1$, one's academic performance is observed at the end of t .

If $d_{it} = 1$, with a location-specific probability, one is arrested. An arrested individual may be sentenced to serve $\tau \geq 0$ periods in jail, where the sentence length τ is a stochastic function of one's previous criminal record and one's age.

Because all individuals in our sample were consistently enrolled with no criminal record between Grades 1 and 4, we start the model at age 11, so that $t = 1, \dots, T$ corresponds to ages 11 to 18, and the initial schooling G_{i0} is 4 for all i .

4.1.1 GPA and Grade Progression

We model GPA_{it} as a stochastic function of family characteristics X_i , school characteristics W_{kit} , the interaction between X_i and W_{kit} , whether or not one is repeating the current grade, and ability a_i , given by:²⁰

$$\begin{aligned} GPA_{it} &= \gamma_{0G_{it}} + X_i\gamma_1 + W_{kit}\gamma_2 + X_iW_{kit}\gamma_3 + \gamma_4(1 - g_{it}) + a_i + \epsilon_{it}^{gpa} \\ &\equiv \overline{GPA}_{it} + a_i + \epsilon_{it}^{gpa}, \end{aligned} \quad (4)$$

where G_{it} is i 's grade level at time t , and $\gamma_{0G_{it}}$ is a grade-specific constant. $g_{it} = 1$ if i successfully progressed to the current grade from the last period, so γ_4 is the effect of repeating a grade. Because ability is not directly observed, we pin its unit to that of GPA and normalize the coefficient in front of a_i to 1. Finally, ϵ_{it}^{gpa} is an i.i.d. shock drawn from $N(0, \sigma_{\epsilon^{gpa}}^2)$. A student cannot separately observe a_i and ϵ_{it}^{gpa} . Let \overline{GPA}_{it} be the part of GPA net of $a_i + \epsilon_{it}^{gpa}$.

As described in Section 2.2, grade retention is largely determined by GPA, but schools have some flexibility in implementing the Ministry's guideline. As such, we model grade progression as a stochastic function of school characteristics and the student's GPA and type, subject to a normally distributed shock $\epsilon^g \sim N(0, \sigma_{\epsilon^g}^2)$ that captures events unforeseen by the teenager. In particular, the probability that a student with GPA_{it} progresses to the next grade in $t + 1$ is

²⁰In the final specification, we include only the interaction between parental education and school quality; adding other interactions do not meaningfully improve the likelihood. Therefore, we choose the simpler specification.

given by

$$\Pr(g_{it+1} = 1 | W_{kit}, GPA_{it}, \chi_i) = \Phi\left(\frac{GPA_{it} + \theta_{0\chi_i} + W_{kit}\theta_1}{\sigma_{\epsilon g}}\right). \quad (5)$$

Importantly, we introduce type-specific parameters $\theta_{0\chi_i}$ to account for factors affecting grade progression that are known to the teenager and the school but not to the researcher.

Two points are worth mentioning. First, besides providing information about a_i , GPA is directly valuable to a teenager: Together with one's completed grade (G_{it}), GPA measures one's academic achievement. As we will show in the teenager's problem, academic achievement contributes to one's contemporaneous utility and future value at the end of T . Second, we assume that, conditional on GPA and type, grade progression does not depend on a_i and hence does not provide the teenager with further information about a_i .²¹ This is plausible if, like the teenager, the school does not know a_i either. A similar assumption is made for primary-secondary school transition in Section 4.1.3.

4.1.2 Beliefs

The teenager does not know his ability a_i , but knows that $a_i \sim N(\bar{a}_{\chi_i}, \sigma_a^2)$. He then uses GPA information to update his belief. Let (EA_{it}, VA_{it}) be i 's updated beliefs about the mean and the variance of his (normally distributed) ability at the end of t , they are given by

$$(EA_{it}, VA_{it}) = \begin{cases} (EA_{it-1}, VA_{it-1}) & \text{if } GPA_{it} \text{ is not available,} \\ \left(EA_{it-1} \frac{\sigma_{\epsilon gpa}^2}{\sigma_{\epsilon gpa}^2 + VA_{it-1}} + (GPA_{it} - \overline{GPA}_{it}) \frac{VA_{it-1}}{\sigma_{\epsilon gpa}^2 + VA_{it-1}}, \frac{\sigma_{\epsilon gpa}^2 VA_{it-1}}{\sigma_{\epsilon gpa}^2 + VA_{it-1}} \right) & \text{otherwise.} \end{cases} \quad (6)$$

When GPA_{it} is not available (e.g., if $e_{it} = 0$), one's belief keeps unchanged. Otherwise, one uses the Bayes rule to update his belief. The speed at which one updates his belief is governed by the relative dispersion of GPA shocks ($\sigma_{\epsilon gpa}$) versus that of the ability distribution (σ_a). Because of the randomness in GPA realizations, ex ante identical teenagers may have different beliefs about themselves and choose different paths accordingly.

Our model allows students to face uncertainty about their schooling ability. Ideally, one would allow teenagers to also learn about their (heterogeneous) criminal abilities. However, our data contain no information about *uncaught* crimes *at the individual level*. As such, a model

²¹A similar assumption is made in Arcidiacono et al. (2016), where one updates beliefs based on the continuous GPA variable but not the binary (non)graduation variable. For studies on the link between grade retention and crime, see, for example, Eren et al. (2017) and Eren et al. (2018).

with heterogeneous criminal abilities (e.g., in avoiding arrests or making criminal earnings) is fundamentally unidentifiable with our data.

4.1.3 Transition to High School

Between Grades 8 and 9, a transition happens between primary school and high school. A student can always get a seat in the high school system; yet, whether or not one will attend high school is his choice. Let the probability that i can transit to school k' be $p_{k'i}^{tr}$; we model $p_{k'i}^{tr}$ as dependent on X_i , χ_i , one's Grade 8 GPA (GPA_{i4}), the characteristics of one's primary school W_{ki0} and the destination school $W_{k'}$, such that

$$p_{k'i}^{tr} = P^{tr}(X_i, \chi_i, GPA_{i4}, W_{ki0}, W_{k'}) = \begin{cases} \frac{\exp(f(X_i, \chi_i, GPA_{i4}, W_{ki0}, W_{k'}))}{\sum_{k' \in K(l_i)} \exp(f(X_i, \chi_i, GPA_{i4}, W_{ki0}, W_{k'}))} & \text{if } k' \in K'(l_i) \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where $K'(l_i)$ is the set of high schools that are *feasible* for a student living in location l_i (l_i is one element of X_i);²² $f(\cdot)$ is a linear function of student, primary school and secondary school characteristics and their interactions. The direct effect of X_i and GPA_{i4} on p^{tr} reflect the fact that schools can select based on family background and GPA. To allow for non-random matching for factors known to the school and the student but not to the researcher, we introduce χ_i into $p^{tr}(\cdot)$. One's primary school affects transition probabilities via two channels. First, it affects the transition indirectly via its effect on achievement (GPA). Second, it can also affect the transition directly via W_{ki0} and the interaction between W_{ki0} and $W_{k'}$. For example, transition to higher-quality high schools may be easier if the quality of one's primary school is also high. Through either effect, inequality at the primary school stage will lead to further inequality at the high school stage.

Without data on applications and admissions, we model the transition to high school in a reduced form way, which is a limitation. In our effort to address some of the selection issues, we include χ_i , which is known to the teenager but unobservable to the researcher, in the stochastic transition function, and we estimate this transition process *within* the model.

²²Empirically, we define $K'(l_i)$ as the collection of all high schools with at least one attendee from location l_i , which includes all high schools in l_i and some high schools in other, mostly nearby, locations.

4.1.4 Timing

For each $t = 1, \dots, T$ (ages 10-18), the within-period timing of events is as follows:

1. Choice-specific payoff shocks $v_{it} = \{v_{it}^{ed}\}_{(e,d) \in \{0,1\}^2}$ are realized, which are i.i.d. Type-1 extreme-value distributed.
2. A teenager chooses $(e_{it}, d_{it}) \in \{(0,0), (0,1), (1,0), (1,1)\}$.
3. If $e_{it} = 1$, one's GPA_{it} and grade progression (g_{it+1}) are realized. One's belief about a_i is updated. If $d_{it} = 1$, one's arrest/non-arrest result is realized. If arrested, a jail sentence ($\tau_{it} \geq 0$) is prescribed, following a stochastic function of one's age and criminal record.

4.1.5 State Variables

The vector of state variables at time t , Ω_{it} , contains the following:

1. Permanent variables: χ_i, X_i .
2. Dynamic variables: $EA_{it-1}, VA_{it-1}, J_{it-1}, G_{it-1}, g_{it}, e_{it-1}, k_{it}$.
 - 1) EA_{it-1}, VA_{it-1} : i 's belief about his ability, evolving according to (6).
 - 2) $J_{it-1} = [J_{it-1,1}, J_{it-1,2}]$. $J_{it-1,1}$: the total number of past arrests. $J_{it-1,2}$: the total length of sentences received in the past.
 - 3) G_{it-1} : the highest grade completed by end of $t - 1$.
 - 4) g_{it} : whether or not one progressed one grade between $t - 1$ and t .
 - 5) e_{it-1} : whether or not one was enrolled last period.
 - 6) k_{it} : the ID of the one's most recent school, which is k_{i0} if $G_{it-1} < 9$ and the ID of one's high school if $G_{it-1} \geq 9$.²³
3. Transitory shocks to choice-specific payoffs: $v_{it} = \{v_{it}^{ed}\}$

The evolution from Ω_{it} to Ω_{it+1} depends on not only one's choices (e_{it}, d_{it}) but also a vector Υ_{it} of factors realized at the *end* of t . The vector Υ_{it} consists of a) GPA_{it} (if $e_{it} = 1$), which affects belief updating (EA_{it}, VA_{it}); b) the grade progression shock (if $e_{it} = 1$), which, together with GPA_{it} , affects g_{it+1} and G_{it} ; c) the realization of arrest and sentencing outcomes (if $d_{it} = 1$), which affect J_{it} . To save on notation, we will write Ω_{it+1} instead of $(\Omega_{it+1} | \Omega_{it}, e_{it}, d_{it}, \Upsilon_{it})$.²⁴

²³Within primary school or high school stage, school transfers are rare. We assume that transfers only happen between the two stages. If a student transferred schools between Grades 5 and 8 (Grades 9 and 12), we use the first primary school (high school) as his school.

²⁴The state space is large because the dynamic beliefs (EA_{it-1}, VA_{it-1}) are continuous variables. We solve the model via backward induction with Monte Carlo integration and interpolation (Keane and Wolpin 1994); details are in the Online Appendix.

4.2 Teenager's Problem

At $t \leq T$, unless he is in jail and hence unable to make choices, a teenager's problem is

$$V_t(\Omega_{it}) = \max_{(e,d) \in \{(0,0), (0,1), (1,0), (1,1)\}} \{V_t^{ed}(\Omega_{it})\},$$

where $V_t^{ed}(\cdot)$ denotes the choice-specific value function. We specify $V_t^{ed}(\cdot)$ for regular cases first, and then for the case of the last grade of primary education, where, if one chooses $e_{it} = 1$, he may face the transition to secondary education at the end of the period.

4.2.1 Value Functions: Regular Cases

e = 0, d = 0 :

$$V_t^{00}(\Omega_{it}) = v_{it}^{00} + \beta E_v V_{t+1}(\Omega_{it+1}), \quad (8)$$

where the mean utility of being inactive in both activities is normalized to 0, v_{it}^{00} is the payoff shock and β is the discount factor. The expectation for the next period value is taken over payoff shocks $\{v_{it+1}^{ed}\}$.

e = 1, d = 0 : Before describing the value function, we introduce the contemporaneous utility from attending school, which is given by

$$U^{sch}(GPA_{it}, X_i, \chi_i, W_{kit}, d_{it}, g_{it}, J_{it-1}). \quad (9)$$

That is, students care about their performance in school (GPA), but schooling utility may vary with one's backgrounds X_i , type χ_i and school characteristics W_{kit} . Criminal participation d_{it} may also affect schooling utility if there is a utility cost of engaging in both activities. In addition, to capture potential scar effects, we allow schooling utility to depend on whether or not one is repeating a grade ($1 - g_{it}$), and on one's criminal record (J_{it-1}).

Let $F_{it}^s(z)$ and $E_{it}^s(z)$ be the CDF and the expected value of some variable z , viewed by teenager i at the beginning of t , given his subjective belief (EA_{it-1}, VA_{it-1}) . The value of choosing (1, 0) is given by

$$V_t^{10}(\Omega_{it}) = \left(v_{it}^{10} - \varphi(1 - e_{it-1}) + \int \left(U^{sch}(GPA_{it}, X_i, \chi_i, W_{kit}, d_{it}, g_{it}, J_{it-1}) + \beta \sum_{m=0}^1 \Pr(g_{it+1} = m | W_{kit}, GPA_{it}, \chi_i) E_v V_{t+1}(\Omega_{it+1}) \right) dF_{it}^s(GPA_{it}) \right). \quad (10)$$

The parameter φ , if positive, captures inertia effects or psychic costs for a non-enrollee to return to school. At the beginning of period t , one has to form expectation about his outcome GPA_{it} (hence the integral over $F_{it}^s(GPA_{it})$), which affects one's schooling utility $U^{sch}(\cdot)$ and the probability that he will proceed to the next grade $\Pr(g_{it+1} = 1 | W_{kit}, GPA_{it}, \chi_i)$.

e = 0, d = 1 :

$$V_t^{01}(\Omega_{it}) = \left(v_{it}^{01} + r_{\chi_i} + (1 - \rho_{l_i}) [R(l_i, X_i) + \beta E_v V_{t+1}(\Omega_{it+1})] + \rho_{l_i} E_{\tau | (J_{it-1}, t)} [u^j(J_{it-1}) + \beta^{\tau+1} E_v V_{t+\tau+1}(\Omega_{it+\tau+1})] \right). \quad (11)$$

If one engages in criminal activities ($d = 1$), he receives a type-specific utility for criminal activities (r_{χ_i}). With a location-specific probability ($1 - \rho_{l_i}$), one is not arrested, in which case he enjoys the payoff $R(l_i, X_i)$, and proceeds to the next period without additional criminal record. With probability ρ_{l_i} , one is arrested and serves a jail time ($\tau \geq 0$) drawn from a distribution that shifts with one's criminal record J_{it-1} and age t (detailed in Appendix A1.4). In this case, he receives the utility of being arrested $u^j(J_{it-1})$. While in jail, one is not allowed to make decisions until $t + \tau + 1$ when one is out of jail and proceeds with an updated criminal record.²⁵

To be more specific, the payoff from uncaught crimes is a function of local crime rate (cr_l), local average income (INC_l), and one's family income (inc_i), given by²⁶

$$R(l_i, X_i) = \omega_1 cr_{l_i} + INC_{l_i} (\omega_2 + \omega_3 I(inc_i = low) + \omega_4 I(inc_i = middle)). \quad (12)$$

²⁵The sentence τ may not be an integer (year), in which case, we round τ to count the length of the inaction period.

²⁶Our model is silent about why criminal payoffs may vary with aggregate conditions such as crime rate and average income, which does not affect our ability to study *individual* choices and how an individual would respond to counterfactual policies. See Fu and Wolpin (2018) as a recent example of studies on crime in an equilibrium setting.

If arrested, one's utility is given by

$$u^j(J_{it-1}) = \mu_1 + \mu_2 I(J_{it-1} = 0), \quad (13)$$

where μ_2 is the additional disutility for first-time arrests.²⁷

$\mathbf{e} = \mathbf{1}, \mathbf{d} = \mathbf{1}$:

$$V_t^{11}(\Omega_{it}) = \left(\begin{array}{l} \int U^{sch}(GPA_{it}, X_i, \chi_i, W_{kit}, d_{it}, g_{it}, J_{it-1}) dF_{it}^s(GPA_{it}) + v_{it}^{11} + r_{\chi_i} - \varphi(1 - e_{it-1}) + \\ (1 - \rho_{l_i}) \left[\begin{array}{l} R(l_i, X_i) + \\ \int (\sum_{m=0}^1 \Pr(g_{it+1} = m | W_{kit}, GPA_{it}, \chi_i) \beta E_v V_{t+1}(\Omega_{it+1})) dF_{it}^s(GPA_{it}) \end{array} \right] \\ + \rho_{l_i} E_{\tau|(J_{it-1}, t)} [u^j(J_{it-1}) + \beta^{\tau+1} E_v V_{t+\tau+1}(\Omega_{it+\tau+1})] \end{array} \right). \quad (14)$$

When engaging in both activities, one enjoys the expected contemporaneous utility from both (Line 1 of (14)). If not arrested, one enjoys the criminal payoff $R(\cdot)$, observes g_{it+1} , and continues to $t + 1$ with updated beliefs (Line 2). The last line of (14) describes the case when one gets arrested.

4.2.2 Value Functions: the Last Grade of Primary Education

When the state variables are such that $G_{it-1} = 7$ and $g_{it} = 1$, the teenager is allowed to progress to Grade 8. In this case, the value functions $V_t^{10}(\cdot)$ and $V_t^{11}(\cdot)$ need to be modified: At the end of t , if one can proceed to the next grade ($g_{it+1} = 1$), he will face the primary-secondary school transition, which involves an additional layer of expectation over the probabilities of transferring to different high schools. In particular, $V_t^{10}(\cdot)$ and $V_t^{11}(\cdot)$ in this special period differ from the regular value functions in that $\sum_{m=0}^1 \Pr(g_{it+1} = m | W_{kit}, GPA_{it}, \chi_i) \beta E_v V_{t+1}(\Omega_{it+1})$ in Equations (10) and (14) is replaced by

$$\Pr(g_{it+1} = 0 | W_{kit}, GPA_{it}, \chi_i) \beta E_v V_{t+1}(\Omega_{it+1}) + \Pr(g_{it+1} = 1 | W_{kit}, GPA_{it}, \chi_i) \sum_{k'} p_{k'i}^{tr} E_v V_{t+1}(\tilde{\Omega}_{it+1}, k'),$$

where $\tilde{\Omega}_{it+1}$ is the vector of state variables excluding the high school ID.

²⁷In an alternative specification, we allow utility $u^j(\cdot)$ to depend not only on J_{t-1} , but also on jail time. The added parameter is estimated to be essentially zero, we therefore choose the simpler specification.

4.2.3 Terminal Value

One’s terminal value is a function of his characteristics (X_i, χ_i) , his education achievement G_{iT} , his criminal records J_{iT} , and his up-to-date belief about his own ability, given by

$$V_{T+1}(X_i, \chi_i, G_{iT}, J_{iT}, E_{iT}(a_i)). \quad (15)$$

Given that we can only follow the teenagers up to age 18, we use the reduced-form $V_{T+1}(\cdot)$ function to close the model at age 18 (T). The age-18 outcome variables included in $V_{T+1}(\cdot)$, such as education achievement and criminal records, are arguably highly relevant for one’s future outcomes. Moreover, we also allow for interactions between one’s characteristics and outcomes in this terminal value function. For example, a high school degree provides one with the option to enroll in college. However, this option value will be lower for a teenager if, for example, he has low tastes for schooling (as captured by one’s type), believes he has low schooling ability or his family cannot afford college education.

4.3 Information on one’s Endowment

Table 5: Information on one’s Endowment

Obs. to both	Obs. to neither	Obs. to teenager	Obs. to researcher
$X_i, k_{i0}, \{GPA_{it}\}_{t=-3}^0$	a_i	χ_i	s_i

Table 5 shows the information structure of the teenager’s endowment. Besides X_i and one’s primary school ID k_{i0} , both the researcher and the teenager also observe the teenager’s GPA ($\{GPA_{it}\}_{t=-3}^0$) from Grade 1 to Grade 4, before the model starts. Neither the teenager nor the researcher observes a_i . The teenager knows his χ_i , which is not known to the researcher. In contrast, as mentioned in Section 2.2, the Grade-4 SIMCE score s_i is observed by the researcher but not by the teenager. We make further specifications as follows.

Beliefs (EA_{i0}, VA_{i0}) : By $t = 1$, the teenager’s should have already used $\{GPA_{it}\}_{t=-3}^0$ to update his initial belief. Therefore, we model (EA_{i0}, VA_{i0}) as the Bayes output starting from $a_i \sim N(\bar{\alpha}_{\chi_i}, \sigma_a^2)$ and updated with $\{GPA_{it}\}_{t=-3}^0$ following (6).

The unobservable (χ_i, a_i) : We model the probability that $\chi_i = 1$ as a logistic function of one’s observable endowment X_i and $W_{k_{i0}}$, where $W_{k_{i0}}$ is included to capture potential non-random matching between students and primary schools arising from factors unknown to the

researcher. Given that $a_i \sim N(\bar{a}_{\chi_i}, \sigma_a^2)$, (χ_i, a_i) is distributed as

$$f(\chi_i, a_i | X_i, W_{k_{i0}}) = \Pr(\chi_i | X_i, W_{k_{i0}}) \phi\left(\frac{a_i - \bar{a}_{\chi_i}}{\sigma_a}\right), \quad (16)$$

with $\Pr(\chi_i = 1 | X_i, W_{k_{i0}}) = \Psi(X_i, W_{k_{i0}})$.

Score s_i (Grade 4 SIMCE), being unobservable to the teenager, does not enter the dynamic decision model; rather, it serves as an additional source of information for the researcher. We model it as

$$s_i = \delta_0 + X_i \gamma_1 + W_{k_{i0}} \gamma_2 + X_i W_{k_{i0}} \gamma_3 + \delta_1 a_i + \xi_i, \quad (17)$$

where $\xi_i \sim N(0, \sigma_\xi^2)$ is a random noise. Equations (4) and (17) share the same vector γ , i.e., the effects of family and school characteristics on GPA and on SIMCE are assumed to be the same, which will be useful for identification (Section 5.1.1).

5 Estimation and Identification

We assume that the sentencing length τ is drawn from an ordered probit distribution that may shift with one's age and criminal records (arrests and jail time), and estimate this distribution outside of the model (Appendix A1.6). For identification reasons (Section 5.1.2), we assume all teenagers in the same municipality face the same arrest probability, calculated as the ratio of the aggregate arrests over crimes in that municipality. We also pre-set the annual discount factor β at 0.9.²⁸

We estimate all the other model parameters using the maximum likelihood method. The vector Θ to be estimated includes parameters governing preferences, the production of GPA and SIMCE, grade progression, the distribution of type and ability, and primary-to-secondary school transition probabilities. Teenager i 's contribution to the likelihood is the sum of his type-specific likelihood, weighted by his type probabilities, i.e.,

$$L_i(\Theta) = \sum_{m=1}^2 \Pr(\chi_i = m | X_i, W_{k_{i0}}; \Theta) L_{im}(\Theta),$$

where L_{im} is the likelihood conditional on i being type m . As detailed in Appendix C, L_{im} takes into account i 's history of choices, academic performance, high school transition, as well

²⁸The monthly discount factor is estimated to be 0.99 (i.e., 0.89 annually) in Imai and Krishna (2004).

as one's Grade 1 to Grade 4 GPAs and Grade 4 SIMCE score. The overall log likelihood is $\mathcal{L}(\Theta) = \sum_{i=1}^N \ln L_i(\Theta)$.

5.1 Identification

We focus our discussion on the complications arising from the two major components in our model that are absent in a bare-bone dynamic discrete choice model: unobserved types and learning about one's own ability.

5.1.1 The GPA Production Function

The identification of the GPA production function (4) involves two complications: C1: GPA is observed conditional on enrollment, a classical selection problem; and C2: the unobservable a_i is correlated with observable inputs in the production of GPA via one's type χ_i .

Without C2, the researcher would have all the information the teenager has about a_i (Table 5), and hence knows his belief. The GPA production function would be identified because the enrollment decision is based on *beliefs* about a_i rather than a_i directly. In particular, without C2, the expected ability based on one's belief, $E_{it}^s(a_i)$, is a weighted sum of all the past GPAs, and it can serve as a control function in estimating the GPA function to deal with C1 (Arcidiacono et al. 2016).

With C2, the control function approach becomes insufficient because $E_{it}^s(a_i)$ contains information known to the teenager but not to the researcher (χ_i). However, the identification is facilitated by the additional information of one's Grade 1 to Grade 4 GPAs and s_i , which we observe for *all* teenagers since all of them were enrolled in Grades 1-4. In particular,

$$\begin{aligned}
& E[GPA_{it}|X_i, W_{kit}, W_{k_{i0}}, \{GPA_{it}\}_{t=-3}^0] \\
&= \gamma_0 G_{it} + X_i \gamma_1 + W_{kit} \gamma_2 + X_i W_{kit} \gamma_3 + \gamma_4 (1 - g_{it}) + E[a_i | X_i, W_{k_{i0}}, \{GPA_{it}\}_{t=-3}^0] \quad (18) \\
&= \left(\gamma_0 G_{it} - \frac{\delta_0}{\delta_1} \right) + \left(1 - \frac{1}{\delta_1} \right) X_i \gamma_1 + \left(W_{kit} - \frac{W_{ik_0}}{\delta_1} \right) \gamma_2 + \left(X_i W_{kit} - \frac{X_i W_{k_{i0}}}{\delta_1} \right) \gamma_3 \\
&\quad + \gamma_4 (1 - g_t) + \frac{1}{\delta_1} E[s | X_i, W_{k_{i0}}, \{GPA_{it}\}_{t=-3}^0].
\end{aligned}$$

The last equality holds because

$$E[s | X, W_{k_0}, \{GPA_{it}\}_{t=-3}^0] = \delta_0 + X_i \gamma_1 + W_{ik_0} \gamma_2 + X_i W_{k_{i0}} \gamma_3 + \delta_1 E[a | X_i, W_{k_{i0}}, \{GPA_{it}\}_{t=-3}^0]. \quad (19)$$

Although all of the structural parameters are identified jointly, Line 2 in Equation (18) implies that including the additional information on $\{GPA_{it}\}_{t=-3}^0$ and s_i in the likelihood function greatly facilitates the identification of δ_1 and $\{\gamma_n\}_{n=1}^4$. The constants $\{\gamma_{0G_t}\}$ and δ_0 are not all identified, we therefore normalize γ_{0G_t} in one grade to 0.

5.1.2 Type Distribution

One major source for identifying the distribution of unobserved types is one's academic outcomes. First, since ability distributions are type-specific, conditional on (X, W_k) , the distributions of ability measures (s and GPAs) will shift as the ability distribution shifts with one's type, exhibiting different modes. Given the assumption that type-specific ability distributions and the mean-zero score noise distributions are all uni-modal and symmetric, the observed modes as well as the distribution of (X, W_k) surrounding each mode are informative about type-specific mean ability \bar{a}_χ and the correlation between type and (X, W_k) . Second, conditional on (X, W_k) , cross-individual dispersion of test scores arise from both their differential ability and test score noises. Using within-individual comparison of multiple measures of his ability (s and GPAs), we learn about the dispersion of test score noises. A comparison of cross-individual and within-individual test score dispersions informs us of the dispersion of abilities. The two dispersions are key parameters governing the speed of learning, as in (6). Third, like ability distribution, the probability of being retained also varies with types. As such, some students will be retained more often than others with the same GPA and school type, which gives information about their types.

As in all dynamic-choice models, another major source of information is one's choices over time. Relative to dynamic models without criminal behaviors, our model is subject to one additional identification complication: Except for studies that use survey data on self-reported criminal activities, uncaught crimes are unobservable at the individual level. This data limitation makes a model with criminal choices unidentifiable if arrest rates are allowed to differ across individuals and/or to change with one's criminal experience. As such, we have to follow the crime literature and impose the strong assumption that individuals within the same municipality l face the same arrest rate ρ_l , which is calculated using aggregate crimes and arrests. With this assumption, individual-level data on (non)arrests gives direct information on the probability that one is involved in crime. As such, the joint distribution of dynamic choices and characteristics (e, d, X, W) becomes available, as is the case in other dynamic models, and hence the usual identification argument applies.

6 Results

6.1 Parameter Estimates

We report a selected set of parameter estimates in this section (standard errors are in parentheses) and the others in Appendix Tables A2-A3.²⁹

Table 6: GPA, Ability and Grade Progression

A. GPA Production: $GPA_{it} = \gamma_0 G_{it} + X_i \gamma_1 + W_{k_{it}} \gamma_2 + X_i W_{k_{it}} \gamma_3 + \gamma_4 (1 - g_{it}) + a_i + \epsilon_{it}^{gpa}$			
X_i :Parent Edu = Low	-0.16 (0.002)	$W_{k_{it}}$: Average SIMCE ^b	0.40 (0.001)
Parent Edu = High	0.11 (0.002)	Ave. SIMCE*(parent edu Low)	-0.07(0.002)
Income = Low ^a	-0.20 (0.002)	Ave. SIMCE*(parent edu high)	-0.01 (0.002)
Income = Middle	-0.10 (0.002)	Public	-0.03 (0.001)
Welfare Enrollee	-0.04 (0.002)	School SES = Low	0.07 (0.003)
Attended Pre-school	0.08 (0.002)	School SES = Middle	0.03 (0.002)
$(1 - g_{it})$:Retained	-0.15 (0.01)	$\sigma_{\epsilon^{gpa}}$	0.63 (0.01)
B. Ability Distribution $a_i \sim N(\bar{a}_{X_i}, \sigma_a^2)$			
\bar{a}_1 :Mean ability (Type1)	0.56 (0.01)	σ_a	0.25 (0.02)
\bar{a}_2 :Mean ability (Type2)	-0.50 (0.01)		
C. Grade Progression $\Pr(g_{it+1} = 1 \cdot) = \Phi\left(\frac{GPA_{it} + \theta_{0X_i} + W_{k_{it}} \theta_1}{\sigma_{\epsilon^g}}\right)$			
θ_{01} :Type 1	0.40 (0.01)	$W_{k_{it}}$:High school	-0.02 (0.01)
θ_{02} :Type 2	0.50 (0.01)	Public primary	0.04 (0.01)
σ_{ϵ^g}	0.59 (0.003)	Public High School	-0.29 (0.01)

^aFamily income levels (low, middle, high) are defined by income terciles.

^bSchool average SIMCE score in the 4th (10th) grade for primary school (high school).

Panel A of Table 6 shows the estimated parameters governing the GPA production function, where ability (a_i) enters with a normalized coefficient of 1. GPA increases with parental education and family income; it is lower for children whose family is on welfare and higher for pre-school enrollees. Grade retention is found to negatively impact GPA. School quality (as measured by school average SIMCE scores) also matters for one's achievement but less so for students with low parental education. In addition, we find that all else being equal, public schools are less effective in producing GPA. Panel B shows that the mean ability among Type 1 individuals is much higher than that among Type 2 individuals. Panel C shows that compared

²⁹To derive the standard errors, we numerically calculate the Hessian of the log likelihood.

to Type 1 individuals enrolled in the same type of schools, Type 2's are slightly more likely to progress to the next grade *if they achieve* the same GPA.

Panel A of Table 7 shows teenagers' in-school utility parameters. A teenager does care about his performance in school: Schooling utility drops significantly if one has a low GPA (below the 25th percentile) and if one is repeating a grade. Everything else being equal, Type 2 individuals, who have lower ability on average, enjoy schools slightly more than their Type 1 counterpart. That is, if Type 2's enjoy schools less, it is due to their lower ability and hence poorer GPA, rather than pure tastes. The next three rows show that the same tuition imposes a larger burden on the lowest-income group, relative to the other income groups. We also find that private schools, high-SES schools and higher-quality schools are more enjoyable to attend.³⁰ Finally, we find that returning to school is quite costly if one was not enrolled in the previous period. This is one of channels via which a temporary bad shock that led one off the track in the past can leave persistent effects.

Panel B shows that Type 2's have a higher taste for criminal activities relative to Type 1's, which, combined with their lower learning ability, makes Type 2 teenagers more prone to crimes. Attending schools and committing crimes at the same time is costly, especially during primary school stage. Moreover, consistent with previous studies, as reviewed by Levitt and Lochner (2000), we also find a significant age effect on criminal tendency. The middle part of Panel B shows that the payoff of uncaught crimes does not vary much with local crime rates in the entire population, but it decreases with local average income except for those from low-income families. Finally, a teenager incurs disutility if caught doing crime, but this deterrence effect is concentrated on first-time arrests.

Panel A of Table 8 shows the estimated parameters governing the type distribution. Teenagers with better family backgrounds and higher-quality primary schools are more likely to be Type 1's (with higher average ability).³¹ The latter reflects the initial sorting between households and schools based on unobservables. Using these estimates, Panel B reports the percentage of Type 1 teenagers in our sample. Overall, there are 45.5% Type 1 teenagers. This fraction is about 10 percentage points higher among teenagers with high-education parents than among those with low-education parents. The disparity is even larger across teenagers attending different primary schools grouped by school SES and by school-level SIMCE scores.

³⁰We have estimated other versions with finer categories of school SES and SIMCE scores, which do not improve upon the more parsimonious specification reported in Table 6 .

³¹The likelihood does not improve if we add other variables in the type distribution (e.g., parental education) and the associated parameters are close to zero. We therefore choose the simpler specification.

Table 7: Preference Parameters

A. Schooling Utility ($e_t = 1$)		B. Crime Utility ($d_t = 1$)	
GPA	0.002 (0.05)	I(type=1)	1.35 (0.45)
I(GPA = low) ^a	-1.43 (0.13)	I(type=2)	2.00 (0.44)
(1 - g_{it}) : Retained	-1.00 (0.03)	$e_t \times d_t$:primary school	-4.39 (0.14)
I(type=2)	0.33 (0.05)	$e_t \times d_t$:high school	-1.55 (0.08)
tuition * I(inc=low) ^b	-0.20 (0.03)	I(high school age)	2.18 (0.20)
tuition * I(inc=middle)	-0.09 (0.03)	Payoff if not caught: $R(l_i, X_i)$	
tuition * I(inc=high)	-0.06 (0.03)	Local crime rate	0.03 (0.01)
I(private school)	0.28 (0.02)	Local mean inc	-0.39 (0.07)
I(high school)	-0.83 (0.07)	Local mean inc*I(inc =low)	0.52 (0.05)
I(school SES< high)	-0.53 (0.05)	Local mean inc*I(inc =middle)	0.32 (0.05)
I(school ave. SIMCE low) ^c	-0.21 (0.06)	Utility if arrested: $u^j(J_{it-1})$	
I(criminal record>0)	-0.001 (0.09)	I(arrest)	-0.10 (2.27)
Back to School Cost φ	1.31 (0.04)	I(first time arrest)	-32.3 (1.04)
Constant	4.08 (0.10)		

^aGPA is low if it is below the 25th percentile in the individual GPA distribution.

^bAnnual tuition in 200,000 pesos. Family income groups are defined by income terciles.

^cSchool ave. SIMCE is low if it is below the 25th percentile among all primary/secondary schools.

Table 8 Type Distribution

A. $\Pr(\chi_i = 1 \cdot) = \Psi(X_i, W_{k_{i0}})$		B. % Type 1 in the Sample (Model Predicted)	
Constant	-0.24 (0.06)	Overall	45.5
Welfare Enrollee	-0.08 (0.03)	Parent Edu=low	40.7
Income (1,000 peso/person-month)	0.02 (0.02)	Parent Edu=high	50.6
Private Insurance	0.04 (0.03)	Family Income = low	41.3
Job contract = low	-0.08 (0.04)	Family Income = high	50.6
Job contract = middle	0.07 (0.04)	Primary Sch SES= low	38.6
Attended Preschool	0.002 (0.05)	Primary Sch SES= high	52.4
Big Family	-0.08 (0.02)	Sch Ave. Score: 1st quartile	37.4
Primary School Ave. SIMCE	0.31 (0.02)	Sch Ave. Score: above median	51.2

6.2 Model Fit

Overall, the model fits the data well. Table 9 shows the fit for outcomes of teenagers by parental education and by primary school quality. The model captures the heterogeneous outcomes across these groups reasonably well, although the model underpredicts the cross-group difference in the fraction of the ever arrested.

Table 9: Model Fit: Outcomes by Background

	Data	Model	Data	Model	Data	Model
Parental Education:	Low		Middle		High	
Ever arrested %	5.6	5.2	3.2	3.4	1.3	1.7
Always enrolled, 0 arrest %	72.2	72.0	81.9	81.1	87.1	87.6
GPA (standardized)	-0.39	-0.43	-0.01	-0.01	0.43	0.45
Retention %	7.3	7.8	5.2	5.3	3.8	3.9
Grade Completed by T	11.1	11.3	11.4	11.5	11.5	11.7
Primary School-Level Ave. SIMCE Score:	1st quartile		2nd quartile		above median	
Ever arrested %	5.6	4.3	3.4	3.5	1.9	2.8
Always enrolled, 0 arrest %	72.5	73.5	79.4	79.1	86.4	85.3
GPA (standardized)	-0.45	-0.48	-0.16	-0.17	0.34	0.36
Retention %	6.9	7.2	5.8	5.8	4.2	4.4
Grade Completed by T	11.3	11.1	11.5	11.3	11.6	11.5

Table 10 shows the model fit for the age-status profile, where with normal grade progression, $t = 1$ to 4 corresponds to the age when one should be enrolled in Grade 5 to 8, and $t = 5$ to 8 corresponds to high-school age. Consistent with the data, the model predicts that both non-enrollment and arrests are low frequency events, but grow over time. In particular, arrests happen almost only in high school age. Furthermore, using model-simulated data, we run the same regressions (1) and (2) as we did in Section 3.1. Table 11 shows that the model is able to replicate the correlation between enrollment and past GPA residuals. Finally, Table 12 shows the fit for primary-to-secondary school transition. In both the data and the model, transition rates are low between the bottom quartile and the upper half.

Table 10: Model Fit: Status Over Time

%	t	1	2	3	4	5	6	7	8
Not enrolled, not arrested	Data	0.5	0.7	0.8	0.5	1.9	6.0	5.1	9.2
	Model	0.7	0.7	0.8	1.4	3.1	4.1	4.8	7.8
Arrested	Data	0.0	0.0	0.0	0.1	0.4	1.0	1.4	1.3
	Model	0.0	0.0	0.0	0.0	0.9	1.0	1.2	1.0

Table 11: Model Fit: Regression of Enrollment on Lagged GPA Residuals

	Spec 1		Spec 2		Spec 3	
	Data	Model	Data	Model	Data	Model
ε_{it-1}	0.018	0.008	0.017	0.010	0.020	0.012
ε_{it-2}			0.004	0.005	0.003	0.007
ε_{it-3}					0.008	0.003

Notes: Enrollment indicator regressed on individual and grade dummies and lagged GPA residuals.

Table 12: Model Fit: Primary School-High School Transition

(%)	Secondary School-Average SIMCE					
	1st quartile		2nd quartile		Above median	
Primary School-Average SIMCE	Data	Model	Data	Model	Data	Model
1st quartile	45.1	50.3	29.3	27.0	25.6	22.7
2nd quartile	29.5	35.9	32.9	27.7	37.6	36.4
Above median	12.7	14.1	18.9	17.3	68.3	68.6

7 Counterfactual Experiments

We use our estimated model to first understand the importance of various factors in explaining the difference in teenagers' outcomes by their backgrounds. Then, we conduct a set of relatively easy-to-implement counterfactual interventions targeted at disadvantaged teenagers, who are more likely to wander astray.

7.1 The Importance of Various Factors

We compare outcomes of three groups of teenagers in the baseline and in counterfactual scenarios with respect to initial endowment, institutional friction, and information friction. These

three groups $g = 1, 2, 3$ are representative of teenagers with disadvantaged, middle, and advantaged backgrounds, whose (primary school SES, parental education) takes the value of (low, low), (middle, middle), and (high, high), respectively.³² To focus on cross-group differences, we assign all teenagers within each Group $g \in \{1, 2, 3\}$ the same primary school and family characteristics and initial achievement $(W_{k_{g0}}, X_g, \{GPA_{gt}\}_{t=-3}^0)$, set at the within-group sample median. Each Group- g teenager draws his type and ability from the distribution $f(\chi_i, a_i | X_g, W_{k_{g0}})$ as in (16), and each is subject to i.i.d. contemporary shocks.³³

The baseline outcomes, shown in the upper panel of Table 13, exhibit a clear correlation between outcomes and backgrounds: The fraction of ever-arrested teenagers is 4.3% in the low group, 1.8% in the middle group, and below 1% in the high group. Similarly, between the low group and the high group, there is an over 20 percentage-point gap in the fraction of teenagers consistently enrolled with no arrest. To understand the importance of various factors underlying these differences, we contrast the baseline outcomes with those in the following five counterfactual scenarios, where one and only one factor is changed at a time.

F1 (primary school environment): We set $W_{k_{g0}} = W_{k_{30}}$ for all g , starting from $t = 1$. As such, for Groups 1 and 2, we improve one’s learning environment (hence GPA and in-school utility) from Grade 5 to Grade 8 and one’s primary-to-secondary school transfer opportunities, indirectly by affecting one’s GPA, and directly via the role of W_{k_0} in the transfer process.

F2 (transfer opportunities): We change primary-to-secondary school transfer probabilities for Groups 1 and 2 to those faced by their counterparts in Group 3, i.e., a teenager in Groups 1 and 2 would be treated equally as a Group-3 teenager with the same GPA and type.³⁴

F3 (unobservable endowment): We change the distribution of (type, ability, initial belief) in Groups 1 and 2 to that in Group 3.

F4 (family background): We set $X_g = X_3$ for all g .

F5 (information friction): Each teenager knows his ability starting from $t = 1$.

³²All three groups are located in the same (the largest) municipality and therefore face the same local conditions and set of feasible schools.

³³Given that a_i is unknown to the teenager and that the observable endowment is common within each group, teenagers of the same type within a group share the same initial belief about their ability.

³⁴That is, each teenager i is given the transfer probability vector $\{P^{tr}(X_3, \chi_i, GPA_i, W_{k_{30}}, W_{k'})\}_{k'}$, where i ’s own characteristics (type and GPA) remain but family and primary school characteristics are replaced by $(X_3, W_{k_{30}})$ (only in $P^{tr}(\cdot)$).

Table 13. The Importance of Various Factors

Baseline	Ever arrested (%)	Always enroll, 0 arrest (%)	Grade finished by T
Low	4.30	70.28	11.28
Mid	1.84	86.08	11.71
High	0.96	91.60	11.76

	New	$\frac{\text{new-base}}{\text{base}}$	New	$\frac{\text{new-base}}{\text{base}}$	New	$\frac{\text{new-base}}{\text{base}}$
F1: Low	3.90	-9.30%	73.54	4.64%	11.11	-1.48%
Mid	1.84	0.0%	86.74	0.77%	11.65	-0.46%
F2: Low	3.80	-11.63%	79.80	13.55%	11.27	-0.10%
Mid	1.76	-4.35%	91.40	6.18%	11.75	0.32%
F3: Low	3.80	-11.63%	73.62	4.75%	11.42	1.26%
Mid	1.76	-4.35%	87.10	1.18%	11.77	0.48%
F4: Low	1.26	-70.70%	82.02	16.70%	11.60	2.80%
Mid	0.98	-46.74%	87.60	1.77%	11.75	0.31%
F5: Low	4.36	1.40%	70.42	0.20%	11.28	-0.01%
Mid	1.86	1.09%	85.94	-0.16%	11.71	0.002%
High	0.94	-2.08%	91.54	-0.07%	11.76	-0.01%

The lower panel of Table 13 shows outcomes under each of the counterfactual scenarios and percentage changes relative to the baseline.³⁵ Between the two school-related factors, primary-secondary school transfer opportunities (F2) appear to be more important in explaining cross-group differences than Grade 5-8 school environment (F1), especially in the fraction of consistently enrolled teenagers. Interestingly, improving these two factors for Groups 1 and 2 can *reduce* the final grade finished by time T , because a low-performing student is more likely to be retained in private schools than in public schools (Table 6).³⁶ Results under F3 shows that unobservable endowment has similar explanatory power as transfer opportunities (F2) and slightly more explanatory power than primary school environment (F1). Among all five factors, family background (F4) is by far the most important because it plays a role in every part of a teenager's life.³⁷ In particular, when we endow everyone with an advantaged family background, the fraction of ever-arrested teenagers becomes much more similar across the three groups (1.26%, 0.98% and 0.96%). Finally, the last three rows show that giving full information

³⁵For F1–F4, outcomes of the high group are not shown, because they are the same as those in the baseline.

³⁶In the baseline, Groups 1 and 2 attend public primary schools and Group 3 attend private primary schools.

³⁷Changing X will affect the distribution of types, preferences, GPA production, terminal values, and primary-to-secondary school transfer probabilities.

to teenagers has minimal impacts on their outcomes, *on average*. This is not surprising: In the baseline, noisy ability signals have made some teenagers overly optimistic and some others overly pessimistic about their abilities.

7.2 Counterfactual Policy Interventions

Simulations in Table 13 are informative but hard or impossible to implement in practice. In this section, we conduct a set of relatively easy-to-implement counterfactual interventions targeted at disadvantaged teenagers, who are more likely to wander astray.³⁸ These teenagers, accounting for 10% of our sample, are initially enrolled in low-SES primary schools and have low parental education. The average monthly family income per person among the targeted teenagers is 25,460 pesos, as compared to 63,114 pesos among all teenagers in the sample. The upper panel of Table 14 shows that, in the baseline, relative to an average teenager in the sample, a targeted teenager is more likely to have some arrest records (5.38% versus 3.36%) and much less likely to be consistently enrolled (69.31% versus 80.82%).

To make our counterfactual experiments more realistic and policy relevant, we have made three choices. First, rather than imposing restrictions directly on behaviors, the interventions aim at inducing behavioral responses by changing one’s opportunities and/or incentives. For example, even with improved and/or free access to schools, a teenager can still choose not to enroll. Second, the interventions we consider are all mild and relatively easy to implement. For example, the targeted teenagers will be given schooling opportunities faced by those from middle rather than rich family backgrounds. Third, as our model is silent on the formation of a teenager’s initial endowment, we hold fixed a teenager’s ability, type and initial belief about himself and start the intervention at $t = 1$ (Grade 5). Therefore, the effects should be interpreted as those on the cohort of teenagers who have not anticipated these interventions. It should also be noted that, as is true in any individual decision model, results from our experiments abstract from potential equilibrium impacts. Even though the equilibrium impacts may be limited from our experiments, which are targeted at a very small fraction of the population, the policy impact (both benefits and costs) should still be interpreted only at the individual level.

The first two interventions, motivated by F1 and F2 in the previous section, aim at reducing

³⁸Notice that, unlike the study of representative teenagers in the previous section, in this section, each targeted teenager i has his own characteristics $(W_{k_{i0}}, X_i, \{GPA_{it}\}_{t=-3}^0)$ as observed in the data.

institutional frictions that have led to poor schooling opportunities for the targeted teenagers.

Policy 1 (better primary schools): Policy 1 is similar to F1 in the previous section but milder. It improves the initial primary school environment for a targeted teenager i from that of his baseline primary school $W_{k_{i0}}$ to W_{k_m} , starting from $t = 1$, where W_{k_m} refers to a public primary school with median school-average SIMCE scores, high SES and zero tuition. Notice that free public schools similar to or better than k_m already exist, however, they may not be easily accessible for the targeted teenagers due to frictions such as selective admissions. Such frictions also exist in other countries.³⁹

Policy 2 (better primary schools and fairer transfer opportunities): In addition to improving one’s primary school environment to W_{k_m} , Policy 2 further improves primary-to-secondary school transfer opportunities for a targeted teenager to those faced by their counterpart from middle-level family backgrounds (i.e., Policy 2 is similar to F1 plus F2 but milder). That is, a targeted teenager would be treated equally as someone with the same GPA and type but whose family income, parental education, and family SES are all at the middle level.⁴⁰ Notice that Policy 2 only requires a fairer transfer process, rather than (unrealistically) change one’s family background. Affirmative action, where admissions favor disadvantaged students, would be a more progressive but perhaps harder-to-implement treatment.

The first row in the lower panel of Table 14 shows that by improving one’s primary school environment from Grade 5 ($t = 1$), Policy 1 would lead to a 0.23 percentage point (ppt) or 4.4% decrease in the fraction of ever-arrested targeted teenagers. It would also lead to a 5.3 ppt or 7.7% increase in the fraction of the targeted teenagers who stay on the right track, i.e., who are consistently enrolled throughout primary and secondary education with no criminal record. The second row shows that the effect of Policy 2 is slightly larger.⁴¹

Given our finding that low-income families are more sensitive to tuition costs than others (Table 7), our next intervention completely lifts the tuition burden for the targeted teenagers.

³⁹For example, in many countries, including the U.S., although public schools are all free to attend, their quality differs substantially across neighborhoods. With neighborhood-priority rules used in most cities to allocate students to public schools, children living in poor neighborhoods have little chance of getting into good public schools.

⁴⁰Specifically, let X_m denote a vector of middle-level family background. A targeted teenager i is now given the transfer probability vector $\{P^{tr}(X_m, \chi_i, GPA_i, W_{k_m}, W_{k'})\}_{k'}$, where i ’s own characteristics (type and GPA) remain but family characteristics X_i is replaced by X_m (only in $P^{tr}(\cdot)$).

⁴¹Exploiting variation from Chicago public high school admission random lotteries, Cullen et al. (2006) find that winning a lottery to a top-quartile school significantly reduces *self-reported* arrests in high-school years from 8.9% to 2.8%, but winning a lottery to any school has no effect. Although the arrest measures used in their paper differ from those in ours, our results are not at odds with theirs: We find rather mild effects on arrests when targeted teenagers’ schooling opportunities are improved moderately.

Policy 0v (free schools): Policy 0v (v for voucher) gives additional vouchers to the targeted teenagers on top of the vouchers that already exist in the baseline (Section 2.2), so that high schools become totally free for them to attend. The purpose of this exercise is to examine the effect of tuition intervention without changing the primary-school environment for these teenagers. Therefore, additional vouchers are needed only at the high school stage, since these teenagers attend public primary schools that are already free. Policy 0v would reduce the fraction of ever-arrested teenagers by 0.2 ppts or 3.5%; it would increase, by 2.6 ppts or 3.7%, the fraction of those who stay on the right track. The last column of Table 14 shows that Policy 0v would involve, an average (additional) voucher of 25,650 pesos (32 USD) per targeted teenager per high school year. Relative to high school vouchers that already existed in our sample period (440,000 to 650,000 pesos per enrollee-year), the (additional) cost of Policy0v is small but so is its effect.

Table 14. Counterfactual Policy Interventions

Baseline	Ever arrested (%)		Always enroll, 0 arrest (%)		Grade finished by T		
All Teens	3.36		80.82		11.50		
Targeted Teens	5.38		69.31		11.22		
Targeted Teens	New	$\frac{\text{new-base}}{\text{base}}$	New	$\frac{\text{new-base}}{\text{base}}$	New	$\frac{\text{new-base}}{\text{base}}$	Cost*
Policy 1	5.15	-4.4%	74.66	7.7%	11.30	0.6%	0
Policy 2	5.08	-5.6%	75.54	9.0%	11.31	0.8%	0
Policy 0v	5.20	-3.5%	71.88	3.7%	11.26	0.3%	25,650
Policy 1v	4.89	-9.1%	77.47	11.8%	11.33	1.0%	96,877
Policy 2v	4.79	-11.0%	78.47	13.2%	11.34	1.1%	123,739
Policy 3v	4.31	-19.9%	85.10	22.8%	11.34	1.1%	389,004

*Additional high school voucher per targeted teenager per high school year (pesos).

The next three policies improve schooling opportunities and make schools free.

Policies 1v, 2v and 3v (better and free schools): Policy 1v (Policy 2v) is the counterpart of Policy 1 (Policy 2) with the additional feature that high school tuition is fully covered for the targeted teenagers via extra vouchers. Table 14 shows that Policies 1v and 2v are about twice as effective as their counterpart. Under Policy 2v, which is more effective than Policy 1v, the fraction of the ever-arrested decreases by 0.6 ppts or 11%, and the fraction of the consistently-enrolled increases by 9.2 ppts or 13.2%. Policy 1v and Policy 2v would cost 96,877 pesos (123 USD) and 123,739 pesos (157 USD) per targeted teenager per high school

year, respectively. Costs of these two policies are significantly higher than that of Policy 0v but still modest.

Our parameter estimates imply that relative to their public-school counterpart, private schools are more enjoyable to attend and more effective in producing achievement. We therefore implement Policy 3v, which enhances Policy 2v by guaranteeing the targeted teenager a seat in a decent private high school (with full tuition vouchers).⁴² The last row of Table 14 shows that Policy 3v would lead to a 20% reduction in the fraction of ever-arrested teenagers and a 23% increase in the fraction of consistently-enrolled teenagers. That is, Policy 3v is almost twice effective as Policy 2v but also three times as costly.

8 Conclusion

We have developed a model of teenagers' dynamic choices of schooling and crime, incorporating four groups of factors underlying the paths taken by different teenagers, namely endowment, institutional friction, information friction and transitory shocks. We have estimated the model using a rare administrative data set from Chile that links school records and teenage-year criminal records.

We use the estimated model to study a set of moderate counterfactual policy interventions, targeted at teenagers with disadvantaged backgrounds. We find that making schools free alone or improving schooling opportunities alone have limited impacts. Interventions that combine these two types of measures would be twice as effective as either type of measures on their own. The cost of these interventions are relatively low, both financially and in terms of the degree to which the schooling opportunities are improved. Although our data do not allow us to investigate longer-run effects, some studies suggest that these effects can be much larger. For example, using data from North Carolina, Deming (2011) finds that, for high-risk youths, winning the lottery to attend one's first-choice public school had no significant effect on crime reduction between ages 16 and 18 but reduced early-adulthood crimes by 50%.

Given that frictions featured in the Chilean setting exist in many other countries in the same or similar formats (e.g., the unequal access to good schools), lessons learned from our exercise can be useful elsewhere. Our framework can be extended along several dimensions for future research. First, we have taken teenagers' ability and unobserved types as pre-determined

⁴²This moderate private high school has middle SES, and an average SIMCE score of 0.08, which is higher than the average among all high schools (0.04), but lower than the average among voucher-private high schools (0.31).

endowments that are correlated with their family backgrounds. An important extension is to bring into our framework early childhood investment that shapes one’s initial conditions by teenage years.⁴³ This extension would allow for an investigation of how interventions in childhood interact with those in teenage years, but would require additional information on parental investment such as time inputs. A second extension is to open the black box of the “terminal value” function by modeling one’s choices and outcomes after age 18. When these additional data become available, this extension will allow for a broader view on how pre-job market intervention and intervention targeted at low-skilled workers may interact in helping those from disadvantaged backgrounds to move up the ladder.

References

- [1] Altonji, J. (1993) “The Demand for and Return to Education When Education Outcomes Are Uncertain,” *Journal of Labor Economics*, 11(1): 48–83.
- [2] Anderson, M. (2014) “In School and Out of Trouble? The Minimum Dropout Age and Juvenile Crime.” *Review of Economics and Statistics* 96 (2): 318–31.
- [3] Arcidiacono, P. (2004) “Ability Sorting and the Returns to College Major,” *Journal of Econometrics*, 121(1–2): 343–375.
- [4] Arcidiacono, P., E. Aucejo, A. Maurel, T. Ransom (2016). “College Attrition and the Dynamics of Information Revelation?” Working paper, Duke University.
- [5] Aucejo, E. and J. James. (2019) “Catching Up to Girls: Understanding the Gender Imbalance in Educational Attainment Within Race” *Journal of Applied Econometrics*, 34 (4):502–525.
- [6] Becker, G. (1968) “Crime and Punishment: An Economic Approach.” *Journal of Political Economy*, 76(2) 169-217.
- [7] Billings, S, D. Deming, and J. Rockoff (2014) “School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg.” *Quarterly Journal of Economics*, 129 (1): 435-476.
- [8] Blumstein, A., J. Cohen, J. Roth and C. Visher, Editors (1986). *Criminal Careers and “Career Criminals”?* Vols I and II. Washington, DC: National Academy Press.
- [9] Bordon., P and C. Fu (2015). “College-Major Choice to College-Then-Major Choice,” *Review of Economic Studies*, 82 (4): 1247-1288.
- [10] Broidy, L., D. Nagin, R. Tremblay, J. Bates, B. Brame, K. Dodge, D. Fergusson, J. Horwood, R. Loeber, R. Laird, D. Lynam, T. Moffitt, G. Pettit, and F. Vitaro (2003). “Developmental Trajectories of Childhood Disruptive Behaviors and Adolescent Delinquency: A Six-Site, Cross-National Study.” *Developmental Psychology*, 39, 222-245.

⁴³See, for example, Cunha and Heckman (2007), Cunha et al. (2010) and Del Boca et al. (2013) for studies on parental investment on children.

- [11] Case, A. and L. Katz (1991). “The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths,” NBER Working Paper No. 3705.
- [12] Clay, K., J. Lingwall, and M. Stephens (2012). “Do Schooling Laws Matter? Evidence from the Introduction of Compulsory Attendance Laws in the United States.” NBER Working Paper 18477.
- [13] Contreras, D., P. Sepulveda, and S. Bustos (2010). “When Schools Are the Ones that Choose: The Effects of Screening in Chile.” *Social Science Quarterly*, 91(5): 1350-1368.
- [14] Cook, P. and S. Kang (2016). “Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation.” *American Economic Journal: Applied Economics* 8(1): 33–57.
- [15] Cortés, T., N. Grau and J. Rivera (2020). “Juvenile Incarceration and Adult Recidivism,” Working paper, Universidad de Chile.
- [16] Cullen, J., B. Jacob, and S. Levitt (2006), “The effect of school choice on participants: Evidence from randomized lotteries” *Econometrica.*, 74.5: 1191-1230
- [17] Cunha, F. and J. Heckman (2007). “The Technology of Skill Formation,” *The American Economic Review*, 97(2): 31-47.
- [18] Cunha, F., J. Heckman and S. Navarro (2005). “Separating Uncertainty from Heterogeneity in Life Cycle Earnings.” *Oxford Economic Papers*, 57(2): 191-261.
- [19] Cunha, F., J. Heckman, and S. Schennach (2010). “Estimating the technology of cognitive and noncognitive skill formation,” *Econometrica*, 78(3): 883–931.
- [20] Del Boca, D., C. Flinn, and M. Wiswall (2013), “Household choices and child development,” *Review of Economic Studies*, 81(1): 137–185.
- [21] Deming, D (2011), “Better schools, less crime?” *Quarterly Journal of Economics*, 126.4: 2063-2115.
- [22] Di Tella, R. and E. Schargrodsky (2004). “Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack,” *American Economic Review*, 94(1), 115–133.
- [23] Díaz, J., Grau, N., Reyes T., and J. Rivera (2021). “The impact of grade retention on juvenile crime ”. *Economics of Education Review*, Volume 84.
- [24] Donohue, J., and S. Levitt (2001) “The Impact of Legalized Abortion on Crime,” *Quarterly Journal of Economics*, 116 (2), 379–420.
- [25] Eren, O., B. Depew, and S. Barnes (2017) “Test-based promotion policies, dropping out, and juvenile crime,” *Journal of Public Economics*, 153: 9-31.
- [26] Eren, O., M. Lovenheim, and N. Mocan. (2018) “The effect of grade retention on adult crime: Evidence from a test-based promotion policy,” NBER working paper 25384.
- [27] Freeman, R. (1996). “Why Do So Many Young American Men Commit Crimes and What Might We Do about It?” *Journal of Economic Perspectives*, 10(1):25-42.
- [28] Fu, C. and K. Wolpin (2018) “Structural Estimation of a Becker-Ehrlich Equilibrium Model of Crime: Allocating Police Across Cities to Reduce Crime”. *Review of Economic Studies*, 85 (4): 2097–2138.

- [29] Gauri, V. (1999). "School Choice In Chile: Two Decades of Educational Reform." Pitt Latin American Studies. University of Pittsburgh Press.
- [30] Gaviria, A. and S. Raphael (2001) "School-Based Peer Effects and Juvenile Behavior," *Review of Economics and Statistics*, 83 (2), 257–268.
- [31] Glaeser, E., B. Sacerdote (1999) "Why is There More Crime in Cities?," *Journal of Political Economy*, 107(2), S225–S258.
- [32] Glaeser, E., B. Sacerdote and J. Scheinkman (1996) "Crime and Social Interactions," *Quarterly Journal of Economics*, 111(2), 507–548.
- [33] Heckman, J. and S. Mosso (2014). "The Economics of Human Development and Social Mobility," Working paper, University of Chicago.
- [34] Hjalmarsson, R. (2008), "Criminal Justice Involvement and High School Completion," *Journal of Urban Economics* 63, 613-630.
- [35] Hjalmarsson, R., H. Holmlund, and M. Lindquist. (2015) "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro-Data." *Economic Journal* 125 (587): 1290–1326.
- [36] Imai, S. and K. Krishna (2004). "Employment, Deterrence, and Crime in a Dynamic Model." *International Economic Review* 45(3): 845-872.
- [37] Jacob, B., and L. Lefgren. (2003) "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime." *American Economic Review* 93 (5): 1560–77.
- [38] Katz, L., S. Levitt, and E. Shustorovich (2003) "Prison Conditions, Capital Punishment and Deterrence," *American Law and Economics Review*, 5(2), 318–343.
- [39] Keane, M., and K. Wolpin (1994): "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence." *The Review of Economics and Statistics*, 76: 648-672.
- [40] Kling, J., J. Ludwig, and L. Katz (2005) "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment," *Quarterly Journal of Economics*, 120(1), 87–130.
- [41] Levitt, S. (1996) "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation," *Quarterly Journal of Economics*, 111 (2), 319–351.
- [42] Levitt, S. (1997) "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, 87 (3), 270–290.
- [43] Levitt, S. (1998). "Juvenile Crime and Punishment," *Journal of Political Economy* 106, 1156–85.
- [44] Levitt, S and L. Lochner (2000). "The Determinants of Juvenile Crime." *Risky Behavior among Youths: An Economic Analysis*, Edited by J Gruber, University of Chicago Press.
- [45] Lochner, L. (2004). "Education, Work, and Crime: A Human Capital Approach." *International Economic Review*, 45(3):811-843.
- [46] Lochner, L. (2011). "Education Policy and Crime. In Controlling Crime: Strategies and Tradeoffs." edited by P. Cook, J. Ludwig, and J. McCrary. University of Chicago Press.

- [47] Lochner, L. and E. Moretti (2004). “The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports.” *American Economic Review*, 94(1):155-189.
- [48] Machin, S., O. Marie, and S. Vujic (2011) “The Crime Reducing Effect of Education.” *Economic Journal* 121 (552): 463–84.
- [49] Mancino, M, S. Navarro and D. Rivers (2015). “Separating State Dependence, Experience, and Heterogeneity in a Model of Youth Crime and Education.” Working Paper, University of Western Ontario.
- [50] Meckes, L. and R. Carrasco (2010). “Two decades of simce: an overview of the national assessment system in chile.” *Assessment in Education: Principles, Policy and Practice*. 17(2): 233-248.
- [51] Merlo, A. and K. Wolpin (2009). “The Transition from School to Jail: Youth Crime and High School Completion Among Black Males.” Working Paper, Rice University.
- [52] Mizala, A. and P. Romaguera (2000). “School performance and choice: The chilean experience.” *The Journal of Human Resources* 35(2), pp. 392-417.
- [53] Munyo, I. (2015). “The Juvenile Crime Dilemma.” *Review of Economic Dynamics*, 18: 201–211.
- [54] Nagin, D., D. Farrington, and T. Moffitt (1995). “Life-Course Trajectories of Different Types of Offenders.” *Criminology*, 33, 111-139.
- [55] Nagin, D. and K. Land (1993). “Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric, Mixed Poisson Model.” *Criminology*, 31, 327-362.
- [56] Nagin, D. and R. Paternoster (1991). “On the Relationship of Past to Future Participation in Delinquency.” *Criminology*, 29, 163-189.
- [57] Sampson, R. and J. Laub (1995). *Crime in the Making: Pathways and Turning Points through Life*. Cambridge, MA: Harvard University Press.
- [58] Stange, K. (2012) “An Empirical Examination of the Option Value of College Enrollment.” *American Economic Journal: Applied Economics*, 4(1): 49–84.
- [59] Valenzuela, J. , C. Bellei, and D. Rios (2014). “Socioeconomic School Segregation in a Market-oriented Educational System: the Case of Chile.” *Journal of Education Policy* 29(2), 217–241.
- [60] Wilson, J. and R. Herrnstein (1985). *Crime and Human Nature*. New York, NY: Simon & Schuster.

Appendix

A1. Functional Forms

Throughout this section, we denote SES_i as student i 's SES status, while SES_k^s as school k 's SES classification by the Ministry of Education.

A1.1 Utility of Schooling

$$U^{sch}(G_{it}, GPA_{it}, X_i, \chi_i, W_k, g_{it}, J_{it-1}) =$$

$$\begin{aligned} & \alpha_0 + \alpha_1 GPA_{it} + \alpha_2 I(GPA_{it} < GPA^*) + \alpha_3 (1 - g_{it}) \\ & + \alpha_4 I(G_{it} > 8) + \alpha_5 I(k \text{ is private}) + \alpha_6 I(S_k = \text{low}) + \alpha_7 I(SES_k^s < 3) \\ & + p_k (\alpha_8 I(\text{inc}_i = \text{low}) + \alpha_9 I(\text{inc}_i = \text{mid}) + \alpha_{10} I(\text{inc}_i = \text{high})) + \alpha_{11} I(\chi_i = 2) \\ & + \alpha_{12} I(J_{it-1,1} > 0) + d_{it} (\alpha_{13} I(G_{it} \leq 8) + \alpha_{14} I(G_{it} > 8)). \end{aligned} \quad (20)$$

where GPA^* is the 25th percentile of GPA over all students, so $I(GPA_{it} < GPA^*)$ indicates low GPA. α_3 is disutility for being retained. The second line specifies the utility associated with different school characteristics: high school ($G_{it} > 8$) or primary school, private or not, whether or not the school quality (average score) is low (*i.e.*, below the 25th percentile), and the school's SES status (1 being low and 3 being high). In the third row, $p_k \in W_k$ is the tuition charged for attending school k , $\text{inc}_i \in X_i$ is i 's family income level. We introduce α_8 to α_{10} to capture the idea that the trade-off between education attainment and tuition costs may depend on family resources. α_{11} is a type-2 specific constant term. In the last row, α_{12} is the disutility of attending school for someone with a criminal record. α_{13} (α_{14}) is the disutility of attending primary school (high school) while committing crimes.

A1.2 School Transition:

The probability of student i in primary school k transferring to high school k' is given by:

$$P^{tr}(X_i, \chi_i, GPA_i, W_k, W_{k'}) = \begin{cases} \frac{\exp(f(X_i, \chi_i, GPA_i, W_k, W_{k'}))}{\sum_{k' \in K(l_i)} \exp(f(X_i, \chi_i, GPA_i, W_k, W_{k'}))} & \text{if } k' \in K'(l_i) \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

where $K'(l_i)$ is the set of high schools that are *feasible* for a student living in location l_i (l_i is one element of X_i). Empirically, we define $K'(l_i)$ as the collection of all high schools with at least one attendee from location l_i , which includes all high schools in l_i and some high schools in other, mostly nearby, locations. The function $f(\cdot)$ in (21) is given by

$$f(X_i, \chi_i, GPA_i, W_k, W_{k'}) =$$

$$(S_{k'} - S_S^*) \left[\begin{array}{l} \psi_1(S_k - S_P^*)_+ + \psi_2(S_k - S_P^*)_- + \\ \psi_3(GPA_i - M_G^*)_+ + (\psi_4 + \psi_5 I(\chi_i = 2))(GPA_i - M_G^*)_- \end{array} \right] \quad (22)$$

$$\begin{aligned} &+ I(k' \text{ private})(\psi_6 + \psi_7 I(k \text{ private})) \\ &+ \sum_{n=1}^3 I(SES_k^s = n)(\psi_{8n} I(SES_{k'}^s = 1) + \psi_{9n} I(SES_{k'}^s = 3)) \\ &+ \sum_{n=1,3} I(SES_i = n)(\psi_{8'n} I(SES_{k'}^s = 1) + \psi_{9'n} I(SES_{k'}^s = 3)) \end{aligned} \quad (23)$$

S_k is the average SIMCE score in school k , S_S^* (S_P^*) is the median of S_k within secondary (primary) schools, M_G^* is the median (standardized) GPA among all students. $(S_k - S_P^*)_+ = (S_k - S_P^*)I(S_k > S_P^*)$, $(S_k - S_P^*)_- = (S_k - S_P^*)I(S_k \leq S_P^*)$; and the other terms are similarly defined. The first row of (22) captures the transition likelihood associated with quality differences between a secondary school and one's primary school as well as one's own GPA. In particular, $(S_{k'} - S_S^*)$ measures the quality of a given secondary school k' relative to a median secondary school. This term is interacted with the quality of one's primary school relative to a median primary school as well as one's own GPA relative to a median student. The coefficients are allowed to differ depending on whether or not the primary school or the student is above the median, and on student unobserved type (χ_i). The second row captures the ease/difficulty to transfer to a private secondary school for all and for those enrolled in private primary schools. The third row captures the transition likelihood associated with the interaction of the SES status of the secondary school and the origin primary school. The last row interacts the SES status of a high school and a student's own SES.

A1.3 Terminal Value

$$V_{T+1}(X_i, \chi_i, G_{iT}, J_{iT}, E_{iT}(a_i)) =$$

$$\begin{aligned} &G_{iT}\lambda_1 + \lambda_2 I(J_{iT1} > 0) + \\ &I(G_{iT} = 12) \left[\begin{array}{l} \lambda_3 + \lambda_4 I(\chi_i = 2) + \lambda_5 \exp(E_{iT}(a_i)) \\ + \lambda_6 I(\text{parent edu}=\text{mid}) + \lambda_7 I(\text{parent edu}=\text{high}) \end{array} \right]. \end{aligned}$$

The first row represents the value of grade completed (G_{iT}) and whether or not one is ever arrested ($I(J_{iT1} > 0)$). The second row specifies the value of finishing high school, which may

differ by one's type, one's belief about his ability, and one's parental education. We take the exponential $\exp(E_{iT}(a_i))$ because $E_{iT}(a_i)$ can be negative, which arises from the fact that test scores are standardized to have a mean of 0 and that a_i and test scores are on the same scale.

A1.4 Distribution of Sentencing Length

We measure τ in units of 6 months, with $\tau \in \{0, 1, 2, \dots, 11\}$. We assume that τ is drawn from an ordered probit distribution, with the continuous latent variable (τ^*) given by

$$\tau_{it}^* = \varrho_0 + \varrho_1 age_{it} + \varrho_2 J_{it-1,1} + \varrho_3 J_{it-1,2} + \varepsilon_{it}^\tau,$$

where $J_{it-1,1}$ ($J_{it-1,2}$) is the total arrests (total sentences) in the past. To estimate ϱ and the cutoff parameters from our DPP data, we assign $\tau = 0$ if no jail time was prescribed, $\tau = 1$ if the observed sentence was positive but no greater than 6 months, ..., $\tau = 10$ if it was longer than 54 months but no greater than 60, and $\tau = 11$ if it was 60 months or longer.

B. Standardization of GPA

For a primary school k , we estimate its grading parameters (a_k^{raw}, b_k^{raw}) using its students' Grade 4 raw GPAs and Grade 4 SIMCE scores in the following regression

$$GPA_{ki0}^{raw} = a_k^{raw} + b_k^{raw} SIMCE_{ki0} + \varepsilon_{ki0}^{raw}.$$

The standardized GPA in primary school k is then given by $GPA_{kit} = \frac{GPA_{kit}^{raw} - a_k^{raw}}{b_k^{raw}}$. To standardize secondary school GPAs, we follow the same procedure but school grading parameters are estimated by comparing Grade 10 raw GPAs with Grade 10 SIMCE scores.

C. The Likelihood Function

Model parameters to be estimated (Θ) include 1) preference parameters Θ_u , 2) GPA production and grade progression parameters Θ_G , 3) type distribution parameters Θ_χ , 4) type-specific ability distribution parameters Θ_a , 5) school transition probability parameters Θ_{tr} , and 6) the parameter that relates standardized test score (unknown to student) and ability (σ_ξ). Given observables (X_i, k_{i0}) , the estimated Θ should maximize the probability of observing each teen's school enrollment, GPAs, grade progress and arrests over time, i.e., $O_{it} = \{e_{it}, GPA_{it}, g_{it+1}, arrest_{it}\}_t$, his school transition (k'_i) and his performances before Grade

$G_0 (\{GPA_{it}, g_{it}\}_{t=-t_0}^0, s_i)$.⁴⁴ Teenager i 's contribution is given by

$$L_i(\Theta) = \sum_m \Pr(\chi_i = m | X_i, k_{i0}; \Theta_\chi) L_{im}(\Theta \setminus \Theta_\chi),$$

$$L_{im}(\Theta \setminus \Theta_\chi) = \prod_{t=1}^T L_{imt}(\Theta \setminus \Theta_\chi) \times L_{im0}(\Theta_G, \Theta_a, \sigma_\xi) \times p^{tr}(\cdot, \chi_i = m | \Theta_{tr}).$$

L_{im} is the likelihood conditional on i being type m , which is the product of period-specific likelihoods $L_{imt}(\Theta \setminus \Theta_\chi)$, complemented by the additional information provided by one's pre- G_0 performance ($L_{im0}(\cdot)$), and the school transition probability as specified in (7).

$L_{imt}(\Theta \setminus \Theta_\chi)$ for $t \geq 1$: Let Ω_{imt} be the vector of state variables for i of Type- m at time t , and $\bar{\Omega}_{imt}$ be the part net of payoff shocks.⁴⁵ $arrest_{it} = 1$ implies $d_{it} = 1$. When $arrest_{it} = 0$, one may not have committed a crime or may have done so but failed to be arrested.⁴⁶ Therefore,

$$L_{imt}(\Theta \setminus \Theta_\chi) = \Pr(O_{it}; \Theta \setminus \Theta_\chi | \bar{\Omega}_{imt})$$

$$= \begin{cases} \rho_{li} \Pr(e_{it}, d_{it} = 1; \Theta \setminus \Theta_\chi) L_{imt}^G(d_{it} = 1; \Theta_G, \Theta_a) & \text{if } arrest_{it} = 1 \\ \left((1 - \rho_{li}) \Pr(e_{it}, d_{it} = 1; \Theta \setminus \Theta_\chi) L_{imt}^G(d_{it} = 1; \Theta_G, \Theta_a) \right. \\ \left. + \Pr(e_{it}, d_{it} = 0; \Theta \setminus \Theta_\chi) L_{imt}^G(d_{it} = 0; \Theta_G, \Theta_a) \right) & \text{if } arrest_{it} = 0 \end{cases}. \quad (24)$$

The two major ingredients in $L_{imt}(\Theta \setminus \Theta_\chi)$ are the choice probability $\Pr(e_{it}, d_{it} | \cdot)$ and the academic outcome probability $L_{imt}^G(d_{it}; \cdot)$.

$$\Pr(e_{it}, d_{it}; \Theta \setminus \Theta_\chi | \cdot) = \frac{\exp\left(\bar{V}_{it}^{e_{it}d_{it}}(\bar{\Omega}_{imt})\right)}{\sum_{ed \in \{0,1\} \times \{0,1\}} \exp\left(\bar{V}_{it}^{ed}(\bar{\Omega}_{imt})\right)}, \quad (25)$$

where $\bar{V}_{it}^{ed}(\bar{\Omega}_{imt}) \equiv V_{it}^{ed}(\Omega_{imt}) - v_{it}^{ed}$ is the part of the value function net of the payoff shock.

$L_{imt}^G(d_{it}; \Theta_G, \Theta_a) =$

$$\phi\left(\frac{GPA_{it} - \overline{GPA}_{it} - \bar{a}_\chi}{\sqrt{\sigma_a^2 + \sigma_{\epsilon gpa}^2}}\right) \times \begin{bmatrix} g_{it} \Pr(g_{it} = 1 | GPA_{it}, W_{k_{it}}, \chi_i = m) + \\ (1 - g_{it}) (1 - \Pr(g_{it} = 1 | \cdot)) \end{bmatrix}. \quad (26)$$

⁴⁴Notice that two outcomes in the model do not enter the likelihood directly: 1) crime, because it is observed only via arrest; 2) jail sentences, which, conditional on arrests, do not depend on Θ and hence will not contribute to the likelihood.

⁴⁵The following state variables are directly observed in the data: $X_i, J_{it-1}, G_{it-1}, e_{it-1}, k_{it-1}, g_{it}$. Given the observed history of GPA's, one's beliefs EA_{it-1}, VA_{it-1} only depends on model parameters and one's type.

⁴⁶The likelihood of observing a missing value of an outcome is 1 in non-applicable cases. For example, $O_{it} = \cdot$ if i is in jail from previous sentence; $GPA_{it} = \cdot$ if $e_{it} = 0$.

The first part specifies the likelihood of observing GPA_{it} , with \overline{GPA}_{it} defined in (4). The second part is the probability of observing the grade progress result (g_{it}) conditional on GPA_{it} .

$\mathbf{L}_{im0}(\Theta_G, \Theta_a, \sigma_\xi)$ consists of the likelihood of one's pre- G_0 performance:

$$L_{im0}(\Theta_G, \sigma_\xi) = \prod_{t=-t_0}^0 L_{imt}^G(d_{it}; \Theta_G) \times \phi\left(\frac{s_i - \bar{a}_\chi}{\sqrt{\sigma_a^2 + \sigma_\xi^2}}\right). \quad (27)$$

We assume that young children do not commit crimes before they enter our model, so $L_{imt}^G(d_{it}; \Theta_G)$ is as specified in (26) with $d_{it} = 0$. The second part is the additional information that is available to the researcher but not to the student (s_i).

Table A1. Distribution of Attendance Rates

$\frac{\text{Attendance days}}{\text{School days}}$	0	(0, 0.5)	[0.5, 0.6)	[0.6, 0.7)	[0.7, 0.8)	[0.8, 1]
Age 11 ($t = 1$)	0.4%	0.1%	0.1%	0.1%	0.9%	98.4%
Age 18 ($t = T$)	8.8%	0.7%	0.8%	1.9%	5.6%	82.2%

Attendance rates at Age 11 ($t = 1$) and Age 18 ($t = T$) are shown as examples.

One is treated as being enrolled in t if his attendance rate is at least 0.5 in t .

Table A2. Parameter Estimates: Primary School k to Secondary School k' Transfer Probability

School SIMCE Scores ^a : S_k vs $S_{k'}$	Student SES _{i} vs HS $SES_{k'}^s$
$(S_{k'} - S_S^*)(S_k - S_P^*)_+$	0.37 (0.02) $I(SES_i = 1, SES_{k'}^s = 1)$ 0.56 (0.04)
$(S_{k'} - S_S^*)(S_k - S_P^*)_-$	-0.25 (0.02) $I(SES_i = 1, SES_{k'}^s = 3)$ -0.32 (0.07)
$I(\text{Type 2})(S_{k'} - S_S^*)(S_k - S_P^*)_-$	0.67 (0.03) $I(SES_i = 3, SES_{k'}^s = 1)$ -0.61 (0.03)
^a : $S_P^*(S_S^*)$: median SIMCE in primary (high) schools.	$I(SES_i = 3, SES_{k'}^s = 3)$ 0.66 (0.04)
Student GPA _{i} vs HS SIMCE ^b	Primary School SES_k^s vs HS $SES_{k'}^s$
$(S_{k'} - S_S^*)(GPA_i - M_G^*)_+$	0.55 (0.01) $I(SES_k^s = 1, SES_{k'}^s = 1)$ 1.12 (0.04)
$(S_{k'} - S_S^*)(GPA_i - M_G^*)_-$	0.35 (0.01) $I(SES_k^s = 1, SES_{k'}^s = 3)$ -1.76 (0.09)
^b : M_G^* median Grade 8 GPA across students..	$I(SES_k^s = 2, SES_{k'}^s = 1)$ -0.14 (0.02)
Private High School	$I(SES_k^s = 2, SES_{k'}^s = 3)$ -1.84 (0.04)
$I(k' \text{ private})$	-0.72 (0.02) $I(SES_k^s = 3, SES_{k'}^s = 1)$ -0.85 (0.03)
$I(k, k' \text{ both private})$	1.35 (0.03) $I(SES_k^s = 3, SES_{k'}^s = 3)$ -0.46 (0.03)

Table A3. Other Parameter Estimates

A. Initial SIMCE Score		C. Grade-specific GPA constant	
Constant (δ_0)	0.07 (0.01)	γ_{05}	-0.03 (0.005)
Ability (δ_1)	0.95 (0.01)	γ_{06}	-0.04 (0.007)
σ_ξ	0.76 (0.01)	γ_{07}	-0.04 (0.008)
B. Terminal Value Function		γ_{08}	0 (normalized)
Grade level G_{iT}	0.0001 (0.12)	γ_{09}	0.02 (0.007)
$I(G_{iT} = 12)$	0.005 (0.42)	γ_{10}	0.05 (0.007)
$I(G_{iT} = 12)I(\text{Parent edu} = \text{mid})$	0.59 (0.13)	γ_{11}	-0.002 (0.007)
$I(G_{iT} = 12)I(\text{Parent edu} = \text{high})$	0.40 (0.17)	γ_{12}	0.05 (0.008)
$I(G_{iT} = 12) \times \exp(E_{iT}(a_i))$	0.84 (0.22)		
$I(G_{iT} = 12)I(\text{Type 2})$	0.16 (0.29)		
$I(\text{ever arrested})$	-4.18 (0.37)		