

# Wandering astray: Teenagers' choices of schooling and crime

CHAO FU

Department of Economics, University of Wisconsin and NBER

NICOLÁS GRAU

Faculty of Economics and Business, University of Chile

JORGE RIVERA

Faculty of Economics and Business, University of Chile

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.

**KEYWORDS.** Teenage crime, education, information friction, institutional friction, dynamic model, structural estimation.

**JEL CLASSIFICATION.** I2, K42.

## 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

---

Chao Fu: [cfu@ssc.wisc.edu](mailto:cfu@ssc.wisc.edu)

Nicolás Grau: [ngrau@fen.uchile.cl](mailto:ngrau@fen.uchile.cl)

Jorge Rivera: [jrivera@fen.uchile.cl](mailto:jrivera@fen.uchile.cl)

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.

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.<sup>1</sup>

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.<sup>2</sup> Second, a teenager's payoff from schooling may depend directly on one's family background.<sup>3</sup> 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.

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, that is, 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

---

<sup>1</sup>Our model focuses on teenagers' decisions and takes the endowment by teenage years as predetermined. The word ability throughout this paper refers to one's ability predetermined 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.

<sup>2</sup>For 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.

<sup>3</sup>For 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.

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%.<sup>4</sup> 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.<sup>5</sup> Closer to our pa-

<sup>4</sup>As 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, Deming, and Rockoff (2014), who study the effect of a large-scale school desegregation policy on education and crime outcomes.

<sup>5</sup>For 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, Levitt, and Shustorovich (2003), and Di Tella and Schargrodsky (2004)), social capital, social interactions, and peer effects (Case and Katz (1991), Glaeser, Sacerdote, and Scheinkman (1996), Gaviria and Raphael (2001), Jacob and Lefgren (2003), Kling, Ludwig, and Katz (2005), and Sickles and Williams (2008)), and family circumstances and structure (Glaeser and Sacerdote (1999), Donohue and Levitt (2001)).

per 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.<sup>6</sup> [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.<sup>7</sup> For example, [Nagin and Paternoster \(1991\)](#), [Nagin and Land \(1993\)](#), [Nagin, Farrington, and Moffitt \(1995\)](#), and [Broidy et al. \(2003\)](#) find that both unobserved individual types and state dependence are important. [Merlo and Wolpin \(2015\)](#), 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. \(2016\)](#) 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.<sup>8</sup> 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.<sup>9</sup> 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 out-

<sup>6</sup>[Jacob 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, Marie, and Vujic \(2011\)](#), [Clay, Lingwall, and Stephens \(2012\)](#), [Anderson \(2014\)](#), [Hjalmarsson, Holmlund, and Lindquist \(2015\)](#), and [Cook and Kang \(2016\)](#)).

<sup>7</sup>See [Blumstein et al. \(1986\)](#), [Gottfredson and Hirschi \(1990\)](#), [Sampson and Laub \(1995\)](#), and [Wilson and Herrnstein \(1985\)](#) for examples in criminology.

<sup>8</sup>Some studies have questioned the assumption of rationality and forward-looking. For example, [Wilson and Herrnstein \(1985\)](#), [Katz, Levitt, and Shustorovich \(2003\)](#), 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.

<sup>9</sup>[Munyo \(2015\)](#) 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, Grau, and Rivera \(2020\)](#) find that arrest and incarceration prior to age 16 reduces the probability of high school graduation.

comes. 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. 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.<sup>10</sup> 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.<sup>11</sup> 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 (nonvoucher-private). In 2014, 54.2% of the schools were public, 41% were voucher-private, and 4.8% were nonvoucher-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

<sup>10</sup>The 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).

<sup>11</sup>In 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.

“elite” family backgrounds, were enrolled in nonvoucher-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 primary 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.<sup>12</sup> 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”.<sup>13</sup>

**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.<sup>14</sup> 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.

<sup>12</sup>Many schools, especially nonpublic schools, required parents to report their income and present their marriage certificates; some schools even conduct parental interviews. See Contreras, Sepulveda, and Bustos (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).

<sup>13</sup>*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. (2014).

<sup>14</sup>Table B1 in the Appendix in the Online Supplementary Material (Fu, Grau, and Rivera (2022)) 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.<sup>15</sup> 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, that is, 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 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.

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,

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

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 (1000 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		1221		

Family background	Primary school type		Primary school Ave. SIMCE		
	Public	Voucher-private	1st quar	2nd quar	Above median
Monthly income/person (1000 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	8627	8708	17,448

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

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.<sup>16</sup>

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

<sup>16</sup>For 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.



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	6911	19,494	8379	8627	8708	17,448

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.<sup>17</sup> 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.

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 nontrivial uncertainties about their learning abilities ([Altonji \(1993\)](#), [Arcidiacono \(2004\)](#), [Cunha, Heckman, and Navarro \(2005\)](#), [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 ( $DI_{it}$ ) and grade dummies ( $DG_{G_{it}}$ ):

$$GPA_{it} = DI_{it} + DG_{G_{it}} + \varepsilon_{it}, \tag{1}$$

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  $\varepsilon_{it}$  arguably captures the deviation of

<sup>17</sup>The fraction of teenagers with  $\frac{\text{attendance days}}{\text{school days}} \in (0, 0.5)$  is very small, for example, 0.1% at age 11 and 0.7% at age 18 (see Table A1 in the [Appendix](#)).

TABLE 4. Regression of current enrollment on lagged GPA residuals.

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

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

one’s GPAs from the expected level. Then we regress enrollment status ( $e_{it}$ ) on individual dummies ( $DI_{i2}$ ) and the lagged GPA residuals  $\{\widehat{\varepsilon}_{it'}\}_{t' < t}$ .

$$e_{it} = DI_{i2} + \sum_{n>0} b_n \widehat{\varepsilon}_{it-n} + \nu_{it},$$

(2)

where  $\nu_{it}$  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.<sup>18</sup>

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  $k \in K'$ ).<sup>19</sup>

**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  $\chi_i$  but not  $a_i$ . Ability is normally distributed with type-specific means:

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

(3)

One has initial belief given by (3), and updates it over time as new information (GPA) comes in. The researcher observes neither  $\chi_i$  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.

<sup>18</sup>For 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  $t' < t$ .

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

The variable  $\chi_i$  captures heterogeneity known to teenagers. Different types of teenagers differ in the distributions of their abilities and preferences.

**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  $\tau$  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  $\text{GPA}_{it}$  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<sup>20</sup>

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

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  $\gamma_4$  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,  $\epsilon_{it}^{\text{gpa}}$  is an i.i.d. shock drawn from  $N(0, \sigma_{\epsilon^{\text{gpa}}}^2)$ . A student cannot separately observe  $a_i$  and  $\epsilon_{it}^{\text{gpa}}$ . Let  $\overline{\text{GPA}}_{it}$  be the part of GPA net of  $a_i + \epsilon_{it}^{\text{gpa}}$ .

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  $\text{GPA}_{it}$  progresses to the next grade in  $t + 1$  is given by

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

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

<sup>20</sup>In 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.

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  $a_i$ .<sup>21</sup> 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

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

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 ( $\sigma_{\epsilon_{\text{gpa}}}$ ) versus that of the ability distribution ( $\sigma_a$ ). 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 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  $p_{k'i}^{\text{tr}}$ ; we model  $p_{k'i}^{\text{tr}}$  as dependent on  $X_i, \chi_i$ , one's grade 8 GPA ( $GPA_{i4}$ ), the characteristics of one's primary school  $W_{k_{i0}}$  and the destination school  $W_{k'}$ , such that

$$p_{k'i}^{\text{tr}} = P^{\text{tr}}(X_i, \chi_i, GPA_{i4}, W_{k_{i0}}, W_{k'}) = \begin{cases} \frac{\exp(f(X_i, \chi_i, GPA_{i4}, W_{k_{i0}}, W_{k'}))}{\sum_{k' \in K(l_i)} \exp(f(X_i, \chi_i, GPA_{i4}, W_{k_{i0}}, W_{k'}))} & \text{if } k' \in K'(l_i), \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

<sup>21</sup>A 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, Depew, and Barnes (2017) and Eren, Lovenheim, and Mocan (2018).

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  $X_i$ );<sup>22</sup>  $f(\cdot)$  is a linear function of student, primary school and secondary school characteristics and their interactions. The direct effect of  $X_i$  and  $\text{GPA}_{i4}$  on  $p^{\text{tr}}$  reflect the fact that schools can select based on family background and GPA. To allow for nonrandom matching for factors known to the school and the student but not to the researcher, we introduce  $\chi_i$  into  $p^{\text{tr}}(\cdot)$ . 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  $\chi_i$ , 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.

**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  $v_{it} = \{v_{it}^{\text{ed}}\}_{(e,d) \in \{0,1\}^2}$  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  $\text{GPA}_{it}$  and grade progression ( $g_{it+1}$ ) are realized. One's belief about  $a_i$  is updated. If  $d_{it} = 1$ , one's arrest/nonarrest 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  $t$ ,  $\Omega_{it}$ , contains the following:

1. Permanent variables:  $\chi_i, X_i$ .
2. Dynamic variables:  $\text{EA}_{it-1}, \text{VA}_{it-1}, J_{it-1}, G_{it-1}, g_{it}, e_{it-1}, k_{it}$ .
  - (1)  $\text{EA}_{it-1}, \text{VA}_{it-1}$ :  $i$ '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.

<sup>22</sup>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.

- (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  $G_{it-1} \geq 9$ .<sup>23</sup>

3. Transitory shocks to choice-specific payoffs:  $v_{it} = \{v_{it}^{\text{ed}}\}$

The evolution from  $\Omega_{it}$  to  $\Omega_{it+1}$  depends on not only one's choices ( $e_{it}, d_{it}$ ) but also a vector  $Y_{it}$  of factors realized at the *end* of  $t$ . The vector  $Y_{it}$  consists of a)  $\text{GPA}_{it}$  (if  $e_{it} = 1$ ), which affects belief updating ( $\text{EA}_{it}, \text{VA}_{it}$ ); b) the grade progression shock (if  $e_{it} = 1$ ), which, together with  $\text{GPA}_{it}$ , 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  $\Omega_{it+1}$  instead of  $(\Omega_{it+1} | \Omega_{it}, e_{it}, d_{it}, Y_{it})$ .<sup>24</sup>

## 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

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

where  $V_t^{\text{ed}}(\cdot)$  denotes the choice-specific value function. We specify  $V_t^{\text{ed}}(\cdot)$  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  $\beta$  is the discount factor. The expectation for the next period value is taken over payoff shocks  $\{v_{it+1}^{\text{ed}}\}$ .

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

$$U^{\text{sch}}(\text{GPA}_{it}, X_i, \chi_i, W_{k_{it}}, d_{it}, g_{it}, J_{it-1}). \quad (9)$$

That is, students care about their performance in school (GPA), but schooling utility may vary with one's backgrounds  $X_i$ , type  $\chi_i$ , and school characteristics  $W_{k_{it}}$ . Criminal participation  $d_{it}$  may also affect schooling utility if there is a utility cost of engaging in both

<sup>23</sup>Within 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.

<sup>24</sup>The state space is large because the dynamic beliefs ( $\text{EA}_{it-1}, \text{VA}_{it-1}$ ) are continuous variables. We solve the model via backward induction with Monte Carlo integration and interpolation (Keane and Wolpin (1994)); details are in the Appendix in the Online Supplementary Material (Fu, Grau, and Rivera (2022)).

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

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

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

**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 ( $r_{\chi_i}$ ). 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  $\rho_{l_i}$ , 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 A.1.4). 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.<sup>25</sup>

To be more specific, the payoff from uncaught crimes is a function of local crime rate ( $cr_l$ ), local average income ( $\text{INC}_l$ ), and one's family income ( $\text{inc}_i$ ), given by<sup>26</sup>

$$R(l_i, X_i) = \omega_1 cr_{l_i} + \text{INC}_{l_i}(\omega_2 + \omega_3 I(\text{inc}_i = \text{low}) + \omega_4 I(\text{inc}_i = \text{middle})). \quad (12)$$

If arrested, one's utility is given by

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

<sup>25</sup>The sentence  $\tau$  may not be an integer (year), in which case, we round  $\tau$  to count the length of the inaction period.

<sup>26</sup>Our 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.



where  $\mu_2$  is the additional disutility for first-time arrests.<sup>27</sup>

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

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

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  $\sum_{m=0}^1 \Pr(g_{it+1} = m | W_{k_{it}}, \text{GPA}_{it}, \chi_i) \beta E_v V_{t+1}(\Omega_{it+1})$  in equations (10) and (14) is replaced by

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

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

**4.2.3 Terminal value** One's terminal value is a function of his characteristics  $(X_i, \chi_i)$ , 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,

<sup>27</sup>In 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, and we therefore choose the simpler specification.

TABLE 5. Information on one's endowment.

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

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 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 ( $\{\text{GPA}_{it}\}_{t=-3}^0$ ) from grade 1 to grade 4, before the model starts. Neither the teenager nor the researcher observes  $a_i$ . The teenager knows his  $\chi_i$ , 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** ( $\text{EA}_{i0}, \text{VA}_{i0}$ ): By  $t = 1$ , the teenager's should have already used  $\{\text{GPA}_{it}\}_{t=-3}^0$  to update his initial belief. Therefore, we model  $(\text{EA}_{i0}, \text{VA}_{i0})$  as the Bayes output starting from  $a_i \sim N(\bar{a}_{\chi_i}, \sigma_a^2)$  and updated with  $\{\text{GPA}_{it}\}_{t=-3}^0$  following (6).

**The unobservable** ( $\chi_i, a_i$ ): 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  $a_i \sim N(\bar{a}_{\chi_i}, \sigma_a^2)$ ,  $(\chi_i, a_i)$  is distributed as

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

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

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

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

5. ESTIMATION AND IDENTIFICATION

We assume that the sentencing length  $\tau$  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 A1.4). 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 preset the annual discount factor  $\beta$  at 0.9.<sup>28</sup>

We estimate all the other model parameters using the maximum likelihood method. The vector  $\Theta$  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, that is,

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

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

### 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  $\chi_i$ .

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 ( $\chi_i$ ). 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,

$$\begin{aligned} & E[\text{GPA}_{it} | X_i, W_{k_{it}}, W_{k_{i0}}, \{\text{GPA}_{it}\}_{t=-3}^0] \\ &= \gamma_0 G_{it} + X_i \gamma_1 + W_{k_{it}} \gamma_2 + X_i W_{k_{it}} \gamma_3 + \gamma_4 (1 - g_{it}) + E[a_i | X_i, W_{k_{i0}}, \{\text{GPA}_{it}\}_{t=-3}^0] \end{aligned}$$

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

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

The last equality holds because

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

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  $\delta_1$  and  $\{\gamma_n\}_{n=1}^4$ . The constants  $\{\gamma_{0G_t}\}$  and  $\delta_0$  are not all identified, we therefore normalize  $\gamma_{0G_t}$  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 unimodal and symmetric, the observed modes as well as the distribution of  $(X, W_k)$  surrounding each mode are informative about type-specific mean ability  $\bar{a}_\chi$  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  $\rho_l$ , 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.

TABLE 6. GPA, ability, and grade progression.

A. GPA production: $\text{GPA}_{it} = \gamma_0 G_{it} + X_i \gamma_1 + W_{kit} \gamma_2 + X_i W_{kit} \gamma_3 + \gamma_4 (1 - g_{it}) + a_i + \epsilon_{it}^{\text{gpa}}$			
$X_i$ :	Parent Edu = Low	−0.16 (0.002)	$W_{kit}$ : Average SIMCE <sup>b</sup> 0.40 (0.001)
	Parent Edu = High	0.11 (0.002)	Ave. SIMCE* (parent edu low) −0.07 (0.002)
	Income = Low <sup>a</sup>	−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 preschool	0.08 (0.002)	School SES = Middle 0.03 (0.002)
	(1 − $g_{it}$ ): Retained	−0.15 (0.01)	$\sigma_{\epsilon^{\text{gpa}}}$ 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)	$\sigma_a$ 0.25 (0.02)
$\bar{a}_2$ :	Mean ability (Type2)	−0.50 (0.01)	
C. Grade progression $\Pr(g_{it+1} = 1   \cdot) = \Phi\left(\frac{\text{GPA}_{it} + \theta_0 X_i + W_{kit} \theta_1}{\sigma_{\epsilon^g}}\right)$			
$\theta_{01}$ :	Type 1	0.40 (0.01)	$W_{kit}$ : High school −0.02 (0.01)
$\theta_{02}$ :	Type 2	0.50 (0.01)	Public primary 0.04 (0.01)
$\sigma_{\epsilon^g}$		0.59 (0.003)	Public high school −0.29 (0.01)

<sup>a</sup>Family income levels (low, middle, high) are defined by income terciles.  
<sup>b</sup>School average SIMCE score in the 4th (10th) grade for primary school (high school).

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.<sup>29</sup>

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 preschool 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 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

<sup>29</sup>To derive the standard errors, we numerically calculate the Hessian of the log likelihood.

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) <sup>a</sup>	−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) <sup>b</sup>	−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) <sup>c</sup>	−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 $\varphi$	1.31 (0.04)	I (first time arrest)	−32.3 (1.04)
Constant	4.08 (0.10)		

<sup>a</sup>GPA is low if it is below the 25th percentile in the individual GPA distribution.  
<sup>b</sup>Annual tuition in 200,000 pesos. Family income groups are defined by income terciles.  
<sup>c</sup>School ave. SIMCE is low if it is below the 25th percentile among all primary/secondary schools.

schools are more enjoyable to attend.<sup>30</sup> 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).<sup>31</sup> 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

<sup>30</sup>We 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.  
<sup>31</sup>The 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 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 (1000 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

even larger across teenagers attending different primary schools grouped by school SES and by school-level SIMCE scores.

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 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 grades 5 to 8, and  $t = 5$  to 8 corresponds to high-school age. Consistent with the data, the model predicts that both nonenrollment 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

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. 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

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.

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.<sup>32</sup> 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.<sup>33</sup>

TABLE 11. Model fit: regression of enrollment on lagged GPA residuals.

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

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

<sup>32</sup>All three groups are located in the same (the largest) municipality and, therefore, face the same local conditions and set of feasible schools.  
<sup>33</sup>Given 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.

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

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, that is, a

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%

teenager in groups 1 and 2 would be treated equally as a group-3 teenager with the same GPA and type.<sup>34</sup>

**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  $t = 1$ .

The lower panel of Table 13 shows outcomes under each of the counterfactual scenarios and percentage changes relative to the baseline.<sup>35</sup> Between the two school-related factors, primary-secondary school transfer opportunities (F2) appear to be more important in explaining cross-group differences than grades 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).<sup>36</sup> 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.<sup>37</sup> 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 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.<sup>38</sup> 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

<sup>34</sup>That is, each teenager  $i$  is given the transfer probability vector  $\{P^{\text{tr}}(X_3, \chi_i, \text{GPA}_i, W_{k_{30}}, W_{k'})\}_{k'}$ , 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  $P^{\text{tr}}(\cdot)$ ).

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

<sup>36</sup>In the baseline, groups 1 and 2 attend public primary schools and group 3 attends private primary schools.

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

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

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	<u>new-base</u> <u>base</u>	New	<u>new-base</u> <u>base</u>	New	<u>new-base</u> <u>base</u>	Cost <sup>*</sup>
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

<sup>\*</sup> Additional high school voucher per targeted teenager per high school year (pesos).

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 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.<sup>39</sup>

<sup>39</sup>For 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

**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.<sup>40</sup> 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, that is, 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.<sup>41</sup>

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.

**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 Policy 0v is small but so is its effect.

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

---

allocate students to public schools, children living in poor neighborhoods have little chance of getting into good public schools.

<sup>40</sup>Specifically, 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  $P^{tr}(\cdot)$ ).

<sup>41</sup>Exploiting variation from Chicago public high school admission random lotteries, Cullen, Jacob, and Levitt (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.

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).<sup>42</sup> 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 predetermined endowments that are correlated with their family backgrounds. An important extension is to bring into our framework early childhood investment that

<sup>42</sup>This 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).

shapes one's initial conditions by teenage years.<sup>43</sup> 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.

## APPENDIX: APPENDIX

### A.1 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.

#### A.1.1 Utility of schooling

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

where  $\text{GPA}^*$  is the 25th percentile of GPA over all students, so  $I(\text{GPA}_{it} < \text{GPA}^*)$  indicates low GPA.  $\alpha_3$  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, that is, 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  $\alpha_8$  to  $\alpha_{10}$  to capture the idea that the trade-off between education attainment and tuition costs may depend on family resources.  $\alpha_{11}$  is a type-2 specific constant term. In the last row,  $\alpha_{12}$  is the disutility of attending school for someone with a criminal record.  $\alpha_{13}$  ( $\alpha_{14}$ ) is the disutility of attending primary school (high school) while committing crimes.

**A.1.2 School transition** The probability of student  $i$  in primary school  $k$  transferring to high school  $k'$  is given by

$$P^{\text{tr}}(X_i, \chi_i, \text{GPA}_i, W_k, W_{k'})$$

<sup>43</sup>See, for example, Cunha and Heckman (2007), Cunha, Heckman, and Schennach (2010) and Del Boca, Flinn, and Wiswall (2013) for studies on parental investment on children.



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

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

$$\begin{aligned} & f(X_i, \chi_i, \text{GPA}_i, W_k, W_{k'}) \\ &= (S_{k'} - S_S^*) \left[ \begin{aligned} & \psi_1(S_k - S_P^*)_+ + \psi_2(S_k - S_P^*)_- \\ & + \psi_3(\text{GPA}_i - M_G^*)_+ + (\psi_4 + \psi_5 I(\chi_i = 2))(\text{GPA}_i - M_G^*)_- \end{aligned} \right] \quad (22) \\ & + I(k' \text{ private})(\psi_6 + \psi_7 I(k \text{ private})) \\ & + \sum_{n=1}^3 I(\text{SES}_k^s = n)(\psi_{8n} I(\text{SES}_{k'}^s = 1) + \psi_{9n} I(\text{SES}_{k'}^s = 3)) \\ & + \sum_{n=1,3} I(\text{SES}_i = n)(\psi_{8'n} I(\text{SES}_{k'}^s = 1) + \psi_{9'n} I(\text{SES}_{k'}^s = 3)) \quad (23) \end{aligned}$$

$S_k$  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 ( $\chi_i$ ). 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.

### A.1.3 Terminal value

$$\begin{aligned} & V_{T+1}(X_i, \chi_i, G_{iT}, J_{iT}, E_{iT}(a_i)) \\ &= G_{iT} \lambda_1 + \lambda_2 I(J_{iT1} > 0) \\ & + I(G_{iT} = 12) \left[ \begin{aligned} & \lambda_3 + \lambda_4 I(\chi_i = 2) + \lambda_5 \exp(E_{iT}(a_i)) \\ & + \lambda_6 I(\text{parent edu} = \text{mid}) + \lambda_7 I(\text{parent edu} = \text{high}) \end{aligned} \right]. \end{aligned}$$

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.

**A.1.4 Distribution of sentencing length** We measure  $\tau$  in units of 6 months, with  $\tau \in \{0, 1, 2, \dots, 11\}$ . We assume that  $\tau$  is drawn from an ordered probit distribution, with the continuous latent variable ( $\tau^*$ ) given by

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

where  $J_{it-1,1}$  ( $J_{it-1,2}$ ) is the total arrests (total sentences) in the past. To estimate  $\varrho$  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.

## APPENDIX B: STANDARDIZATION OF GPA

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

$$\text{GPA}_{ki0}^{\text{raw}} = a_k^{\text{raw}} + b_k^{\text{raw}} \text{SIMCE}_{ki0} + \varepsilon_{ki0}^{\text{raw}}.$$

The standardized GPA in primary school  $k$  is then given by  $\text{GPA}_{\text{kit}} = \frac{\text{GPA}_{\text{kit}}^{\text{raw}} - a_k^{\text{raw}}}{b_k^{\text{raw}}}$ . 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.

## APPENDIX C: THE LIKELIHOOD FUNCTION

Model parameters to be estimated ( $\Theta$ ) include (1) preference parameters  $\Theta_u$ , (2) GPA production and grade progression parameters  $\Theta_G$ , (3) type distribution parameters  $\Theta_\chi$ , (4) type-specific ability distribution parameters  $\Theta_a$ , (5) school transition probability parameters  $\Theta_{\text{tr}}$ , and (6) the parameter that relates standardized test score (unknown to student) and ability ( $\sigma_\xi$ ). Given observables ( $X_i, k_{i0}$ ), the estimated  $\Theta$  should maximize the probability of observing each teen's school enrollment, GPAs, grade progress, and arrests over time, that is,  $O_{it} = \{e_{it}, \text{GPA}_{it}, g_{it+1}, \text{arrest}_{it}\}_t$ , his school transition ( $k'_i$ ), and his performances before grade  $G_0$  ( $\{\text{GPA}_{it}, g_{it}\}_{t=-t_0}^0, s_i$ ).<sup>44</sup> Teenager  $i$ 's contribution is given

<sup>44</sup>Notice 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  $\Theta$ , and hence will not contribute to the likelihood.

TABLE A1. Distribution of attendance rates.

Attendance days 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%

Note: 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  $t$ .

by

$$L_i(\Theta) = \sum_m \Pr(\chi_i = m | X_i, k_{i0}; \Theta_\chi) L_{im}(\Theta \setminus \Theta_\chi),$$
$$L_{im}(\Theta \setminus \Theta_\chi) = \prod_{t=1}^T L_{\text{imt}}(\Theta \setminus \Theta_\chi) \times L_{im0}(\Theta_G, \Theta_a, \sigma_\xi) \times p^{\text{tr}}(\cdot, \chi_i = m | \Theta_{\text{tr}}).$$

$L_{im}$  is the likelihood conditional on  $i$  being type  $m$ , which is the product of period-specific likelihoods  $L_{\text{imt}}(\Theta \setminus \Theta_\chi)$ , 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).

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

$$L_{\text{imt}}(\Theta \setminus \Theta_\chi)$$

TABLE A2. Parameter estimates: primary school  $k$  to secondary school  $k'$  transfer probability.

School SIMCE Scores <sup>a</sup> : $S_k$ vs. $S_{k'}$		Student SES <sub><math>i</math></sub> vs. HS SES <sup>s</sup> <sub><math>k'</math></sub>	
$(S_{k'} - S_S^*)(S_k - S_P^*)_+$	0.37 (0.02)	$I(\text{SES}_i = 1, \text{SES}_{k'}^s = 1)$	0.56 (0.04)
$(S_{k'} - S_S^*)(S_k - S_P^*)_-$	−0.25 (0.02)	$I(\text{SES}_i = 1, \text{SES}_{k'}^s = 3)$	−0.32 (0.07)
$I(\text{Type } 2)(S_{k'} - S_S^*)(S_k - S_P^*)_-$	0.67 (0.03)	$I(\text{SES}_i = 3, \text{SES}_{k'}^s = 1)$	−0.61 (0.03)
<sup>a</sup> : $S_P^*$ ( $S_S^*$ ): median SIMCE in primary (high) schools		$I(\text{SES}_i = 3, \text{SES}_{k'}^s = 3)$	0.66 (0.04)
Student GPA <sub><math>i</math></sub> vs. HS SIMCE <sup>b</sup>		Primary School SES <sub><math>k</math></sub> <sup>s</sup> vs. HS SES <sub><math>k'</math></sub> <sup>s</sup>	
$(S_{k'} - S_S^*)(\text{GPA}_i - M_{G^*})_+$	0.55 (0.01)	$I(\text{SES}_k^s = 1, \text{SES}_{k'}^s = 1)$	1.12 (0.04)
$(S_{k'} - S_S^*)(\text{GPA}_i - M_{G^*})_-$	0.35 (0.01)	$I(\text{SES}_k^s = 1, \text{SES}_{k'}^s = 3)$	−1.76 (0.09)
<sup>b</sup> : $M_{G^*}$ median grade 8 GPA across students		$I(\text{SES}_k^s = 2, \text{SES}_{k'}^s = 1)$	−0.14 (0.02)
Private high school		$I(\text{SES}_k^s = 2, \text{SES}_{k'}^s = 3)$	−1.84 (0.04)
$I(k' \text{ private})$	−0.72 (0.02)	$I(\text{SES}_k^s = 3, \text{SES}_{k'}^s = 1)$	−0.85 (0.03)
$I(k, k' \text{ both private})$	1.35 (0.03)	$I(\text{SES}_k^s = 3, \text{SES}_{k'}^s = 3)$	−0.46 (0.03)

<sup>45</sup>The 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 GPAs, one's beliefs  $\text{EA}_{it-1}, \text{VA}_{it-1}$  only depends on model parameters and one's type.

<sup>46</sup>The likelihood of observing a missing value of an outcome is 1 in nonapplicable cases. For example,  $O_{it} = \cdot$  if  $i$  is in jail from previous sentence;  $\text{GPA}_{it} = \cdot$  if  $e_{it} = 0$ .

TABLE A3. Other parameter estimates.

A. Initial SIMCE score		C. Grade-specific GPA constant	
Constant ( $\delta_0$ )	0.07 (0.01)	$\gamma_{05}$	−0.03 (0.005)
Ability ( $\delta_1$ )	0.95 (0.01)	$\gamma_{06}$	−0.04 (0.007)
$\sigma_\xi$	0.76 (0.01)	$\gamma_{07}$	−0.04 (0.008)
B. Terminal value function		$\gamma_{08}$	0 (normalized)
Grade level $G_{iT}$	0.0001 (0.12)	$\gamma_{09}$	0.02 (0.007)
$I(G_{iT} = 12)$	0.005 (0.42)	$\gamma_{10}$	0.05 (0.007)
$I(G_{iT} = 12)I(\text{Parent edu} = \text{mid})$	0.59 (0.13)	$\gamma_{11}$	−0.002 (0.007)
$I(G_{iT} = 12)I(\text{Parent edu} = \text{high})$	0.40 (0.17)	$\gamma_{12}$	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)		

$$\begin{aligned}
&= \Pr(O_{it}; \Theta \setminus \Theta_\chi | \bar{\Omega}_{\text{imt}}) \\
&= \begin{cases} \rho_{l_i} \Pr(e_{it}, d_{it} = 1; \Theta \setminus \Theta_\chi) L_{\text{imt}}^G(d_{it} = 1; \Theta_G, \Theta_a) & \text{if } \text{arrest}_{it} = 1, \\ \left( (1 - \rho_{l_i}) \Pr(e_{it}, d_{it} = 1; \Theta \setminus \Theta_\chi) L_{\text{imt}}^G(d_{it} = 1; \Theta_G, \Theta_a) \right. \\ \quad \left. + \Pr(e_{it}, d_{it} = 0; \Theta \setminus \Theta_\chi) L_{\text{imt}}^G(d_{it} = 0; \Theta_G, \Theta_a) \right) & \text{if } \text{arrest}_{it} = 0. \end{cases} \quad (24)
\end{aligned}$$

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

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

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

$$\begin{aligned}
&L_{\text{imt}}^G(d_{it}; \Theta_G, \Theta_a) \\
&= \phi\left(\frac{\text{GPA}_{it} - \overline{\text{GPA}}_{it} - \bar{a}_\chi}{\sqrt{\sigma_a^2 + \sigma_{\epsilon^{\text{gpa}}}^2}}\right) \times \left[ g_{it} \Pr(g_{it} = 1 | \text{GPA}_{it}, W_{k_{it}}, \chi_i = m) \right. \\
&\quad \left. + (1 - g_{it})(1 - \Pr(g_{it} = 1 | \cdot)) \right]. \quad (26)
\end{aligned}$$

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

$L_{\text{im}0}(\Theta_G, \Theta_a, \sigma_\xi)$  consists of the likelihood of one's pre- $G_0$  performance:

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

We assume that young children do not commit crimes before they enter our model, so  $L_{\text{imt}}^G(d_{it}; \Theta_G)$  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 ( $s_i$ ).

## REFERENCES

- Altonji, Joseph (1993), “The demand for and return to education when education outcomes are uncertain.” *Journal of Labor Economics*, 11 (1), 48–83. [395]
- Anderson, Mark (2014), “In school and out of trouble? The minimum dropout age and juvenile crime.” *Review of Economics and Statistics*, 96 (2), 318–331. [390]
- Arcidiacono, Peter (2004), “Ability sorting and the returns to college major.” *Journal of Econometrics*, 121 (1–2), 343–375. [395]
- Arcidiacono, Peter, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom (2016), “College attrition and the dynamics of information revelation?” Working paper, Duke University. [395, 398, 404]
- Aucejo, Esteban and Jonathan James (2019), “Catching up to girls: Understanding the gender imbalance in educational attainment within race.” *Journal of Applied Econometrics*, 34 (4), 502–525. [390]
- Becker, Gary (1968), “Crime and punishment: An economic approach.” *Journal of Political Economy*, 76 (2), 169–217. [390, 395]
- Billings, Stephen, David Deming, and Jonah 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. [389]
- Blumstein, Alfred, Jacqueline Cohen, Jeffrey Roth, and Christy Visser (1986), *Criminal Careers and “Career Criminals” ? Vols I and II*, Editors edition. National Academy Press, Washington, DC. [390]
- Bordon, Paola and Chao Fu (2015), “College-major choice to college-then-major choice.” *Review of Economic Studies*, 82 (4), 1247–1288. [395]
- Broidy, Lisa, Daniel Nagin, Richard Tremblay, John Bates, Bobby Brame, Kenneth Dodge, David Fergusson, John Horwood, Rolf Loeber, Robert Laird, Donald Lynam, Terrie Moffitt, Gregory Pettit, and Frank Vitaro (2003), “Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study.” *Developmental Psychology*, 39, 222–245. [390]
- Case, Anne and Lawrence Katz (1991), “The company you keep: The effects of family and neighborhood on disadvantaged youths.” NBER Working Paper No. 3705. [389]
- Clay, Karen, Jeff Lingwall, and Melvin Stephens (2012), “Do schooling laws matter? Evidence from the introduction of compulsory attendance laws in the United States.” NBER Working Paper 18477. [390]

Contreras, Dante, Paulina Sepulveda, and Sebastián Bustos (2010), "When schools are the ones that choose: The effects of screening in Chile." *Social Science Quarterly*, 91 (5), 1350–1368. [392]

Cook, Philip and Songman 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. [390]

Cortés, Tomás, Nicolás Grau, and Jorge Rivera (2020), "Juvenile incarceration and adult recidivism." Working paper, Universidad de Chile. [390]

Cullen, Julie, Brian Jacob, and Steven Levitt (2006), "The effect of school choice on participants: Evidence from randomized lotteries." *Econometrica*, 74 (5), 1191–1230. [413]

Cunha, Flavio, James Heckman, and Salvador Navarro (2005), "Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic Papers*, 57 (2), 191–261. [395]

Cunha, Flavio and James Heckman (2007), "The technology of skill formation." *The American Economic Review*, 97 (2), 31–47. [415]

Cunha, Flavio, James Heckman, and Susanne Schennach (2010), "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, 78 (3), 883–931. [415]

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall (2013), "Household choices and child development." *Review of Economic Studies*, 81 (1), 137–185. [415]

Deming, David (2011), "Better schools, less crime?" *Quarterly Journal of Economics*, 126 (4), 2063–2115. [414]

Di Tella, Rafael and Ernesto Schargrodsky (2004), "Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack." *American Economic Review*, 94 (1), 115–133. [389]

Díaz, Juan, Nicolás Grau, Reyes Tatiana, and Jorge Rivera (2021), "The impact of grade retention on juvenile crime." *Economics of Education Review*, 84. [392]

Donohue, John and Steven Levitt (2001), "The impact of legalized abortion on crime." *Quarterly Journal of Economics*, 116 (2), 379–420. [389]

Eren, Ozkan, Briggs Depew, and Stephen Barnes (2017), "Test-based promotion policies, dropping out, and juvenile crime." *Journal of Public Economics*, 153, 9–31. [398]

Eren, Ozkan, Michael Lovenheim, and Naci Mocan (2018), "The effect of grade retention on adult crime: Evidence from a test-based promotion policy." NBER working paper 25384. [398]

Freeman, Richard (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. [390]

Fu, Chao, Nicolás Grau, Jorge Rivera (2022), "Supplement to 'Wandering astray: Teenagers' choices of schooling and crime.'" *Quantitative Economics Supplemental Material*, 13, <https://doi.org/10.3982/QE1722>. [392, 400]

- Fu, Chao and Kenneth 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. [401]
- Gauri, Varun (1999), *School Choice in Chile: Two Decades of Educational Reform*. Pitt Latin American Studies. University of Pittsburgh Press. [391]
- Gaviria, Alejandro and Steven Raphael (2001), "School-based peer effects and juvenile behavior." *Review of Economics and Statistics*, 83 (2), 257–268. [389]
- Glaeser, Edward, Bruce Sacerdote, and Jose Scheinkman (1996), "Crime and social interactions." *Quarterly Journal of Economics*, 111 (2), 507–548. [389]
- Glaeser, Edward and Bruce Sacerdote (1999), "Why is there more crime in cities?" *Journal of Political Economy*, 107 (2), S225–S258. [389]
- Gottfredson, Michael and Travis Hirschi (1990), *A general theory of crime*. Stanford University Press, Stanford. [390]
- Heckman, James and Stefano Mosso (2014), "The economics of human development and social mobility." *Annual Review of Economics*, 6 (1), 689–733. [388]
- Hjalmarsson, Randi (2008), "Criminal justice involvement and high school completion." *Journal of Urban Economics*, 63, 613–630. [390]
- Hjalmarsson, Randi, Helena Holmlund, and Matthew Lindquist (2015), "The effect of education on criminal convictions and incarceration: Causal evidence from micro-data." *The Economic Journal*, 125 (587), 1290–1326. [390]
- Imai, Susumu and Kala Krishna (2004), "Employment, deterrence, and crime in a dynamic model." *International Economic Review*, 45 (3), 845–872. [390, 404]
- Jacob, Brian and Lars Lefgren (2003), "Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime." *American Economic Review*, 93 (5), 1560–1577. [389, 390]
- Katz, Lawrence, Steven Levitt, and Ellen Shustorovich (2003), "Prison conditions, capital punishment and deterrence." *American Law and Economics Review*, 5 (2), 318–343. [389, 390]
- Keane, Michael and Kenneth 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. [400]
- Kling, Jeffrey, Jens Ludwig, and Lawrence 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. [389]
- Lee, David and McCrary Justin (2005), "Crime, Punishment, and Myopia." NBER Working Paper No. 11491. [390]
- Levitt, Steven (1996), "The effect of prison population size on crime rates: Evidence from prison overcrowding litigation." *Quarterly Journal of Economics*, 111 (2), 319–351. [389]



- Levitt, Steven (1997), "Using electoral cycles in police hiring to estimate the effect of police on crime." *American Economic Review*, 87 (3), 270–290. [389]
- Levitt, Steven (1998), "Juvenile crime and punishment." *Journal of Political Economy*, 106, 1156–1185. [389, 390]
- Levitt, Steven and Lance Lochner (2000), *The Determinants of Juvenile Crime* (Jonathan Gruber, ed.), Risky Behavior Among Youths: An Economic Analysis. University of Chicago Press. [389, 407]
- Lochner, Lance (2004), "Education, work, and crime: A human capital approach." *International Economic Review*, 45 (3), 811–843. [390]
- Lochner, Lance (2011), "Education policy and crime." In *Controlling Crime: Strategies and Tradeoffs* (Philip Cook, Jens Ludwig, and Justin McCrary, eds.). University of Chicago Press. [390]
- Lochner, Lance and Enrico Moretti (2004), "The effect of education on crime: Evidence from prison inmates, arrests, and self-reports." *American Economic Review*, 94 (1), 155–189. [390]
- Machin, Stephen, Olivier Marie, and Suncica Vujic (2011), "The crime reducing effect of education." *The Economic Journal*, 121 (552), 463–484. [390]
- Mancino, Maria Antonella, Salvador Navarro, and David Rivers (2016), "Separating state dependence, experience, and heterogeneity in a model of youth crime and education." *Economics of Education Review*, 54. [390]
- Meckes, Lorena and Rafael Carrasco (2010), "Two decades of since: An overview of the national assessment system in Chile." *Assessment in Education: Principles, Policy and Practice*, 17 (2), 233–248. [392]
- Merlo, Antonio and Kenneth Wolpin (2015), "The transition from school to jail: Youth crime and high school completion among black males." *European Economic Review*, 79. [390]
- Mizala, Alejandra and Pilar Romaguera (2000), "School performance and choice: The Chilean experience." *The Journal of Human Resources*, 35 (2), 392–417. [391]
- Munyo, Ignacio (2015), "The juvenile crime dilemma." *Review of Economic Dynamics*, 18, 201–211. [390]
- Nagin, Daniel, David Farrington, and Terrie Moffitt (1995), "Life-course trajectories of different types of offenders." *Criminology*, 33, 111–139. [390]
- Nagin, Daniel and Kenneth Land (1993), "Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model." *Criminology*, 31, 327–362. [390]
- Nagin, Daniel and Raymond Paternoster (1991), "On the relationship of past to future participation in delinquency." *Criminology*, 29, 163–189. [390]

- Sampson, Robert and John Laub (1995), *Crime in the Making: Pathways and Turning Points Through Life*. Harvard University Press, Cambridge, MA. [390]
- Sickles, Robin C. and Jenny Williams (2008), “Turning from crime: A dynamic perspective.” *Journal of Econometrics*, 145, 158–173. [389]
- Stange, Kevin (2012), “An empirical examination of the option value of college enrollment.” *American Economic Journal: Applied Economics*, 4 (1), 49–84. [395]
- Valenzuela, Juan Pablo, Cristián Bellei, and Danae De los Rios (2014), “Socioeconomic school segregation in a market-oriented educational system: The case of Chile.” *Journal of Education Policy*, 29 (2), 217–241. [392]
- Wilson, James and Richard Herrnstein (1985), *Crime and Human Nature*. Simon & Schuster, New York, NY. [390]

---

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 14 September, 2020; final version accepted 1 September, 2021; available online 28 September, 2021.