Document Title: Essays in Applied Microeconomics

Author: Juan Pantano

Document No.: 223028

Date Received: June 2008

Award Number: 2006-IJ-CX-0002

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays in Applied Microeconomics

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Juan Pantano

2008

The dissertation of Juan Pantano is approved.

Daniel A. Ackerberg

Jinyong Hahn

Robert Denis Mare

Moshe Buchinsky, Committee Co-chair

V. Joseph Hotz, Committee Co-chair

University of California, Los Angeles

2008

To my Parents

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

Acknowledgments

I am deeply indebted to my advisors Moshe Buchinsky and V. Joseph Hotz for their continuous guidance and support. I am also thankful to Dan Ackerberg who was extremely helpful and patient. Jin Hahn provided useful suggestions at several stages in the dissertation. I have also benefited from the insightful remarks of Sandy Black, Leah Boustan, Dora Costa, Harold Demsetz, Paul Devereux, Fred Finan, Sebastian Galiani, Matt Kahn, Hugo Hopenhayn, Kathleen McGarry, Rob Mare, Maurizio Mazzocco, Judith Seltzer and Duncan Thomas. All remaining errors are mine. This dissertation research was partially funded with a Graduate Research Fellowship from the National Institute of Justice (Grant Number: 2006-IJ-CX-0002). Additional financial support from the UCLA Department of Economics and the California Center for Population Research are greatly appreciated. The second chapter of this dissertation is based on joint work with V. Joseph Hotz.

VITA

1976	Born, Buenos Aires, Argentina
1989–1994	Colegio Nacional de Buenos Aires, Buenos Aires, Argentina
1996–2000	B.A. (Economics), Universidad Torcuato di Tella, Buenos Aires, Argentina
1998–1999	Research Assistant, FIEL, Buenos Aires, Argentina
2001–2002	Junior Economist, FIEL, Buenos Aires, Argentina
2004–2007	Research and Teaching Assistant, Dept. of Economics, UCLA
2005	Pre-Doctoral Visiting Scholar, Dept. of Economics, Northwestern University
2005	C.Phil (Economics), UCLA

ABSTRACT OF THE DISSERTATION

Essays in Applied Microeconomics

by

Juan Pantano

Doctor of Philosophy in Economics

University of California, Los Angeles, 2008

Professor V. Joseph Hotz, Co-chair

Professor Moshe Buchinsky, Co-chair

This dissertation contains three essays that apply techniques in applied microeconomics to solve scientific puzzles and questions closely related to practical policy issues. The first essay explores the impact of early access to the birth control pill on the future crime rates of the children who are born to mothers who take advantage of this unprecedented improvement in contraceptive technology. The second essay investigates whether changing parenting strategies associated with parental reputation dynamics generate birth order effects in school performance. The last essay develops and estimates a dynamic model of human capital accumulation and criminal behavior. The estimated model is used to evaluate alternative criminal records policies and to shed light on the causal relationship between education and crime

CHAPTER 1

Unwanted Fertility, Contraceptive Technology and Crime: Exploiting a Natural Experiment in Access to the Pill

Donohue and Levitt (2001) claim to explain a substantial part of the recent decline in U.S. crime rates with the legalization of abortion undertaken in the early ’70s. While the validity of these findings remains heavily debated, they point to unwanted fertility as a potentially important determinant of a cohort’s criminality. In that spirit, I exploit a natural experiment induced by policy changes during the ’60s and ’70s. After the introduction of the contraceptive pill in 1960, single women below the age of majority faced restricted access to this new contraceptive method. Mostly as a by-product of unrelated policy changes, these access restrictions were lifted differentially across states during the ‘60s and ‘70s. This differential timing of contraceptive liberalization induces exogenous variation that can be used to identify the causal effect of unwanted fertility on crime. Results are consistent with the arguments of Donohue & Levitt. They indicate that greater flexibility to avoid unwanted pregnancies (through better contraceptive technology) reduces crime about two decades later, when undesired children would have reached their criminal prime.

1.1 Introduction

A blossoming literature in the U.S. examines the role of abortion legalization on the criminality of the cohorts born before and after this controversial law change. In the same spirit, I propose to exploit an alternative natural experiment induced by policy changes during the '60s and '70s during the "Contraceptive Revolution". In particular, after the introduction of the contraceptive pill in 1960, different states maintained some form of required parental consent to obtain a doctor's prescription for women below the age of majority. For a particular group of single women in their late teens, these restrictions were lifted differentially across states during the '60s and '70s. This differential timing of contraceptive liberalization induces exogenous variation that can be used to explore the causal link between unwanted fertility and crime. Greater flexibility to avoid unwanted pregnancies is likely to reduce crime two decades down the road, when undesired children born to these women would have reached their maximum criminal potential. In this hypothesis, "wantedness" is conceptualized as an overall indicator of willingness to invest resources in the future child. Rather than joining the already substantial literature in the abortion-crime debate, the contribution here explores the consequences of a set of completely unrelated policy changes which also induce exogenous variation in prevalence of unwantedness for a given birth cohort.

In addition to its scientific value as a potential determinant of a given birth cohort's criminality, understanding the causal link between unwanted fertility and criminality is relevant to policy makers. Potentially higher levels of criminality induced by more unwanted children is a cost that, in principle, should be taken into account when evaluating policies that restrict contraceptive freedom, or more generally, policies that limit women's ability to avoid unwanted children. In 2005-2006 there has been substantial policy debate over the apparent reluctance by the Federal Drug Administration to allow a new contraceptive device, the "day after" pill (Plan B) to be sold over the counter.

While most of the current debate centers on short run fears of increased teen promiscuity and the spread of sexually transmitted diseases, it is important to keep in mind the long run effects of a given contraceptive policy change.

The rest of the chapter is organized as follows. The next section provides some brief background on the institutional and legal history of the pill. Section 1.3 discusses related literature, causal mechanisms and necessary conditions for pill access to have a negative effect on future crime. Section 1.4 describes the data and Section 1.5 presents the basic empirical strategy, results and tests of the maintained hypothesis. A counterfactual policy extrapolation is conducted in Section 1.6. Conclusions follow.

1.2 Institutional Background

Here I provide a brief overview of the institutional and legal history associated with access to the pill.[1] The pill was introduced in the market in 1960 and quickly diffused among American women, becoming one of their preferred methods of contraception. However, underneath this "Contraceptive Revolution", the adoption of the pill as a contraceptive device by younger women faced a number of state-level legal obstacles. In particular, the pill was only available by prescription, and women below the age of majority required parental consent to receive medical services. During the '60s and '70s, different states liberalized their laws governing access to contraception for young women. This process was accomplished by state legislation that reduced the age of majority and granted mature minors capacity to consent to medical care. In some other states this liberalization took the form of judicial mature "minor" rulings or special family planning legislation. As shown in Table 1.1, the timing of this contraceptive liberalization was different for most states, spanning the period from 1960 to 1977.

[1] For more details see Goldin & Katz (2000, 2002), Hock (2005) and Bailey (2006)

Table 1.1: Access to Contraception Among Single Women in Late Adolescence 1960-1977

1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
ARIZONA IDAHO MONTANA NEVADA NORTH DAKOTA OKLAHOMA UTAH ALASKA	ILLINOIS				KENTUCKY OHIO					KANSAS MISSISSIPPI WASHINGTON	ALABAMA COLORADO CONNECTICUT GEORGIA MARYLAND NEW HAMPSHIRE NEW MEXICO NEW YORK NORTH CAROLINA OREGON PENNSYLVANIA TENNESSEE	ARKANSAS CALIFORNIA DELAWARE FLORIDA LOUISIANA MAINE MICHIGAN NEBRASKA RHODE ISLAND SOUTH CAROLINA SOUTH DAKOTA VERMONT VIRGINIA WEST VIRGINIA WISCONSIN	IOWA INDIANA NEW JERSEY TEXAS WYOMING	DISTRICT OF COLUMBIA MASSACHUSETTS MINNESOTA	HAWAII		MISSOURI

Table 1: Access to Contraception Among Single Women in Late Adolescence 1960-1977

The diagram shows the years in which women 18-19 years old first obtained access to the pill in each state. Hock (2005)

This latter fact induces plausibly exogenous cross-state variation over time that allows me to identify the causal effect of unwanted fertility on crime, in the same spirit of the abortion legalization arguments of Donohue & Levitt (2001). Moreover, note that young women being granted more unrestricted access to this effective contraception technology was by large a by-product of more general legislation drafted to address other unrelated policy concerns. Therefore, the usual threat of policy endogeneity does not appear to be particularly problematic in this context. Bailey (2006) makes a convincing case for the lack of policy endogeneity in the legislative and judicial process that leads, as unintended by-product, to contraceptive liberalization for unmarried teen women. Moreover, federal legislation prohibited individuals from obtaining oral contraceptives by mail shipped from other states. This greatly enhances the reliability of the proposed quasi-experimental design.

1.3 Related Literature

The idea that the levels of criminality of a given cohort can be traced back to how desired or "wanted" were births in that cohort has been around since the seminal contribution by Donohue & Levitt (2001) which exploited abortion legalization as a natural experiment to quantify this effect. In their initial article, Donohue & Levitt claimed that abortion legalization may account for as much as 50 % of the recent decline in crime rates in the U.S.

The pioneering work of Donohue & Levitt was followed by some critiques. In particular, Joyce (2004) casts doubts over the validity of these findings claiming that the authors failed to account for unobserved factors that might vary both across state and over time like the crack cocaine epidemic. A rejoinder by Donohue & Levitt (2004) argued that, if anything, failure to account for the crack epidemic biased the results against and not in favor of their 2001 findings. Other recent challenges to the findings of Donohue & Levitt (2001) include Foote & Goetze (2005), Sykes et al (2006) and Lott & Whitley (2006). A rejoinder by Donohue & Levitt (2006) and a more comprehensive methodological overview of the subject by Ananat et al (2006) address some of these recent challenges and, to some extent, confirm the provocative magnitudes of the 2001 article, although as in Foote & Goetze (2005), this recent work emphasize the fact that most of the effect is coming simply from declines cohort size as opposed to selection into the cohort.

While much has been written about the so-called "Contraceptive Revolution", the exogenous variation in the number of unwanted children induced by policy changes governing teen access to the pill has not been used to investigate the causal relationship between unwanted fertility and crime. The quasi-experimental variation induced by the differential timing of the contraceptive liberalization in different states has been exploited by some researchers to address other questions. In seminal work, Goldin & Katz (2000, 2002) exploited this variation to analyze the career and marriage decisions of women in the '60s and '70s, a period that witnessed substantial change in those dimensions. More recently, Hock (2005) and Bailey (2006) also exploited the variation available in state laws regarding access to the contraceptive pill. Hock (2005) concluded that by lowering the incidence of early fertility, unconstrained access to the pill increased the enrollment rate of college age women by almost 5 percentage points, and it had a less sizable but still positive and significant impact on college completion rates. Bailey (2006) found significant effects of the pill in women's child bearing timing and life cycle labor supply. In other recent contributions, Guldi (2005) examines the relative impacts of the pill and abortion on the fertility patterns of young women and Ananat & Hungerman (2006) explore how the pill changed the characteristics of the average mother.

Finally, the use of quasi-experimental variation in laws governing access to the pill for teen women is specially relevant in my context as there exists prolific literature relating teenage and out-of-wedlock fertility to the levels of criminality of the teenage and/or unmarried mother's offspring. For example, Grogger (1997) shows that young men who were born to young teen mothers are 3.5 percentage points more likely to be incarcerated than sons of older mothers. Hunt (2006) uses international victimization data to investigate the effects between teen fertility and crime and concludes that the high rates of teen births in the U.S. have prevented further declines in some types of crimes relative to other countries. Not surprisingly, criminologists have also looked into this question. Nagin, Farrington & Pogarsky (1997) use the Cambridge Study in Delinquent Development to examine alternative mechanisms or "accounts" through which teen fertility of the mother may have a significant effect in the delinquency levels of the children. They consider life course-immaturity, persistent poor parenting and diminished resources as alternative channels, finding some support for the latter two. More recently, Kendall & Tamura (2006) adopt a more historical, long run perspective to look at the effects of unmarried fertility on crime

1.3.1 Causal Mechanisms

Note that unwanted fertility is not likely to have a direct causal effect on crime. Rather, unwanted fertility will manifest itself as a cumulative process of disadvantage, starting right at the instant of conception. Those cumulated disadvantages are the ones that end up increasing criminal tendencies. While this chapter will not be focusing on disentangling these alternative contributing mechanisms, it is worth mentioning some of them. For example, the early harmful effects of being an unwanted child are likely to be channeled through inadequate prenatal care and child abuse and neglect.[2]. The impact

[2] For the impact of child abuse and neglect on future crime see Currie & Tekin (2006). For the relationship between unwanted fertility and inadequate prenatal care see Joyce & Grossman (1990)

of these initial disadvantages as well as the consequences of further underinvestments are likely to be experienced during childhood and early adolescence, therefore increasing the risk of delinquency onset. Note also that unwantedness might cause maternal risky behaviors during pregnancies. These behaviors are likely to lead to negative birth and infant health outcomes. Poor child health and low socio-emotional development are likely disadvantages to affect unwanted children.[3] Moreover, unintended children may, if born, stall maternal human capital accumulation by both, reducing the mother's formal educational attainment[4] and lowering her life-cycle labor force participation[5]. Unwantedness might lead not only to high incidence of child abuse and neglect but also reduce the levels of parental monitoring, control and supervision. This will certainly propel children's potentially deviant behavior. It could also be the case that unwanted children receive lower parental support (both in terms of time and money) for school. This is important because it is likely that lower education might itself lead to higher criminality.[6]

The impact of the pill might operate through channels other than the selection mechanisms discussed above. Indeed, the pill might reduce the criminality of *wanted* siblings through a "family size" effect. There exist evidence that the pill had an impact on completed fertility. Averted children were not compensated for at later stages of women's reproductive cycle. Therefore siblings of the these (unborn) unwanted children might benefit from a more abundant set of parental resources and also reduce their crime rates. Moreover, extending this argument to society at large, general equilibrium effects might operate through the smaller cohort sizes that pill access induces.

[3]See, for example, Joyce, Kaestner & Korenman (2000)

[4]See for example, Hock (2005) and Ananat & Hungerman (2006)

[5]See Bailey (2006) and Nagin et al. (1997)

[6]See Lochner & Moretti (2004)

In summary, there are many avenues through which higher levels of unwanted fertility can end up leading to higher crime rates. Moreover, many of these avenues or channels have feedback effects between them which will generally reinforce the link to a higher criminal propensity.

1.3.2 Necessary Conditions

Before describing the empirical strategy, it is important to establish whether two necessary conditions for the hypothesis in this chapter to be valid do in fact hold. First, pill access liberalization must lead to increased pill use. If, for whatever reason, access does not translate into actual use, the mechanism advanced in this chapter cannot be set in motion. Second, and most importantly, increased access must lead to a reduction in unwanted fertility. Regarding the first, Goldin & Katz (2002) provide evidence from the National Survey of Young Women showing that early legal access to the pill was indeed associated with greater pill use among young unmarried women. Regarding the second, an even more basic, question like ”Does improvement in contraceptive technology succeed in reducing fertility?” remains somewhat debated. Using a moral hazard argument the answer can be: may be not. Indeed, more available insurance provides an incentive to increase the activity level in the risky behavior, say, unprotected sex.[7]. In fact, some recent empirical evidence suggests that legalized abortion led to a significant increase in sexual activity.[8] If this increase in risky behavior is coupled with a failure of the insurance mechanism like, say, improper pill use, the result might be an increase, rather than a decrease in fertility.[9] Despite these appealing theoreti-

[7] See Akerlof, Yellen and Katz (1996)

[8] See Klick & Stratmann (2003)

[9] An alternative theoretical reason involving strong habit persistence induced by sexual debut is explored in a dynamic structural model of teen sex and contraception by Arcidiacono, Khwaja & Ouyang (2006).

cal arguments, the empirical evidence in Hock (2005), Bailey (2006) and Ananat & Hungerman (2007) is more consistent with the standard effect that can be expected a priori: Improved contraceptive technology leads to a decline in fertility.[10]

Finally, it must be noted that the pill made its initial impact mostly on women of advantaged backgrounds, a group that is less likely to generate criminals regardless of the wantedness status of their pregnancies. This would bias the results not in favor of but against finding a pill effect, as we would be mixing in this group for which the pill really does not matter with women of lower socioeconomic status for which the pill is more likely to make a difference.

1.4 Data

1.4.1 The Pill

As mentioned above, this chapter exploits data on the timing of contraceptive liberalization. In particular, I follow the classification adopted by Hock (2005) to identify the years in which single women 18-19 years old first obtained access to the pill. Hock's methodology differs slightly from the one adopted in the works of Goldin & Katz (2000, 2002) and Bailey (2006).[11]

1.4.2 FBI-UCR Data on Arrests

I compute the arrests per-capita for each age category using state level counts of arrests from the Uniform Crime Reports collected by the Federal Bureau of Investigations. In this chapter I work with a version of the UCR-FBI data maintained by the National

[10] See, however, Guldi (2005) for some evidence that access to both, abortion and pill contraception, actually increased the birth rate.

[11] For more details on these differences, see Hock (2005).

Consortium on Violence Research (NCOVR) at Carnegie Mellon. As pointed out by Maltz & Targonski (2002) FBI-UCR data should be used with caution, due to a number of data quality problems, especially at the county level. Note that these very same FBI-UCR data have been used by Donohue & Levitt (2001) in their much debated contribution.

Using these data I am able to observe the behavior of 33 cohorts. The youngest cohort (born in 1988) is 15 years old in the last year of the sample (2003). The oldest cohort (born in 1956) is 24 years old in the first year of the sample (1980). See Table 1.2. The last years of the sample do not provide interesting variation since cohorts who are 15-24 at that time have been mostly born under liberal contraceptive regimes, regardless of state of birth. This is so except for those in their 20s who were born in Missouri.

While most of the analysis is carried out with state level data, a more finely disaggregated version of the UCR-FBI data is later used to provide a test of the hypothesis linking early access to the pill to future crime.

1.5 Empirical Strategy

In principle, I could look at the aggregate state level crime rates. Then, I would estimate the following panel data model for the per capita crime rate

$$\frac{Crime_{st}}{Pop_{st}} = \beta\, D_{s,t-20} + \lambda_s + \lambda_t + \varepsilon_{st} \tag{1.1}$$

where the dependent variable is the per capita number of crimes in state s and time t, λ_s and λ_t denote state and year specific effects and $D_{s,t-20}$ is a dummy variable indicating whether a liberal contraceptive policy was in place, say 20 years before t.

Now, if the pill is responsible for the reduction in crime, we should observe a decline in the crime rates of those cohorts born under the liberal regime only. The lack of state level crime data by age of the criminal prevents me from testing this hypothesis directly. I therefore turn to FBI-UCR arrest data and estimate the following model for the number of arrests per capita, using age-state-year cells as the unit of observation.

$$\frac{Arrests_{ast}}{Pop_{ast}} = \beta Pill_{t-a-1,s} + \lambda_a + \lambda_s + \lambda_t + \varepsilon_{ast} \tag{1.2}$$

where $a = 15, 16, ..., 24$ indexes single year of age categories, $s = 1, 2, ..., 51$ indexes states and $t = 1980, ..., 2003$ indexes years. λ_t denote year specific effects that capture any national pattern in the time series of percapita arrests which is common across states and age categories. λ_s denote state effects that capture time invariant, unobserved state level characteristics that might affect the arrest rate. Finally, λ_a denote age effects that non-parametrically account for the crime-age profile, one of the most firmly established hard facts in criminology. More importantly, given data constraints (i.e. the fact that FBI arrest data by age is only available from 1980 onwards)

I do not observe the arrest rates for cohorts 5 to 9 before 1980, when their ages range from their mid to their late teens. See Table 1.2.

$Arrests_{ast}$ and Pop_{ast} denote the counts of arrests and population size for individuals of age a in state s in year t. $Pill_{t-a-1,s}$ is a binary indicator which is equal to one if the specific age-state-year combination implies that those individuals were born under a liberal contraceptive regime. In other words, the policy variable $Pill_{t-a-1,s}$ indicates whether a particular cohort that happens to be a years old at calendar year t in state s was born in a state-time combination that allowed single women 18-19 years old to obtain a prescription for contraceptive pills without parental consent.

The coefficient β measures the causal effect of teen access to the pill on the number of arrests per capita. With an estimate of β at hand, back of the envelope calculations can be done to derive an aggregate effect of the pill.

Table 1.2: Cohort Structure of NCOVR Data

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1956	1																										
1957	2	1																									
1958	3	2	1																								
1959	4	3	2	1																							
1960	5	4	3	2	1																						
1961	6	5	4	3	2	1																					
1962	7	6	5	4	3	2	1																				
1963	8	7	6	5	4	3	2	1																			
1964	9	8	7	6	5	4	3	2	1																		
1965	10	9	8	7	6	5	4	3	2	1																	
1966	11	10	9	8	7	6	5	4	3	2	1																
1967	12	11	10	9	8	7	6	5	4	3	2	1															
1968	13	12	11	10	9	8	7	6	5	4	3	2	1														
1969	14	13	12	11	10	9	8	7	6	5	4	3	2	1													
1970	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1												
1971	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1											
1972	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1										
1973	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1									
1974	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1								
1975	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1							
1976	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1						
1977	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1					
1978	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1				
1979	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1			
1980	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1		
1981	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2		
1982	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3		
1983	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4		
1984	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5		
1985	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6		
1986	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7		
1987	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8		
1988	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9		
1989		33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10		
1990			33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11		
1991				33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12		
1992					33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13		
1993						33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14		
1994							33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15		
1995								33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16		
1996									33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17		
1997										33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18		
1998											33	32	31	30	29	28	27	26	25	24	23	22	21	20	19		
1999												33	32	31	30	29	28	27	26	25	24	23	22	21	20		
2000													33	32	31	30	29	28	27	26	25	24	23	22	21		
2001														33	32	31	30	29	28	27	26	25	24	23	22		
2002															33	32	31	30	29	28	27	26	25	24	23		
2003																33	32	31	30	29	28	27	26	25	24		

As explained above, given data limitations, results in following sections will be all in terms of arrests. It would be interesting to extend these results and look at the impact of the pill in actual crime rates since only a very small fraction of crimes end up in an arrest. While there is no reason to believe that the pill might have had an impact on the arrests-to-crimes ratio, I am ultimately interested in understanding the impact of unwanted fertility on crime, so further research is necessary to confirm that the results on arrests in following sections hold robust when the actual outcome is more directly related to the level of criminal activity.

Finally, note that the quasi-experimental variation is over a relatively small group, namely, single women who are 18 or 19 years old. These women account for a relatively small fraction of births in a given birth cohort. Since I am not able to distinguish who among the arrested individuals was born to a single 18 or 19 years old mother, I cannot look at the impact on the arrest rates for that ideal group, say, $\frac{Arrests_{ast}^{s,18-19}}{Pop_{ast}^{s,18-19}}$ where $Arrests_{ast}^{s,18-19}$ and $Pop_{ast}^{s,18-19}$ would denote the counts of arrests and population size for individuals of age a in state s in year t *who were born to single mothers 18 or 19 years* old at the time of conception. However, under mild assumptions, it can be shown that my estimate of β will recover a lower bound (in absolute magnitude) for the true causal effect of the pill on the arrest rates for this unobserved group.

Indeed, let α_{ast} denote the fraction of births due to single, 18 and 19 years old mothers $(\text{Births}_{s,t-a}^{s,18-19})$ taken relative to the total number of births $(\text{Total Births}_{s,t-a})$

$$\alpha_{ast} = \frac{\text{Births}_{s,t-a}^{s,18-19}}{\text{Total Births}_{s,t-a}}$$

Then, I can always decompose $\frac{Arrests_{ast}}{Pop_{ast}}$ as

$$\frac{Arrests_{ast}}{Pop_{ast}} = \alpha_{ast}\left[\frac{Arrests_{ast}^{s,18-19}}{Pop_{ast}^{s,18-19}}\right] + (1-\alpha_{ast})\left[\frac{Arrests_{ast}^{\tilde{}(s,18-19)}}{Pop_{ast}^{\tilde{}(s,18-19)}}\right]$$

where $Arrests_{ast}^{\tilde{}(s,18-19)}$ and $Pop_{ast}^{\tilde{}(s,18-19)}$ denote the count of arrests and population size for individuals of age a in state s in year t who were born to mothers who *were not* single 18-19 years old at the time of conception.

Then consider the following two population regression functions

$$\frac{Arrests_{ast}^{s,18-19}}{Pop_{ast}^{s,18-19}} = \beta^{*} Pill_{t-a-1,s} + \lambda_a + \lambda_s + \lambda_t + \eta_{ast} \tag{1.3}$$

$$\frac{Arrests_{ast}^{\tilde{}(s,18-19)}}{Pop_{ast}^{\tilde{}(s,18-19)}} = \beta^{\tilde{}} Pill_{t-a-1,s} + \lambda_a + \lambda_s + \lambda_t + \eta_{ast}^{\tilde{}} \tag{1.4}$$

Now, if there are no family size or cohort size effects, all the impact of the pill will be channeled trough a selection mechanism that will only impact the crime rates of those born to single 18-19 year old mothers and therefore we have $\tilde{\beta} = 0$. Then multiplying the first equation by α_{ast} and the second one by $1-\alpha_{ast}$ and adding the two we get the regression function that I can actually estimate with the available data, namely,

$$\frac{Arrests_{ast}}{Pop_{ast}} = \beta^{*}\alpha_{ast}Pill_{t-a-1,s} + \lambda_a + \lambda_s + \lambda_t + \varepsilon_{ast} \tag{1.5}$$

If $\alpha_{ast} = \alpha$ then my estimate $\widehat{\beta}$ will be consistently estimating $\alpha\beta^{*}$, a loose lower bound for the causal parameter β^{*} given that $\alpha < 1$ by construction and indeed, only about 0.07 overall in the estimating sample. Moreover, since access to the pill will have an impact on α_{ast} we can relax the above assumption and let $\alpha_{s,t-a} = \alpha + \delta Pill_{t-a-1,s} + \nu_{s,t-a}$ with $\delta < 0$. It can be shown that in this case my estimate $\widehat{\beta}$ will be consistently estimating an even less tight lower bound for the causal parameter of interest β^{*}. Indeed, $\widehat{\beta}$ will be consistent for $\beta^{*}(\alpha+\delta)$,with $0 < (\alpha+\delta) < 1$ and $\alpha+\delta$ close to zero given $\alpha \approx 0.07$ and $\delta < 0$

1.5.1 Basic Estimates

Table 1.3 shows the baseline results. I estimate equation (1.2) by simple OLS. Column 1 shows that the coefficient for β is negative and significant with a point estimate of -0.004.

Noting that the dependent variable on arrests is in annual per-capita terms, the magnitude of this estimated negative causal effect is not minor. For example, for California, this translates into 450000 $\times$ 0.004 = 1800 fewer arrests on average for each year and each age category. Moreover, if we take into account that arrests are only the tip of the iceberg when it comes to measuring the extent of criminal activity, the impact of the pill cannot be understated.

I explore the robustness of this result to two adjustments that deal with some of the limitations of the data used in this article. First, I am able to observe neither the month of the arrest nor the month of birth of the arrested person. Therefore, while $t-a-1$ is most likely the year in which the arrested individual was conceived, it is possible that conception took place on year $t-a-2$ or, less likely, $t-a$. Assuming that births and arrests are uniformly distributed across the calendar year and that all pregnancies end up in births after the normal 9 months period, I construct an alternative indicator of pill access as

$$Pill_{ast} = \left(\frac{9}{24}\right) Pill_{t-a-2,s} + \left(\frac{12}{24}\right) Pill_{t-a-1,s} + \left(\frac{3}{24}\right) Pill_{t-a,s} \tag{1.6}$$

I then estimate equation (1.2) using $Pill_{ast}$ as defined above instead of $Pill_{t-a-1,s}$.

Another implicit assumption maintained in the previous section is that the state of arrest is the same as the state of birth for all individuals contributing to the aggregate arrest data. But this is not likely to be the case. While it is hard to imagine that the cross-state migration pattern would be systematic in a particular way that might threaten the causal interpretation of the pill effect, internal migration could affect the previous results. Note that so far I am abstracting away from internal migration by assuming that all the good or bad consequences of contraceptive liberalization will be felt within the state that adopts the policy change. In particular, I am assuming that arrested individuals were born in the same state that they are arrested. Problems might arise if states with early liberalization have a systematically different pattern of migration into or out of the state relative to states with late liberalization. Donohue & Levitt (2001) faced similar concerns and showed that their results hold robust when adjusting for cross-state mobility. If measurement error is classical, attenuation bias resulting from state mis-classification would bias results against the hypothesis that access to the pill leads to future declines in the arrest rate, implying that the estimated magnitude is a lower bound (in absolute value).[12]

In order to address this issue, I use the 1980, 1990 and 2000 decennial censuses' microdata to compute state of birth probabilities, conditional on state of residence at any age (15-24) for each year.[13] With these probabilities at hand, the adjustment is relatively straightforward. I replace the raw policy indicator $Pill_{t-a-1,s}$ with a weighted version of it,

$$Pill^{W}_{t-a-1,s} = \sum_{s'} p_{at}\left(s'|s\right) Pill_{t-a-1,s'} \tag{1.7}$$

where $p_{at}\left(s'|s\right)$ are the conditional probabilities coming from the appropriate age- and year-specific state-of-birth / state-of-residence transition matrix.

Table 1.3: The Effect of Early Access to the Pill on Future Arrests

	Baseline	Alternative Birth Window	Cross State Mobility
Pill Access	-0.004	-0.005	-0.016
	[0.001]***	[0.001]***	[0.002]***
State effects?	YES	YES	YES
Year Effects?	YES	YES	YES
Age Effects?	YES	YES	YES
Observations	10200	10200	10200
R-squared	0.43	0.43	0.43

Robust standard errors in brackets

*significant at 10%; ** significant at 5%; *** significant at 1%

[12]Measurement error might not be classical, though. See Heckman, Farrar and Todd (1996) for an example of the consequences of non-classical measurement error and selective migration for the analysis of state-of-birth/state-of residence transitions.

[13]I use the PUMS microdata to compute these migration transition matrices for 1980,1990 and 2000 and impute the values for intervening years by interpolation.

Table 1.3 shows the results of the two adjustments described above. Column (1) shows the baseline estimate. As can be seen in column (2), the effect of the pill is robust to an alternative definition of pill access that takes into account the likelihood of conception at the two adjacent years. Column (3) shows that the effect of the pill is up to 4 times higher in magnitude when the adjustment for cross-state mobility is implemented by using the weighted pill indicator described in (1.7)

1.5.2 Abortion

Note that when abortion becomes legal the treatment effect provided by access to the pill is not the same. It is less powerful because it implies less of a change in the ”possibility frontier” to avoid unwanted children. In the same vein, it would be interesting to check whether the results of Donohue & Levitt (2001) are actually picking up part of the pill effect and verify whether results from the previous section on the impact of the pill stand robust when controlling for abortion legal status. Note that the pattern of abortion legalization might be correlated with the process of contraceptive liberalization, say, for political reasons at the state level.

Five states legalized abortion in 1970. These ”early legalizers” provide the variation necessary to identify the impact of abortion on future crime. Abortion becomes legal in the rest of the United States by way of the famous Supreme Court ruling in Roe v. Wade in 1973. I construct an indicator for the availability of legal abortion in the same way I constructed my pill access indicator.

$LegalAbort_{t-a-1,s}$ is a binary indicator which is equal to one if the specific age-state-year combination implies that those individuals were likely to be born under a regime in which abortion was already legal.

To maximize comparability with the results from Donohue & Levitt (2001) I restrict the sample to the same period (1985-1997) used by these authors.[14] Then, I augment the model in (1.2) by including the indicator for legal abortion.

$$\frac{Arrests_{ast}}{Pop_{ast}} = \beta\ Pill_{t-a-1,s} + \gamma\ LegalAbort_{t-a-1,s} + \lambda_a + \lambda_s + \lambda_t + \varepsilon_{ast} \qquad (1.8)$$

Table 1.4 reports the results from estimating Equation (1.8).

Table 1.4: The Effect of Early Access to the Pill and Abortion Legalization on Future Arrests

	1	2	3
Pill Access	-0.007		-0.005
	[0.002]***		[0.002]***
Legal Abort?		-0.009	-0.008
		[0.002]***	[0.002]***
State effects?	YES	YES	YES
Year Effects?	YES	YES	YES
Age Effects?	YES	YES	YES
Observations	6630	6630	6630
R-squared	0.49	0.49	0.49

Robust standard errors in brackets
*significant at 10%; ** significant at 5%; *** significant at 1%

In column (1) we corroborate that the results for the pill hold robust to the new sample period. The coefficient is now higher in magnitude (-0.007) and still significantly negative. Column (2) seems to replicate the well known results of Donohue & Levitt: legal abortion is significantly associated with substantial declines in the future

[14]However, as shown below in Table 1.5, these results stand robust when using the full sample and controlling for state-year effects.

rate of arrests per capita.[15] Finally, the model in column (3) includes both policy indicators simultaneously. Both coefficients are slightly smaller in magnitude relative to columns (1) and (2) but remain negative and significant indicating that both, abortion legalization and contraceptive technology, are valid and quantitatively important channels through which reductions in unwanted fertility yield crime declines in the long run. It is surprising however that magnitudes are so similar because the impact of the pill measures a treatment effect on late teen women only, while abortion legalization affects mothers of all ages.[16] In principle, one would expect the magnitude of the latter to be many times larger.

1.5.3 State-Year Effects

In this subsection I address the potential skepticism that may arise, as in the abortion-crime debate, regarding the causal nature of the previous results. In particular, despite the experimental flavor of the research design, it might be the case that by pure chance, there are some other factors operating at the state level that might generate a spurious correlation between pill access and future crime. I therefore turn to a more demanding identification strategy in which I exploit the single year of age dimension of the data to allow for a full set of state-year effects. These state-year effects can account for any state-specific phenomena that is responsible for fewer arrest in specific years during the '80s and '90s and that might be unfortunately correlated with the timing of pill access across states in the '60s and '70s, thus confounding the estimation of the parameter of interest. The following specification is more stringent in the sense that the variation

[15] This replication is not exact, though, because Donohue & Levitt use effective abortion rates rather than a simple dummy variable on whether abortion is legal or not.

[16] It is difficult to measure the impact of the pill on mothers other than 18-19 because in that case the empirical strategy would have to rely only on "before-and-after" designs around 1960. The usual caveats for inference with this type of design would then apply.

left in the data to identify the causal parameter is much smaller. Specifically, I estimate a more saturated model given by:

$$\frac{Arrests_{ast}}{Pop_{ast}} = \beta Pill_{t-a-1,s} + \gamma LegalAbort_{t-a-1,s} + \lambda_{st} + \lambda_{as} + \lambda_{at} + \varepsilon_{ast} \tag{1.9}$$

where λ_{st} denote state-year effects, λ_{as} denote age-state effects and λ_{at} denote age-year effects. Table 1.5 shows the results of estimating equation (1.9).

Table 1.5: The Effect of Early Access to the Pill on future Arrests Controlling for Abortion Legalization and State-Year Effects

	Basic	Controlling for Abortion and State-Time Effects			
	1	2	3	4	5
Pill Access	-0.004 [0.001]***	-0.011 [0.001]***	-0.006 [0.001]***	-0.007 [0.001]***	-0.002 [0.001]**
Legal Abort?		-0.004 [0.001]***	-0.007 [.0032]**	-0.006 [0.001]**	-0.008 [0.001]***
State effects?	YES	YES	YES	YES	YES
Year Effects?	YES	YES	YES	YES	YES
Age Effects?	YES	YES	YES	YES	YES
State-Year Effects?	NO	YES	YES	YES	YES
Age-Year Effects?	NO	NO	YES	NO	YES
State-Age Effects?	NO	NO	NO	YES	YES
Observations	10200	10200	10200	10200	10200
R-squared	0.43	0.78	0.80	0.93	0.95

Robust standard errors in brackets
*significant at 10%; ** significant at 5%; *** significant at 1%

Column (2) shows that the causal effect of the pill is still statistically and economically significant under the more stringent identification strategy that controls for state-year effects. Moreover, as shown in Columns (3)-(5) the effect remains significant when controlling, in addition, for a full set of state-age and year-age effects that allows the crime-age profile to flexibly vary by state and year. The effect of the pill remains significant, but smaller in magnitude, even in the fully saturated model that includes all the possible interactions and puts the most pressure on the data.

1.5.4 Tests

In this subsection I provide two tests of the proposed causal link between early teen access to the pill and future crime.

1.5.4.1 Relative size of population at risk of treatment

The results so far suggest the existence of a causal link between access to the pill and later crime. However, it would be reassuring to subject these results to further scrutiny in order to provide more credibility to the findings in previous sections. I use data from decennial population censuses to construct a measure of the relative size of the population at risk of treatment. Let $F^{18-19}_{t-a-1,s}$ be the proportion of females who were 18 or 19 years old in state s at time $t-a-1$. Let this proportion to be taken with respect to the total number of female residents of state s in the reproductive age range, say 15-44.[17] I augment the basic model by including this measure of relative size of the population at risk. Moreover, I interact this share with the policy indicator, $Pill_{t-a-1,s}$. If access to the pill is what really drives down crime two decades later, we should expect a more sizeable negative causal effect in those states with a higher fraction of the population at risk of treatment. In other words, the interaction between the fraction of women 18-19 years old and the policy indicator for pill access, should be negative.

[17] To compute $F^{18-19}_{t-a-1,s}$ for years after 1969 I rely on estimates from the Surveillance Epidemiology and End Results (SEER) Program at the National Cancer Institute. I interpolate the years between 1956 and 1968 exploiting the 1950, 1960 and 1970 decennial censuses.

This would provide a further test that the proposed channel is the one actually driving the results. The extended specification would be

$$\begin{aligned}\frac{Arrests_{ast}}{Pop_{ast}} &= \beta Pill_{t-a-1,s} \\ &+\delta_0 F^{18-19}_{t-a-1,s} + \delta_1 \left(F^{18-19}_{t-a-1,s} \times Pill_{t-a-1,s}\right) \\ &+\lambda_{st} + \lambda_{as} + \lambda_{at} + \varepsilon_{ast}\end{aligned} \tag{1.10}$$

where $F^{18-19}_{t-a-1,s}$ is the proportion of women who were 18-19 years old when the cohort which is at age a in state s and time t was conceived. If the results of this test are to be supportive of the unwanted fertility story we expect the coefficient δ_1 on the key interaction term in (1.10) to be negative and statistically significant. This would imply that the effect of the pill was stronger in those states where the relative size of the treatment group was bigger. Similar tests could be conducted with the proportion of single 18-19 females or the fraction of births due to single mothers who were 18-19 years old at the time they got pregnant, say $B^{18-19}_{t-a-1,s}$. A caveat on the validity of this latter test might arise if we allow for the possibility that higher levels of teen fertility across states do not really reflect higher levels of unwantedness. In other words, unmarried teen fertility in Mississippi might be much higher than in California but still the fraction of unwanted births could be lower in the former state than in the latter. Moreover, marital status and fertility are choices that are affected by the policy variation of interest thus inducing potential post-treatment bias in estimation. Therefore I rely on the more crude but cleaner test that relies only on the relative age structure of the female population, using $F^{18-19}_{t-a-1,s}$, which can be considered predetermined.

The impact of the pill is then given by $\beta + \delta_1 F^{18-19}_{t-a-1,s}$. Table 1.6 presents the results of the test. In columns 2 and 4 both the interaction term and the main effect become not significant. However, specifications in Columns 1 and 3 show that the key interaction term, δ_1 is negative and significant. Noting that the variable $F^{18-19}_{t-a-1,s}$ranges from 0.05 to 0.11 over the sample period, the total effect is negative for $F^{18-19}_{t-a-1,s} > 0.07$ in model (1) and $F^{18-19}_{t-a-1,s} > 0.065$ in model (3)

Table 1.6: Size of Treatment Group and the Impact of the Pill on Future Arrests

	1	2	3	4
Pill Access	0.022 [0.012]*	-0.006 [0.013]	0.018 [0.007]***	-0.006 [0.007]
$F^{18\text{-}19}$ x Pill Access	-0.334 [0.138]**	0.096 [0.152]	-0.290 [0.076]***	0.076 [0.078]
State-Year effects ?	YES	YES	YES	YES
Age-Year effects ?	NO	YES	NO	YES
State-Age effects ?	NO	NO	YES	YES
Observations	10170	10170	10170	10170
R-squared	0.78	0.80	0.93	0.95

Note: Robust standard errors in brackets. All models include state-year effects and control for abortion legal status. Pill Access and $F^{18\text{-}19}$ are adjusted for cross-state mobility. * significant at 10%; ** significant at 5%; *** significant at 1%

1.5.4.2 Geographic Spillovers in Access to the Pill and Criminal Activity

It is possible that geographic spillovers in access to the pill and criminal activity exist. The most extreme example of the first type of spillover is given by single teen women living in St. Louis, Missouri, west of the Mississippi. While Illinois liberalized access in 1961, Missouri was the last state to do so in 1977 (See Table 1.1). This creates 16 years of lag in the timing of pill access liberalization within a few miles. Researchers who have investigated the impact of abortion legalization on fertility have addressed similar concerns. In particular, Blank et al (1996) and Levine et al. (1996) emphasize the importance of taking into account cross-state traveling when assessing the effects abortion legalization. On the other hand, this should be less of a concern in the case of the pill because it would require teens to regularly drive out-of-state for checkups and refillings. This would entail a much greater cross-state travel burden relative to the case of abortion which only involves a single trip. Geographic spillovers in criminal activity are also relevant in my context. They involve state criminals residing close to a state boundary and crossing state lines to commit crimes in a nearby out-of-state city. As explained below, I can exploit the testable implications of these spillovers to provide further causal evidence for the link between pill access and crime.

In this section I turn to arrest data from a finer level of geographic disaggregation: metropolitan statistical areas. Crime is, by far, an urban problem. Then, it's not surprising that most of each state's crime is actually committed in the corresponding metropolitan areas. Having this additional margin of variation within states allows me to explore the issue of geographic spillovers in more detail. In particular, these data allow me to compute distances to the nearest neighboring state in which the pill is available. This strategy provides an alternative and potentially helpful source of variation when testing the effects of access to the pill on future crime.[18]

I consider the following model for the number of arrests per capita in age category a, in metropolitan area m within state s, at time t.[19]

$$\begin{aligned}\frac{Arrests_{amst}}{Pop_{amst}} &= \beta Pill_{t-a-1,s} \\ &\quad + \gamma[1 - Pill_{t-a-1,s}]Dist_{t-a-1,m} \\ &\quad + \lambda_a + \lambda_m + \lambda_t + \varepsilon_{amst}\end{aligned} \tag{1.11}$$

[18] Alternatively, one could compare focal states which are surrounded by states with similar policy timing or, more formally, use a spatial model.

[19] I exclude metropolitan areas that cross state borders from the analysis.

where

$$Dist_{t-a-1,m} = \min_{c \in D^*_{t-a-1}} d(m,c)$$

with

$$D^*_{t-a-1} = \left\{s : Pill_{s,t-a-1} = 1\right\}$$

and $d(m,c)$ denotes the geographic distance between metropolitan area m and a county c. Distance minimization is then conducted between a given metropolitan area and the counties belonging to any of the states in the set of states with liberal contraceptive regimes at time $t-a-1$, namely D^*_{t-a-1}.[20]

Table 1.7 presents the results of estimating the model in equation (1.11). We observe that the coefficient γ on the key interaction term $[1-Pill_{t-a-1,s}]\,Dist_{t-a-1,m}$ is positive across specifications. Note that for metropolitan areas in states that by year $t-a-1$ still remain in with conservative contraception regimens $[1-Pill_{t-a-1,s}]\,Dist_{t-a-1,m}$ captures the distance to the closest county with liberal contraception. Since there are no policy reversals, this distance always declines over time as additional states switch from conservative to liberal contraceptive regimes. A by-product of these switches is that they make the distance to liberal contraception closer for those metropolitan that remain in conservative states. Then, it is easier to interpret γ as the impact of declines in this distance. A positive γ implies that declines in the distance to liberal contraception lead to declines in the (future) arrest rate.[21] It is hard to imagine an alternative story to rationalize why the number of arrests per capita would be smaller for some MSAs in such a precise spatial pattern if the timing of pill access is not the one to blame. The fact that γ is positive and significant is consistent with the main-

[20] I am thankful to Leah Boustan and the Minnesota Population Center who kindly provided data and codes to compute these distances.

[21] Noting that the distance is measured in miles, the magnitude of the interaction term is small. It is left for future research to investigate whether these magnitudes are consistent with findings in spatial criminology.

tained hypothesis relating early access to the pill and future crime. If γ is positive, for a metropolitan area in non-liberal state, declines in the distance to a liberal state are associated with declines in the own number of future arrests. This finding implies that the contraceptive liberalization in an adjacent state will bring down future crime in a non-liberalizing state too, specially in metropolitan areas close to the boundary between the two states.

Table 1.7: The Effect of Early Access to the Pill on Future Arrests. Metropolitan Areas. Dependent Variable: Arrests per capita

	1	2	3
Pill Access	0.004**	0.006***	0.004*
	[0.002]	[0.001]	[0.002]
[1-Pill Access]*Dist	0.012***	0.007***	0.006***
	[0.003]	[0.001]	[0.002]
MSA effects?	YES	YES	YES
Year Effects?	YES	YES	YES
Age Effects?	YES	YES	YES
MSA x Year Effects?	NO	YES	YES
MSA x Age Effects?	NO	NO	YES
Observations	34711	34711	34711
R-squared	0.74	0.87	0.88

Robust standard errors in brackets. *significant at 10%; ** significant at 5%; *** significant at 1%. The units used in the distance measure are miles. The interaction coefficient and its standard error have been multiplied by 1000000.

It should be stressed that these findings are consistent with cross-state travel for the pill in the '60s and '70s but it is even more likely that an alternative mechanism is at play: The more "wanted" cohorts born in the adjacent liberalizing states will not be crossing the state line to commit crimes that often two decades later (in the '80s and '90s). However, regardless of the mechanism at play, this evidence is at least suggestive that the pill is really driving future crime down.

1.6 Counterfactual Policy Extrapolation

Consider the following hypothetical scenario: Suppose unrestricted access to the Pill is granted across the board in 1960. We expect the improved wantedness level to induce lower criminality in cohorts born after 1960. How quantitatively important is this effect? How many arrests would have not taken place?

Integrating over ages, years and states, we can compute the counterfactual change in the number of arrests during the period according to the proposed scenario as:

$$\sum_{s=1}^{51}\sum_{t=1980}^{2003}\sum_{a=15}^{24} Pop_{ast}\left(1-D_{t-a,s}\right)\widehat{\beta} \tag{1.12}$$

This simple back of the envelope calculation shows that a counterfactual scenario in which every state grants immediate unrestricted pill access to single teen women in 1960 is consistent with approximately 2 million fewer arrests in the period 1980-2003. To put this number in context, note that over the same period, there are about 97 million arrests reported in the FBI-UCR data. Therefore, the total impact would have been slightly over 2 %. Assuming a crime-to-arrests ratio of 5, about 10 million crimes would have been avoided over the period.

1.7 Conclusions

The evidence presented in this chapter shows that increased flexibility to avoid unwanted pregnancies reduce crime two decades into the future, when cohorts born in more liberal contraceptive regimes reach their criminal prime. These results hold in different samples and stand robust to several adjustments.

While further testing and sensitivity analysis is warranted to place more confidence in these findings, it seems possible to extend the abortion-crime arguments to policies other than abortion legalization, as long as these other policies (i.e. family planning and contraception) also reduce the level of unwanted fertility. However, while results suggest that a selection mechanism is at play, further research is needed to quantify the magnitude of ”family size”, ”cohort size” and ”selection” channels.

1.8 Bibliography

[1] Akerlof, Yellen & Katz (1996) ”An Analysis of Out-of Wedlock Childbearing in the United States”, Quarterly Journal of Economics

[2] Ananat, Gruber, Levine & Staiger (2006), ”Abortion and Selection”, mimeo.

[3] Ananat & Hungerman (2006), ”The Power of The Pill for the Next Generation”, mimeo.

[4] Arcidiacono, Khwaja & Ouyang (2007) ”Habit Persistence and Teen Sex: Could Increased Access to Contraception have Unintended Consequences for Teen Pregnancies?”, mimeo.

[5] Bailey (2006) “More Power to the Pill: The Impact of Contraceptive Freedom on Women’s Life Cycle Labor Supply”, Quarterly Journal of Economics

[6] Blank, George & London (1996), ”State abortion rates. The impact of policies, providers, politics, demographics, and economic environment”, Journal of Health Economics.

[7] Currie & Tekin (2006) " Does Child Abuse Cause Crime? " NBER Working Paper 12171

[8] Donohue & Levitt (2001) "The Impact of Legalized Abortion on Crime", Quarterly Journal of Economics

[9] Donohue & Levitt (2003) "Further Evidence that Legalized Abortion Lowered Crime: A Reply to Joyce", NBER Working Paper 9532.

[10] Donohue & Levitt (2004) "Further Evidence that Legalized Abortion Lowered Crime: A Reply to Joyce", Journal of Human Resources

[11] Donohue & Levitt (2006) "Measurement Error, Legalized Abortion, the Decline in Crime: A Response to Foote and Goetz (2005)", mimeo.

[12] Goldin & Katz (2000) "Career and Marriage in the Age of the Pill", American Economic Review

[13] Goldin & Katz (2002) "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions", Journal of Political Economy

[14] Guldi (2005) "Abortion or The Pill - Which Matters More? The Impact of Access on the Birth Rates of Young Women", mimeo.

[15] Foote and Goetz (2005) "Testing Economic Hypotheses with State-Level Data: A Comment on Donohue and Levitt (2001)". Federal Reserve Bank of Boston working paper. November 2005

[16] Grogger (1997) "Incarceration-Related Costs of Early Child Bearing", in Rebecca Maynard, Ed. Kids Having Kids, Washington D.C.: The Urban Institute Press.

[17] Heckman, Farrar & Todd (1996) ”Human Capital Pricing Equations with an application to estimating the effect of Schooling Quality on Earnings”, Review of Economics and Statistics.

[18] Hock (2005) “The Pill and the College Attainment of American Women and Men,” mimeo, Florida State University.

[19] Hunt (2006) “Do Teen Births Keep American Crime High?”, Journal of Law and Economics.

[20] Joyce & Grossman (1990) ”Pregnancy Wantedness and the Early Initiation of Prenatal Care”, Demography.

[21] Joyce (2004a) “Did Legalized Abortion Reduced Crime?”, Journal of Human Resources

[22] Joyce (2004b) “Further Tests of Abortion and Crime”, NBER

[23] Joyce, Kaestner & Korenman (2000), ”The effect of pregnancy intention on child development”, Demography

[24] Kendall & Tamura (2006) ”Unmarried Fertility, Crime & Social Stigma”. mimeo.

[25] Klick & Stratmann (2003) ”The Effect of Abortion Legalization on Sexual Activity: Evidence form Sexually Transmitted Diseases” Journal of Legal Studies.

[26] Levine, Staiger, Kane & Zimmerman (1996), ”Roe v. Wade and American Fertility”, NBER Working Paper 5615

[27] Lochner & Moretti (2004) , ”The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports”, American Economic Review

[28] Lott & Whitley (2006) ”Abortion and Crime: Unwanted Children and Out-of-Wedlock Births”, Economic Inquiry

[29] Maltz & Targonski (2002) "A Note on the Use of County-Level UCR Data", Journal of Quantitative Criminology

[30] Sykes, Hangartner & Hathaway (2006) , The Ecological Fallacy of the Abortion Crime Thesis, mimeo.

CHAPTER 2

Strategic Parenting, Birth Order and School Performance

Interest on the effects of birth order on human capital accumulation has recently reemerged. The debate about its existence seems to be settled, but identification of the main mechanisms remains somewhat elusive. While the latest research aims at rediscovering dilution theory, we advance complementary economic hypotheses regarding the causal mechanisms underlying birth order effects in education. In particular, we entertain theories of differential discipline in which those who are born later face more lenient disciplinary environments. In such contexts, the later born sibling will be likely to exert lower school effort, thus reaching lower performance levels. We provide robust empirical evidence on substantial attenuation parental restrictions for those with higher birth order (born later). We speculate this may arise a) as a result of parental reputation dynamics and/or b) because of the changing relative cost of alternative monitoring and punishment technologies available to parents as well as increasing enforcing costs that must be afforded when multiple children must be monitored at the same time.

2.1 Introduction and Motivation

Interest on the effects of birth order on human capital accumulation has been reinvigorated by several recent studies (Black, Devereux and Salvanes, 2005; Conley and Glauber, 2006; Gary-Bobo, Prieto and Picard, 2006) which present new empirical evidence of birth order effects. For example, Black, Devereux and Salvanes (2005) (BDS, hereafter) find large and robust effects of birth order on educational attainment with Scandinavian data. However, despite the convincing results, the underlying causal mechanisms generating such findings remain somewhat unknown. Indeed, BDS acknowledge:

"...One important issue remains unresolved: what is causing the birth order effects we observe in the data? Our findings are consistent with optimal stopping being a small part of the explanation. Also, the large birth order effects found for highly educated mothers, allied with the weak evidence for family size effects, suggest that financial constraints may not be that important. Although a number of other theories (including time constraints, endowment effects, and parental preferences) have been proposed in the literature, we are quite limited in our ability to distinguish between these models...."

In thinking about children's behavior it is important to remember that parents can resort to a variety a mechanisms to influence it. In particular, they can limit or grant access to important sources of utility for children.

This chapter advances two channels of influence that have not been previously considered in the generating process for birth order effects in educational outcomes: we consider differential parental disciplining schemes arising from a) the dynamics of a parental reputation mechanism and/or b) the changing constraints in the technologies associated with enforcement of parenting rules and the implementation of punishment schemes. One channel that can generate birth order effects is characterized in Hao, Hotz and Jin (2008). A key insight of this paper is that birth order effects arise endogenously as the result of viewing parent-child interactions as a reputation game in which parents ”play tough” when their older children engage in bad behavior – tougher than caring, or altruistic, parents would prefer – in an attempt to establish a reputation of toughness in order to deter bad behavior amongst their younger children. Thus, we hypothesize that one mechanism that give rise to birth order effects is this form of strategic parenting and responses by their children implied by game-theoretic models of reputation in repeated games, where, in the context of this chapter, parents invest in developing a reputation of severe parenting with those born earlier in the hope of inducing their (paternalistic) preferred school effort levels on those born later.

We also consider a second mechanism of parenting that can generate birth order effects. In this case, parents differentially treat children of different ages because the technology of punishment available to parents might change as children grow up. This can happen because older children, who are initially reared alone, are able to interact with their younger siblings, once the latter are born, and such interactions can change the relative costs of alternative punishment schemes that parents might wish to employ. Similarly, their ability to monitor and enforce compliance with parenting rules may diminish when several children need to be overseen at the same point in time.

2.2 Related Literature

In this section we briefly review the literature on birth order effects and on the links between the effort of students in school and their academic performance and achievement.

There is a substantial literature on birth order effects in education. Zajonc (1976), Olneck & Bills (1979), Blake (1981), Hauser & Sewell (1985), Behrman & Taubman (1986), Kessler (1991), among others, found mixed results that provide support for a variety of birth order theories ranging from the ”no-one-to-teach” hypothesis to the theory of differential genetic endowments. However, with the strong birth order effects found in Behrman & Taubman (1986) and, more recently, in Black, Devereux & Salvanes (2005), the literature seems to be settling on the issue of existence and moving towards consideration and sophisticated testing of alternative mechanisms. Indeed, Price (2008) finds empirical support in time use data for a modern version of dilution theory: at least for a limited time, the first born doesn't have to share the available stock of parental quality time input with other siblings whereas those born later usually enjoy more limited parental input as parents are not able to match the increased demand for their ”quality” time.[1]

In another strand of research, mostly in Psychology, the issue of birth order effects in IQ has been examined. In particular, Rodgers et al. (2000, 2001) have consistently sided against the existence of such a relationship and they have criticized studies for confounding ”within-family” and ”between-family” processes and by attributing to the former, patterns that are actually shaped by the latter. More recently, Black, Devereux & Salvanes (2007) and Bjerkedal, et al. (2007) find strong and significant effects of birth order on IQ within families in a large dataset from Norway but Whichman, Rodgers & McCallum (2006) insist, using a multilevel approach, that the effects only arise between families and they disappear within the family. The debate remains open as Zajonc & Sulloway (2007) criticize Whichman, Rodgers & McCallum (2006) on several grounds and reach the opposite conclusion. Finally, Whichman, Rodgers & McCallum (2007) address the issues raised by Zajonc & Sulloway (2007) and confirm their previous findings.

[1] See Lindert (1977) for a related approach exploiting time use data.

There is also a sizeable literature on the links between students' effort in school and their academic performance (see, for example, Natriello and McDill (1986); Wolters (1999); and Covington (2000); Stinebrickner and Stinebrickner (2006)), in which a common measure of student effort is self-reports of number of hours spent on homework. There appears to be a fairly clear consensus in this literature that greater student effort improves academic performance. For example, Stinebrickner & Stinebrickner (2006) show the importance of actual school effort on school performance. But our understanding of how the factors that lead to greater student effort and how such effort interacts with other features of a student's home and school environments is less clear. Relevant to this chapter, there is a literature on the relationship between parenting and parental involvement and student effort and, ultimately, performance (see Trautwein and Koller (2003); Fan and Chen, (2001); Hoover-Dempsey, et al. (2001)). Most of this literature does not model or account for the endogenous nature of how the amount of school effort exerted by children is affected by parental incentives and policy instruments (i.e. whether it can be manipulated by the more economic, incentive-based side of parenting in the same way that criminal behavior can be influenced by the criminal justice system or savings and labor supply can be manipulated by different tax schemes). An exception to this shortcoming of the literature is a recent paper by De Fraja, D'Oliveira and Zanchi (2005). These authors develop an equilibrium model in which parents, schools and students interact to influence the effort of students and their performance and test this model using data from the British National Child Development Study. At the same time, the De Fraja, D'Oliveira and Zanchi (2005) does not characterize the potential informational problems that parents have in monitoring their children's input and the potential role of strategic behavior on the part of parents in attempting to influence the children's effort. This chapter attempts to fill this deficit in the literature.

As noted above, we draw on the game-theoretic literature on reputation models.

Such models were initially developed in the industrial organization literature in response to the chain store paradox of Selten (1978). In particular, Kreps and Wilson (1982) and Milgrom and Roberts (1982) developed models in which the introduction of a small amount of incomplete information gives rise to a different, more intuitive type of equilibrium. Hao, Hotz & Jin (2008) pioneered the use of this type of models in a family context to analyze teenage risk-taking behavior.

2.3 Theories of Birth Order Effects

There are several alternative causal hypotheses in the literature trying to explain the relationship between birth order and schooling. First, there could be parental time dilution. Under this hypothesis, the earlier born siblings enjoy more parental time than later-born siblings. This may explain why earlier-borns do better in school. Second, there could be genetic endowment effects. Indeed, later-born siblings are born to older mothers so they are more likely to receive a lower quality genetic endowment. Third, first-borns may ”reveal” the utility from parenting. According to this theory, a bad draw (i.e. a difficult to raise, problematic child) may lead to fertility stoppage. This will induce selection in the quality of the last child, being of lower quality than the average. Fourth, closely related to the ”confluence model” of Zajonc, the ”no one to teach” hypothesis postulates that the last born will not benefit from teaching a younger sibling. Without this pedagogic experience, the last born will not develop strong learning skills. Fifth, it may well be possible that the later-born siblings are more affected by family breakdown. BDS (2005) re-estimate their models in a sample of intact families and find no support for such hypothesis.[2]. Last, but not least, first-borns may enjoy higher parental investment for insurance purposes or simply because

[2]See Ginther & Pollak (2003) for an analyis of the relation between family structure and education outcomes.

parents are more likely to enjoy utility from observing their eventual success in life.

While all the above theories predict that earlier born siblings will do better, it is worth noting that it is possible that the effect can go in the other direction. For example, parents might learn to teach better. In this case, parents commit mistakes with those born earlier but they are more proficient, experienced parents when the later born siblings need to be raised. It also can be the case that, if there are financial constraints, the later-born siblings might be raised at time in which parental resources are more abundant.

Without taking away the merits of the previous literature, below we provide a novel, complementary mechanism that can induce birth oder effects in school performance. This channel is more economic, in that it highlights the role of incentives faced by children to perform well in school as well as the reputation concerns of lenient parents.

2.4 A Dynamic Model of Parental Reputation and Child School Performance

Consider a finite horizon game between parents and children being played in families with more than one child. Consider a long lived player (the parent) facing a new short lived player (the child) in every round of the game. In each round, the parent and the child observe the entire history of play. In particular, they observe the choices made by earlier born-siblings and the punishment decisions made by the parents when older siblings realized low levels of school performance. Parents can be of one of two types. With some prior probability, $\widehat{\mu}^T$, the parent is a commitment type (tough parent) that will always punish low school performance. With probability $1-\widehat{\mu}^T$, the parent is lenient and dislikes punishing the children. Of course, the history observed at any given point in time will not be informative about parental type if older siblings have always done well in school and $\widehat{\mu}^T$ will go unrevised. If $\widehat{\mu}^T$ is sufficiently small, the first-born child would prefer to exert low school effort if she is certain that a lenient parent would not punish such behavior. The prior beliefs about parental type are updated sequentially in a Bayesian fashion by siblings that come later in the birth order. In round t, the parent is believed to be tough with probability $\widehat{\mu}^t$.

It can be shown that a sequential equilibrium for this finitely repeated game exists. The critical event in the game is the observation of leniency upon low school performance at any given round t. If parents reveal themselves to be of the lenient type by not punishing the poor school performance of one of their children, $\widehat{\mu}^t$ drops to zero and remains there until the end of the game. From then on, the parents' children will fear no punishment from their revealed-to-be-lenient parent whose threats are no longer credible.

The equilibrium gives rise to 3 phases of the repeated game. In the first phase, played by earlier born siblings, uncertainty about parental type and threat of punishment induces these children to exert high levels of effort in school to deliver good school performance and prevent the triggering of potential punishments coded in the parenting rule. In this phase, bad grades will translate into loss of privileges anyway. If a parent is tough, he will punish by principle. If the parent is lenient, he will punish to invest in and/or maintain a reputation for toughness to prevent later born children from taking advantage of his leniency. As a result, we expect earlier born children playing mostly through this initial phase of the equilibrium to do better in school.[3] As the rounds of the game proceed, the number of remaining children at risk to play the game declines. At some point, the reputation benefits of punishment for a lenient parent do not cover the disutility of witnessing their child's suffering. Depending on how small

[3]Here we rely on results from Stinebrickner & Stinebricker (2006) that emphasize the importance of study effort in determining school performance.

$\widehat{\mu}^T$ is and how few children remain in the sequence, it will be likely for some children in the middle of the sequence to ”test the waters” by exerting low school effort and exploring what happens in response. After the first parental accommodating-behavior is observed, the final phase is triggered in which later born siblings do not put effort in school and go unpunished.

The model delivers some predictions that can be taken directly to the data. In particular, according to the model, earlier-born siblings are more likely to put more effort in school and should end up doing better. Moreover, parents are more likely to establish rules of behavior with the earlier-born, engage in a more systematic monitoring of earlier-born's schoolwork, increase supervision and limit privileges of the earlier-born in the event of low school performance.

2.5 The Data

We exploit data from the Children of NLSY79 female respondents (NLSY-C). In particular, we are able to observe the whole fertility history of NLSY79 females so we can potentially observe all of their children. Crucially, many of these females have 2 or more children so we are able to directly explore birth order effects that arise in these families. Due to limited sample sizes, however, we limit most of our analysis to families that have between 2 and 4 children.

TV watching and, more recently, video gaming are time intensive activities that usually crowd-out, at least partially, the time that could be used for homework or study. Indeed, there exist a vast literature in psychology documenting the detrimental effects of TV watching on school performance. Therefore TV viewing and videogaming are natural places to look for parental discipline schemes. Children value these activities highly and parents may be able to enforce and monitor restrictions on their access.

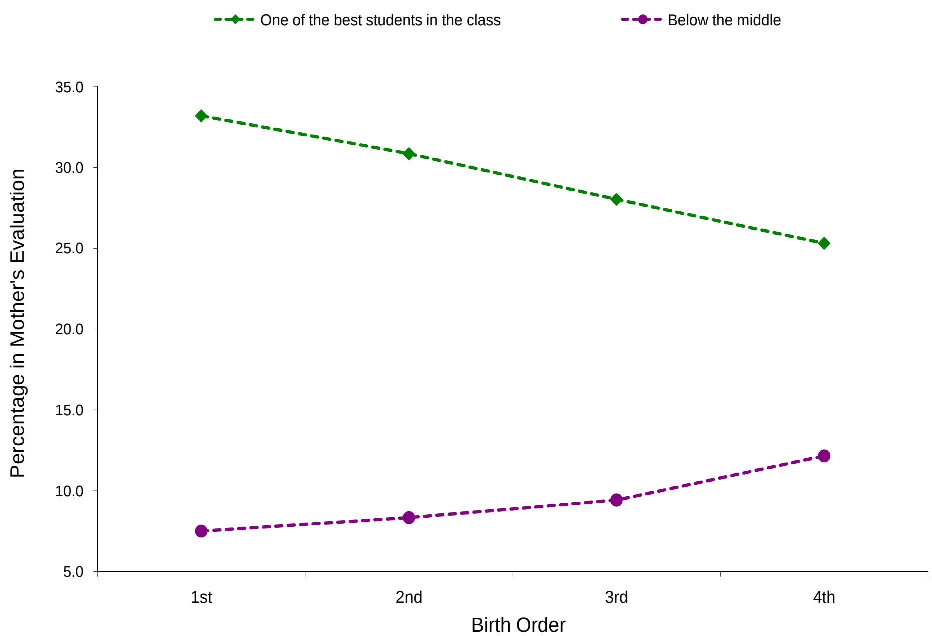
Useful for our purposes, the NLSY-C includes some detailed information on parenting. Several questions ask the mother and/or the children about different features about the parent-child relationship. We also exploit other parenting rules as reported by the children and/or the mother. Crucially, we are able to observe multiple self-reports from the same mother about all of her kids, and we observe those at two and sometimes three points in time. We restrict the analysis to children between the ages of 10 and 14. Since birth order is time invariant, we do not exploit the longitudinal nature of the data in our analysis. However, having repeated observations of parenting rules applied to each child over time allows us to identify changing parenting strategies.

On the other hand, the NLSY-C does not have systematic information on grades except for a specific supplemental school survey fielded in 1995-96 about school years 1994-95. However, the NLSY-C includes a self-report about how the mother thinks each of her children is doing in school. The specific question is: ”Is your child one of the best students in class, above the middle, in the middle, below the middle, or near the bottom of the class?” Useful for our purposes the same questions are asked of the mother separately for each child and in several waves. Note that even when these self-reports could be validated against school transcripts, it can be argued that it is the parental subjective belief about the child's performance what really matters at the end. We do, however, validate mother's perceptions below, exploiting limited transcript data from the 1995-96 School Supplement.

2.6 Birth Order Effects in (Perceptions of) Academic Success

Table 2.1 and Figure 2.1 show that there exists a clear association between school performance (as perceived by the mother) and birth order. Since the NLSY-C has very few observations coming from families with more than four siblings we focus our analysis on families with 2, 3 or 4 children. The table shows that while 33% of first born children are considered ”one of the best in the class” only 25% of those 4th in the birth order reach such recognition. On the other hand, only 7.5% of first-borns are considered ”below the middle or at the bottom of the class”, while 12.2% of 4th-borns are classified in such manner by their mothers.

Figure 2.1: Birth Order and Perceptions of School Performance



One possible concern with the results in Table 2.1 is that there is that they may confound birth order and family size effects, an issue that has been recognized very early in the development of the birth order literature. Figure 2.2 below explores birth order effects within family of specific sizes. Higher birth orders, by construction, belong in families of bigger size. As pointed out by Berhman & Taubman (1986), such families locate themselves at a different locus of the quantity-quality trade-off. Therefore we risk attributing to birth order what really comes from family size. As can be seen in the figure, birth order effects appear to persist in all these families, regardless of size.

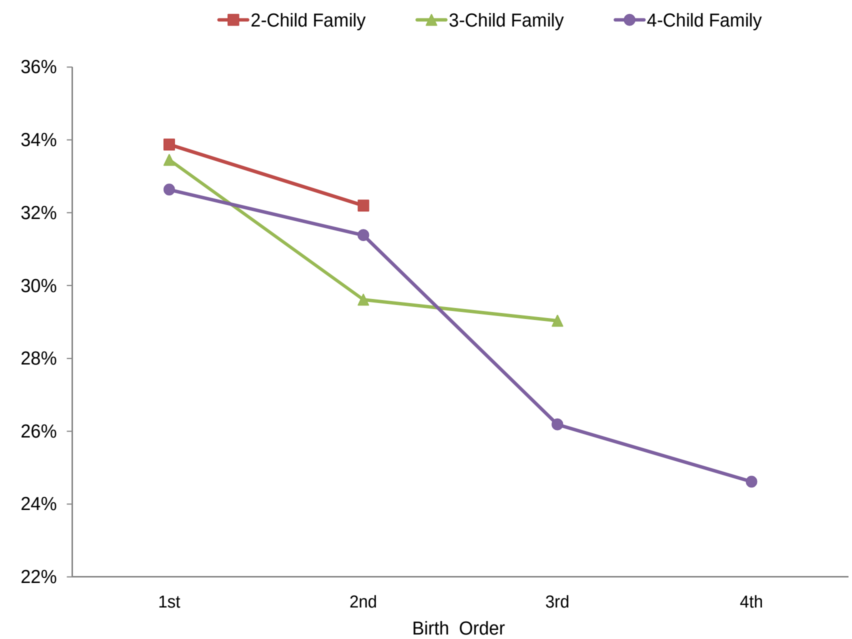
Table 2.1: Mothers Evaluation of Childs Performance by Birth Order

	Birth Order			
	1st	2nd	3rd	4th
One of the best students in the class	33.2	30.9	28.0	25.3
Above the middle	24.9	24.3	24.5	23.1
In the middle	34.4	36.5	38.0	39.5
Below the middle	5.7	6.3	7.3	8.8
Near the bottom of the class	1.9	2.0	2.1	3.4
Total	100	100	100	100

A second concern with the results above is that they show clear evidence of inflation in perceived school performance (i.e her assessments appear to show a mother's Lake Wobegon effect about their own children. However, this need not be a problem, per se, as long as the sign and magnitude of these misperceptions do not vary with birth order. In Table 2.2 below, we validate maternal perceptions. Higher GPAs of children obtained in the School Supplement are associated with significantly lower chances of being perceived to be at the bottom of the class and significantly higher chances to be classified as one of the best students in the class. Re-estimating the same models including birth order measures show that misperceptions (the differences between perceived and actual performance) are not correlated with birth order. Therefore, to the extent that mothers are too optimistic about their own children performance but they are so for all of their own children, we account for this mother specific bias when we include family fixed effects in our models of perceived school performance.

More formally, we follow BDS (2005) and explore birth order effects in academic performance by estimating the following two linear models for the probability that the child i in family h is being considered by his/her mother to be one of the best students in the class. The first specification imposes linearity while the second is more

Figure 2.2: Birth Order, Family Size and Perceptions of School Performance



non-parametric in the sense that it allows different effects for different birth orders

$$\begin{aligned} \text{Best Student}_{ih} &= NYG_i + X_i\beta + \lambda_h + \varepsilon_i \qquad (2.1) \\ \text{Best Student}_{ih} &= \sum_k \alpha_k \text{ Birth Order}_{ki} + X_i\beta + \lambda_h + \varepsilon_{ih} \text{ k=2,3,4} \end{aligned}$$

where X_i includes controls for child's age, survey year (and family size when pooling all families). NYG_i is the number of younger siblings, a measure of birth order that imposes linearity. Birth Order$_{ki}$ is a dummy variable which equals one when respondent i is the k^{th} child born in the the family, and equals zero otherwise. λ_h denote family fixed effects.

Tables 2.3 and 2.4 show the results of estimating the model in (2.1) for all families and then for families with 2, 3 or 4 children separately. In column one the specification imposes linearity of birth order and uses the number of younger siblings as a measure of birth order. In columns 2 to 5, all birth order coefficients are relative to the first

Table 2.2: Mothers Evaluation of Childs Academic Standing by Birth Order

	Ordered Probit		Probit		LPM by OLS	
	linear	non-parametric	linear	non-parametric	linear	non-parametric
GPA	-0.499***		0.188***		0.168***	
	[0.092]		[0.041]		[0.034]	
GPA=2		-0.902***		0.357**		0.191**
		[0.257]		[0.177]		[0.082]
GPA=3		-0.976***		0.423***		0.266***
		[0.259]		[0.158]		[0.079]
GPA=4		-1.870***		0.678***		0.557***
		[0.304]		[0.119]		[0.101]
Birth Order	0.063	0.074	-0.062	-0.065	-0.043	-0.051
	[0.114]	[0.119]	[0.054]	[0.055]	[0.046]	[0.047]
Observations	180	180	180	180	180	180

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%
1=Best, 0=otherwise. Controls include Age and Gender. In non-parametric specifications GPA=1 is the ommited category.

born, which is the omitted category. As can be seen in Table 2.3, there exist strong birth order effects in all families.

Moreover, when we estimate (2.1) controlling for family fixed effects the birth order results remain. See Table 2.4 below.

Table 2.3: Effect of Birth Order on the Probability of Being Perceived as One of the Best Students.(OLS)

	All Families	All Families	2-Child Family	3-Child Family	4-Child Family
# of Younger Sibs	0.054***				
	[0.006]				
Second Child		-0.057***	-0.053***	-0.066***	-0.053**
		[0.010]	[0.014]	[0.017]	[0.026]
Third Child		-0.107***		-0.089***	-0.140***
		[0.015]		[0.020]	[0.028]
Fourth Child		-0.158***			-0.178***
		[0.026]			[0.031]
Observations	11532	11532	4809	4433	2290
Number of Families					

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year. The specifications pooling all families include family size indicators. Linear Probability Models. Dependent Variable =1 if Mother thinks child is one of the best students in the class, =0 otherwise.

Table 2.4: Effect of Birth Order on the Probability of Being Perceived as One of the Best Students.(Family Fixed Effects)

	All Families	All Families	2-Child Family	3-Child Family	4-Child Family
# of Younger Sibs	0.041***				
	[0.011]				
Second Child		-0.049***	-0.072***	-0.037*	-0.049*
		[0.014]	[0.024]	[0.021]	[0.029]
Third Child		-0.085***		-0.027	-0.120***
		[0.024]		[0.034]	[0.042]
Fourth Child		-0.110***			-0.151***
		[0.037]			[0.056]
Observations	11532	11532	4809	4433	2290
Number of Families	2693	2693	1391	915	387

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year. Specifications pooling all families include family size indicators. Linear Probability Models. Dependent Variable =1 if Mother thinks child is one of the best students in the class, =0 otherwise.

2.7 Birth Order Effects in Incentives

In this section, we explore at a descriptive level whether birth order effects may arise because of differential parental treatment. We ask whether the data shows any sign of differential parental toughness by birth order. First, we estimate ordered probit models for our categorical variable on the likelihood of getting TV time limited by parents[4]

$$\text{Limit TV time}_i = \begin{cases} \text{Never} & \text{if} \quad & \text{Limit}_i^* & < \mu_0 \\ \text{Rarely} & \text{if} \quad \mu_0 < & \text{Limit}_i^* & < \mu_1 \\ \text{Sometimes} & \text{if} \quad \mu_1 < & \text{Limit}_i^* & < \mu_2 \\ \text{Often} & \text{if} \quad \mu_2 < & \text{Limit}_i^* & \end{cases} \tag{2.2}$$

where

$$\text{Limit}_i^* = \sum_k \gamma_k \text{Birth Order}_{ki} + \delta X_i + \varepsilon_i \tag{2.3}$$

Table 2.5 shows estimates from this ordered probit model for parental toughness.

As can be seen in the table, the frequency of TV time limitations declines with birth order. Those born first tend to face stricter disciplinary standards regarding this activity (i.e. parents tend to be more tough/severe on them when it comes to the child's TV time) . Similarly, parents seem to be increasingly lenient with those born later. Column 6 includes estimates from the same ordered model for all families but allowing for family random effects. The estimated birth order effects and its precision is little changed.

Table 2.6 shows OLS and Family Fixed Effects estimates of the same model with a dichotomous variable which equals one if parents often limit TV time, and equals zero

[4]OLS and Fixed Effects estimates for models with dichotomous versions of the same dependent variable generate the same pattern of birth order effects.

Table 2.5: Effect of Birth Order on the Frequency TV limitations (Ordered Probit)

	All Families	All Families	2-child Family	3-child Family	4-child Family	All Families (Family Random Effects)
# of Younger Sibs	0.109***					0.112***
	[0.019]					[0.022]
Second Child		-0.138***	-0.151***	-0.088	-0.221**	
		[0.031]	[0.041]	[0.055]	[0.099]	
Third Child		-0.216***		-0.168***	-0.303***	
		[0.044]		[0.058]	[0.099]	
Fourth Child		-0.306***			-0.392***	
		[0.077]			[0.105]	
Observations	6684	6684	2911	2518	1255	6684

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%
Categorical Dependent Variable with 4 categories: Never, Rarely, Often, Always. All models include indicators for child's age and survey year. Specification pooling all families includes family size indicators.

otherwise. The OLS results are, again, strikingly similar. They support the existence of differential disciplinary schemes which are strongly linked to birth order. However, the fixed effect estimates show consistent signs but the effects are no longer significant.

Table 2.6: Effect of Birth Order on the Probability of Having TV time Limited (OLS and Family Fixed Effects)

	OLS					Family Fixed Effects				
	All Families	All Families	2-child Family	3-child Family	4-child Family	All Families	All Families	2-child Family	3-child Family	4-child Family
# of Younger Sibs	0.027***					0.016				
	[0.007]					[0.018]				
Second Child		-0.041***	-0.042***	-0.028	-0.082**		-0.029	-0.015	-0.02	-0.097**
		[0.011]	[0.015]	[0.020]	[0.036]		[0.020]	[0.036]	[0.031]	[0.047]
Third Child		-0.056***		-0.039*	-0.086**		-0.031		-0.002	-0.096
		[0.016]		[0.021]	[0.036]		[0.036]		[0.053]	[0.066]
Fourth Child		-0.068***			-0.093**		-0.042			-0.101
		[0.026]			[0.038]		[0.057]			[0.093]
Observations	6684	6684	2911	2518	1255	6684	6684	2911	2518	1255
Number of Families						2143	2143	1084	738	321

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All models include indicators for child's age and survey year. Specification pooling all families includes family size indicators. Linear Probability Models. Dependent Variable =1 if parents always limit TV time, =0 otherwise

Below we provide additional evidence consistent with some of the predictions delivered by the reputation model. While evidence in Table 2.6 is somewhat mixed, Table 2.7 shows that there are strong birth order effects in the existence of rules about TV watching. In this case, the results are robust to the introduction of family fixed effects. Earlier-born siblings seem to grow up in a more regulated environment regarding TV relative to their later-born counterparts.

Table 2.7: Existence of Rules about Watching TV and Birth Order (OLS and Family Fixed Effects)

	OLS					Family Fixed Effects				
	All Families	All Families	2-Child Family	3-Child Family	4-Child Family	All Families	All Families	2-Child Family	3-Child Family	4-Child Family
# of Younger Sibs	0.047***					0.034**				
	[0.007]					[0.014]				
Second Child		-0.052***	-0.077***	-0.039**	-0.012		-0.028*	-0.014	-0.027	-0.02
		[0.012]	[0.016]	[0.019]	[0.030]		[0.017]	[0.029]	[0.026]	[0.035]
Third Child		-0.103***		-0.118***	0.001		-0.071**		-0.106**	-0.04
		[0.017]		[0.023]	[0.034]		[0.029]		[0.042]	[0.053]
Fourth Child		-0.110***			-0.042		-0.100**			-0.121*
		[0.032]			[0.039]		[0.048]			[0.072]
Observations	9665	9665	4004	3745	1916	9665	9665	4004	3745	1916
Number of Families						2548	2548	1298	874	376

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year. Specifications for all families include family size indicators. Linear Probability Models. Dependent Variable =1 if there are rules about watching TV, =0 otherwise.

In Tables 2.8 and 2.9 we provide evidence of birth order effects in how intensely parents monitor a child's homework. Consistent with the reputation model, earlier born siblings face more intense, systematic parental scrutiny regarding homework. Parents who are prepared to punish in the event of low school performance are more likely to seek more information on how much effort is being exerted by their children on homework. Moreover, more intense monitoring conveys more credibility to the threat of punishment. Table 2.8 report results from Ordered Probit models that fully exploit all the variation in the categorical dependent variable. Table 2.9 shows OLS and Family Fixed Effects estimates based upon a binary version of the dependent variable which equals one when the monitoring is most intense (daily checks on homework)[5].

[5]The actual question is "*How often do your parents check on whether you have done your homework?*" Allowed answers include: Never, Less than once a month, 1-2 times a month, 1-2 times a week, Almost every day, Every day.

Table 2.8: Intensity of Homework Monitoring and Birth Order (Ordered Probit)

	How Often Parents Check Homework is Done?					
	All Families	All Families	2-Child Family	3-Child Family	4-Child Family	All Families (Family Random Effects)
# of Younger Sibs	0.044**					0.052**
	[0.019]					[0.022]
Second Child		-0.073**	-0.083**	-0.031	-0.141	
		[0.032]	[0.042]	[0.057]	[0.100]	
Third Child		-0.089**		-0.098	-0.042	
		[0.045]		[0.061]	[0.098]	
Fourth Child		-0.103			-0.082	
		[0.076]			[0.104]	
Observations	6629	6629	2893	2487	1249	6629

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%
All regressions include indicators for child's age and year. a This specification includes family size indicators. In Linear Probability Models. Categorical Dependent Variable with increasing categories of monitoring intensity.

Table 2.9: Intensity of Homework Monitoring and Birth Order (OLS and Family Fixed Effects)

Table 2.9: Intensity of Homework Monitoring and Birth Order. (OLS and Family Fixed Effects)

	OLS					Family Fixed Effects				
	All Families	All Families	2-Child Family	3-Child Family	4-Child Family	All Families	All Families	2-Child Family	3-Child Family	4-Child Family
# of Younger Sibs	0.016*					0.037*				
	[0.008]					[0.022]				
Second Child		-0.032**	-0.034*	-0.021	-0.058		-0.063**	-0.028	-0.071*	-0.125**
		[0.014]	[0.018]	[0.025]	[0.044]		[0.025]	[0.047]	[0.037]	[0.059]
Third Child		-0.032		-0.028	-0.045		-0.076*		-0.109*	-0.1
		[0.019]		[0.026]	[0.044]		[0.046]		[0.065]	[0.085]
Fourth Child		-0.031			-0.04		-0.091			-0.116
		[0.034]			[0.047]		[0.070]			[0.116]
Observations	6629	6629	2893	2487	1249	6629	6629	2893	2487	1249
Number of Families						2138	2138	1082	735	321

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year.Specifications for all families include family size indicators. In Linear Probability Models. Dependent Variable =1 if parents check homework everyday, =0 otherwise.

Tables 2.10 and 2.11 show strong birth order effects on the likelihood that the parent would increase supervision in the event of low school performance. The question is asked to the mother and is a scenario question, not a self report about behavior. In this sense, it provides an interesting complement to more standard data on observed behavior because it essentially recovers the parental ”reaction function” directly, even in cases in which the child does well in school and never triggers the eventual punishment. Again, Table 2.10 shows estimates from an ordered probit model that exploit the variation in the categorical dependent variable.[6] In Table 2.11 we work with a dichotomous version of the dependent variable which equals one if mother would be very likely to keep a closer eye on the child in the event of low school performance and zero, otherwise. This allows us to easily control for family fixed effects.

[6]The specific question we exploit in this context is the following : ”*If (Child) brought home a report card with grades lower than expected, how likely would you (the mother) be to keep a closer eye on [his/her] activities?”* Allowed answers were: Not At All Likely, Somewhat Unlikely, Not Sure How Likely, Somewhat Likely, Very Likely.

Table 2.10: Likelihood of Increased Supervision in the Event of Low Grades and Birth Order (Ordered Probit)

	How Likely to Supervise more Closely if Low Grades					
	All Families	All Families	2-Child Family	3-Child Family	4-Child Family	All Families (Family Random effects)
# of Younger Sibs	0.060***					0.093***
	[0.019]					[0.026]
Second Child		-0.049	-0.078*	-0.036	0.015	
		[0.032]	[0.046]	[0.053]	[0.077]	
Third Child		-0.121***		-0.159***	-0.003	
		[0.046]		[0.062]	[0.086]	
Fourth Child		-0.192**			-0.13	
		[0.080]			[0.099]	
Observations	10468	10468	4346	4031	2091	10468

Robust standard errors in brackets.* significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year. Specifications for all families include family size indicators. Categorical Dependent Variable denoting increasing likelihood of parents closer supervision in the event of low school performance.

Table 2.11: Likelihood of Increased Supervision the Event of Low Grades and Birth Order (OLS and Family Fixed Effects)

	OLS					Family Fixed Effects				
	All Families	All Families	2-Child Family	3-Child Family	4-Child Family	All Families	All Families	2-Child Family	3-Child Family	4-Child Family
# of Younger Sibs	0.017***					0.021*				
	[0.006]					[0.011]				
Second Child		-0.012	-0.021*	-0.006	0.007		-0.007	-0.021	-0.008	0.019
		[0.009]	[0.012]	[0.015]	[0.026]		[0.012]	[0.021]	[0.019]	[0.028]
Third Child		-0.036***		-0.046**	0.000		-0.039*		-0.068**	0.022
		[0.014]		[0.018]	[0.028]		[0.022]		[0.031]	[0.044]
Fourth Child		-0.053**			-0.035		-0.077**			-0.006
		[0.026]			[0.032]		[0.037]			[0.058]
Observations	10468	10468	4346	4031	2091	10468	10468	4346	4031	2091
Number of Families						2646	2646	1365	900	381

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year. Specification for all families include family size indicators. In Linear Probability Models. Dependent Variable =1 if parents are very likely to supervise more closely in the event of low grades, =0 otherwise.

Finally, Table 2.12 presents estimates for the outcome of whether privileges have been limited because of poor grades. The question here is asked of each different child and since the actual question is more related to observed limitations in privileges we expect only the kids who actually are perceived to be doing not so well in school to be the ones reporting loss of privileges. We then estimate the following model

$$\text{Limit Privileges}_i = \beta_0 + \beta_1 B_i + \beta_2 NYG_i + \beta_3 B_i \times NYG_i + \delta X_i + \varepsilon_i \qquad (2.4)$$

where $B_i = 1$ if the child is perceived to be doing bad in school (below the middle of the class) and NYG_i is our measure of birth order indicating the number of younger siblings. In this context, we expect the coefficient β_3 on the interaction term $B_i \times NYG_i$ to be positive and to capture the prediction of the reputation model: earlier-born siblings should be more likely to report a loss of privileges relative to later-born siblings, *only when they do bad in school.*[7]

As can be seen in the last column of Table 2.12, the family fixed effects estimates of β_3 are positive with a point estimate of 0.124 that is significant at 5%.

[7]Note that the results in columns 6 and 12 of Table 2.12 should be interpreted with caution as they do not control for the endogeneity of B.

Table 2.12: Observed Limits to Privileges because of Low Grades and Birth Order (OLS and Family Fixed Effects)

	OLS						Family Fixed Effects					
	All Families	All Families	2-Child Families	3-Child Families	4-Child Families	All Families	All Families	All Families	2-Child Families	3-Child Families	4-Child Families	All Families
# of Younger Sibs	-0.028***					-0.029***	-0.056***					-0.058***
	[0.008]					[0.008]	[0.019]					[0.020]
Second Child		0.037***	0.026	0.056**	0.041			0.050**	0.079*	0.062*	0.019	
		[0.014]	[0.018]	[0.024]	[0.044]			[0.023]	[0.041]	[0.034]	[0.053]	
Third Child		0.062***		0.084***	0.029			0.114***		0.138**	0.009	
		[0.019]		[0.025]	[0.043]			[0.041]		[0.059]	[0.078]	
Fourth Child		0.061*			0.042			0.171***			0.035	
		[0.033]			[0.046]			[0.062]			[0.107]	
Bad Grades						-0.067						-0.160**
						[0.062]						[0.081]
Bad Grades x # of Younger Sibs						0.051						0.124**
						[0.053]						[0.062]
Observations	6579	6579	2855	2492	1232	6579	6579	6579	2855	2492	1232	6579
Number of Families							2134	2134	1077	738	319	2134

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year. Specifications for all families include family size indicators. In Linear Probability Models. Dependent Variable =1 if child reports that parents limited priviliges when low grades, =0 otherwise

While many of the previous results are consistent with predictions from the reputation model, we can also test for other mechanisms that may generate birth order effects in performance. For example, we can provide a test of the dilution theory by looking at whether the frequency of parental help with homework varies with birth order. Dilution theory predicts that at any given age of the child, parents will help earlier born siblings more frequently. In Table 2.13, we show that there appear to be no birth order effects on the frequency of parental help with homework. Earlier-born siblings seem to benefit from the same level of parental input in this regard.

Table 2.13: How Often Parents Help with Homework and Birth Order.(Ordered Probit)

	How Often Parents Help with homework?					
	All Families	All Families	2-Child Family	3-Child Family	4-Child Family	All Families (Family Random Effects)
# of Younger Sibs	-0.028					-0.022
	[0.018]					[0.022]
Second Child		0.009	-0.02	0.07	-0.023	
		[0.030]	[0.039]	[0.053]	[0.096]	
Third Child		0.062		0.139**	-0.019	
		[0.042]		[0.056]	[0.094]	
Fourth Child		0.09			0.054	
		[0.074]			[0.100]	
Observations	6640	6640	2897	2488	1255	6640

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year. Specifications for all families include family size indicators. Child'sselfreported categorical Dependent Variable with categories denoting increasing frequency of parental help with homework.

In the first five columns of Table 2.14 we show OLS estimates for the same model and, if anything, the children who benefit the most from parental help with homework seem to be later-born siblings. The more robust fixed effects estimates in last five columns, however, show no relation between birth order and frequency of parental help with homework as reported by the children.

Table 2.14: How Often Parents Help with Homework and Birth Order (OLS and Family Fixed Effects)

	OLS					Family Fixed Effects				
	All Families	All Families	2-Child Family	3-Child Family	4-Child Family	All Families	All Families	2-Child Family	3-Child Family	4-Child Family
# of Younger Sibs	-0.012*					0.024				
	[0.007]					[0.017]				
Second Child		0.015	0.01	0.023	0.01		-0.026	0.011	-0.034	-0.053
		[0.011]	[0.015]	[0.018]	[0.034]		[0.019]	[0.037]	[0.028]	[0.042]
Third Child		0.025*		0.048**	-0.017		-0.055		-0.086*	-0.087
		[0.015]		[0.020]	[0.033]		[0.035]		[0.049]	[0.061]
Fourth Child		0.029			0.016		-0.056			-0.088
		[0.026]			[0.035]		[0.055]			[0.086]
Observations	6640	6640	2897	2488	1255	6640	6640	2897	2488	1255
Number of Families						2140	2140	1083	736	321

Robust standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%

All regressions include indicators for child's age and year. Specifications for all families include family size indicators. Linear Probability Models. Dependent Variable =1 if parents often help with homework, =0 otherwise.

2.8 Directions for Future Research

In future work, we plan to examine the consistency of our data with other theories of birth order (i.e. dilution) that might complement the effects arising from differential birth order discipline. In particular, some of the predictions of the reputation model, such as the more intense homework monitoring of earlier born siblings, cannot be distinguished from the predictions of a time dilution model. In that sense, we plan to test whether birth order effects reflect a parental reputation mechanism or merely reflect changes in the relative costs of implementing, enforcing and monitoring a given disciplinary scheme over an increasing number of children of different ages at the same point in time.

Also, while family fixed effects account for time invariant characteristics of the family, they do not capture dynamic processes within the family that vary over time and are correlated with birth order and may affect school outcomes. In particular, later born siblings are more likely to be affected by family breakdown. The NLSY sample provides ample opportunities to control for family structure as a potential determinant of birth order effects.

The available data also allows us to estimate a structural version of the full reputation game. To proceed in that direction, we will have to specify the structure of the dynamic game. In particular, functional forms for payoffs, prior beliefs and the updating rule must be parameterized while allowing for rich forms of observed and unobserved heterogeneity. The solution to the dynamic reputation game can then be embedded in a estimation algorithm that should match the predictions of the game to the observed behavior in the NLSY families. Note that since we are able to observe the parental policy function over the last child we, as econometricians, are in a privileged position of knowing parental type. That is, we can readily identify severe parents when they report to be willing to punish the last born sibling in the event of low school performance. The estimated model can then be used to answer a number of interesting questions. The key counterfactual question of interest would be about school performance when kids are fully informed about parental type. In particular, we expect school performance to be lower as kids take advantage of altruistic-forgiving parents and don't put effort in school given that punishment threat by parents is no longer credible. This counterfactual provides a quantitative measure of the effects of "responsible" parenting in school performance. We also can explore whether birth order effects persist in this "complete information" scenario. We also can use the estimated game to endow children in high-risk groups with different priors, enough to induce their parents to build and maintain severe reputations. We can then ask whether the achievement gap across groups declines.

2.9 Conclusions

We contribute to the literature on birth order effects in human capital accumulation by showing that those born earlier are perceived to perform better in school. A validation of perceptions using actual transcript data shows that these findings do no reflect Lake Wobegon effects or, more importantly, any differential performance misperception by birth order. Our results are robust to controls for family size and, more generally, to the inclusion of family fixed effects.

We provide evidence consistent with parental reputation incentives generating birth order effects in school performance. In particular, earlier born siblings are more likely

- to be subject to rules about TV watching
- to face more intense parental monitoring regarding homework
- to suffer loss of privileges because of low grades.

Moreover, mothers themselves report being more likely to increase supervision in the event of low school achievement when the child in question was born earlier.

While further research is needed to rule out alternative explanations associated with changing cost and technologies of alternative parenting strategies as sibships grow we believe that results indicate that parental reputation dynamics may explain part of the observed birth order effects in school performance.

2.10 Bibliography

[1] Behrman, J. & P. Taubman (1986) ”Birth Order, Schooling & Earnings” Journal of Labor Economics, Vol 4, No 3.

[2] Bjerkedal, T., P. Kristensen, G. Skejeret and J.I. Brevik (2007) ”Intelligence test scores and birth order among young Norwegian men (conscripts) analyzed within and between families”, Intelligence

[3] Black, S., P. Devereux and K. Salvanes (2005) ”The More The Merrier ? The Effect of Family Size and Birth Order on Children’s Education”, Quarterly Journal of Economics, CXX, 669

[4] Black, S. P. Devereux and K. Salvanes (2007) ”Older and Wiser?: Birth Order and IQ of Young Men” NBER working paper 13237

[5] Bonesrønning, H. (1998). “The importance of student effort in education production. empirical evidence from Norway’ Department of Economics. Norwegian University of Science and Technology.

[6] Blake, Judith (1981) ”Family Size and the Quality of Children ” Demography, Vol.18, No 4, 421-442

[7] Conley, D. and R. Glauber (2006) "Parental Education Investment and Children's Academic Risk: Estiamtes of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility" Journal of Human Resources 41(4): 722-37.

[8] Covington, M. (2000), "Goal Theory, Motivation and School Achievement: An Integrative Review,' Annual Review of Psychology 51:171–200.

[9] De Fraja, G., T. D'Oliveira and L. Zanchi (2005). "Must Try Harder. Evaluating the Role of Effort on Examination Results", CEPR Discussion Paper 5048, May 2005.

[10] Fan, X. and M. Chen (2001). "Parental involvement and students' academic achievement: A meta-analysis,' Educational Psychology Review 13:1-22.

[11] Gary-Bobo, R., A. Prieto and N. Picard (2006) "Birth-Order and Sibship Sex Composition Effects in the Study of Education and Earnings" Centre for Economic Policy Research Discussion Paper No. 5514.

[12] Hao, Hotz & Jin (2008) "Games that Parents and Adolescents Play: Risky Behavior, Parental Reputation and Strategic Transfers" Economic Journal, 118(528): 515–555.

[13] Hauser R. M. and W. H. Sewell (1985) "Birth Order and Educational Attainment in Full Sibships" American Educational Research Journal, Vol. 22, no1, 1-23.

[14] Hoover-Dempsey, K. V., A. C. Battiato, J. M. T. Walker, R. P. Reed, J. M. DeJong, and K. P. Jones (2001). "Parental involvement in homework,' Educational Psychologist 36: 195-209.

[15] Kessler (1991) "Birth Order, Family Size & Achievement: Family Structure and Wage Determination", Journal of Labor Economics, Vol. 9, no 4. 413-426.

[16] Kreps, D. and Robert Wilson (1982), "Reputation and Imperfect Information," Journal of Economic Theory, 27, 253-279.

[17] Lindert, Peter (1977) "Sibling Position and Achievement" Journal of Human Resources 12, no. 2, 220-241

[18] Milgrom, Paul and John Roberts (1982), "Predation, Reputation and Entry Deterrence." Journal of Economic Theory, 27, 280-312.

[19] Natriello, G. and E. McDill (1986), "Performance Standards, Student Effort on Homework, and Academic Achievement,' Sociology of Education, 59(1): 18-31.

[20] Olneck M. R. and D. B. Bills (1979) "Family Configuration and Achievement: Effects of Birth Order and Family Size in a sample of Brothers", Social Psychology Review, Vol. 42, No. 135-148

[21] Price, Joseph. (2008) "Parent-Child Quality Time: Does Birth Order Matter?" Journal of Human Resources, 43(1): 240-265.

[22] Rodgers, J. L. , H. H. Cleveland, E. van den Oord and D. C. Rowe (2000), Resolving the Debate Over Birth Order, Family Size and Intelligence, American Psychologist, 55(6) 599-510

[23] Rodgers, J. L. , H. H. Cleveland, E. van den Oord and D. C. Rowe (2001), Birth Order and Intelligence: Together Again for the Last Time?, American Psychologist

[24] Selten, R., 1978, The Chain Store Paradox, Theory and Decision, 9, 127 - 159

[25] Trautwein, U. and O. Koller (2003). "The relationship between homework and achievement: Still much of a mystery,' Educational Psychology Review 15, 115-45.

[26] Whichman, A. L., J. L. Rodgers and R. C. McCallum (2006), A Multi-Level Approach to the relationship between Birth Order and Intelligence, Personality and Social Psychology Bulletin, Vol. 32 No. 1, January, 117-127

[27] Whichman, A. L., J. L. Rodgers and R. C. McCallum (2007), Birth Order has no effect on Intelligence: A Reply and Extension of Previous Findings, Personality and Social Psychology Bulletin, Vol. 33 No. 9, September,1195-1200

[28] Wolters, C. (1999), "The relation between high school students' motivational regulation and their use of learning strategies, effort, and classroom performance, Learning and Individual Differences 11:3: 281-99.

[29] Zajonc R. B. and F.J. Sulloway, (2007), The Confluence Model: Birth Order as a Between Family or Within Family Dynamic? Personality and Social Psychology Bulletin, Vol 33, 1187-1194.

[30] Zajonc, R. B. (1976), "Family Configuration and Intelligence" Science, Vol. 192, 227-236, 1976

CHAPTER 3

On Scarlet Letters and Clean Slates: Criminal Records Policy in a Dynamic Model of Human Capital Accumulation and Criminal Behavior

In this chapter I formulate and estimate a dynamic model of human capital accumulation and criminal behavior. Every period, forward looking individuals optimally decide whether to engage in criminal activities and invest in human capital. They can attend school and accumulate experience through learning by doing in legitimate activities. The estimated model sheds light on the relationship between education and crime and it allows me to explore some of the competing explanations advanced in the literature. Specifically, some researchers have emphasized the inherently custodial nature of schools. Others highlight the role of education in raising the opportunity costs of criminal behavior in terms of time allocation. It could also be the case that a negative correlation arises only because of unobserved heterogeneity in the degree of forward looking behavior. The model developed in this chapter allows for these alternative channels and advances a novel one, based on information and reputation arguments. The model is also used to evaluate alternative policies associated with the availability of criminal records, improving our understanding of criminal record stigma as a cost borne by ex-offenders and also as an expected cost taken into account by those considering to engage in crime. Microdata from the National Youth Survey are used to estimate the parameters of a discrete choice dynamic programming model using a simulated maximum likelihood estimator that adapts importance sampling techniques to reduce computational burden. Results indicate that the informational structure embedded in an open criminal records policy is responsible for part of the observed negative correlation between education and crime. ”Open Records” policies also generate substantial dynamic deterrence via threat of stigmatization and induce an overall decline in crime relative to both, the status-quo and alternative ”Sealed Records” policies.

3.1 Introduction and Motivation

Every year more than half a million prisoners are released in the United States. It is a continuous flow. Moreover, this flow will increase significantly in upcoming years. Many more criminals are put on probation for less serious convictions but cannot avoid the stigmatizing mark of a criminal record. What are the prospects for reintegration into society for these ex-offenders? A substantial fraction of these criminals happen to be high school dropouts. This has led many to examine the negative relationship between crime and education. While some consensus has been established regarding the causal nature of this relationship, the particular mechanisms responsible for it remain somewhat debatable. Some researchers emphasize the custodial nature of schools while others highlight their role in fostering the accumulation of human capital which increases the opportunity cost of time devoted to criminal activities. It is also possible that high levels of crime and low levels of schooling merely reflect unobserved factors such as low rate of time preference, which favors activities with contemporaneous benefits and delayed, uncertain costs (i.e., crime), and discourages activities with early costs and distant benefits (i.e., education). A novel channel, advanced in this chapter, explores whether labeling mechanisms, such as those induced by an open criminal records policy, may be responsible for the negative relationship between education and crime. If individuals are forward looking and criminal records are open to potential employers, additional human capital not only deters criminal activity by increasing its opportunity cost in terms of time allocation, but it does so by putting more earnings potential at risk of being jeopardized for life. More educated individuals suffer from a more sizable labor market stigma (fewer job offers, lower earnings) in the event of conviction. Open criminal records policy therefore makes human capital and criminal capital bad complements in the production of lifetime utility, thus inducing further educational stratification in criminal activity.

The teenage and early adulthood years are very important years in everyone's life. Critical investments in human capital as well as the initiation of legal and criminal careers occur during these years. To better understand the channels that give rise to a negative relationship between education and crime it is useful to formulate a comprehensive model in which individuals make optimal decisions regarding human capital accumulation and criminal behavior. These decisions are inherently dynamic and it seems critical to let individuals take into account future consequences of current actions as well as to allow them to make decisions sequentially, so as to re-optimize with the arrival of new information every period.

Research on the effects of certainty and severity of punishment is vast, both in the fields of criminology and economics. On the other hand, much less research has been devoted to the effects of a particular criminal records policy. While there are other potentially stigmatizing effects at the community level, I will focus on the labor market consequences of a criminal record. If individuals are rational and forward-looking, employers' access to criminal histories has, in principle, two effects:

1. Deterrent or crime-reduction effect, because it deteriorates labor market prospects in the event of conviction and therefore exerts deterrence on individuals evaluating the start-up of a criminal career. This "dynamic deterrence" operates beyond and above the standard deterrent effects associated with the certainty and the severity of punishment by the criminal justice system (CJS).

2. A crime-promoting effect, as those who get caught see their chances of starting a "legal" life reduced by the stigma of their criminal record. This stigma hinders their labor market opportunities after conviction and therefore propels recidivism.

While the second may prevail in the short run, the former is important to understand the longer, potentially unintended consequences of altering criminal records policy. For example, suppose a massive criminal records sealing policy is adopted which effectively improves labor market opportunities for stigmatized ex-offenders. It is likely that we may observe a decline in the crime rate in this subset of the population. However, if individuals are forward looking, this policy may have unintended consequences as new generations grow up facing a new set of incentives that make crime effectively cheaper by eliminating potential labor market stigma. The increased criminal activity of the marginal individuals from these younger generations might outweigh the decline in crime among ex-offenders.

To summarize, the main research questions I seek to address in this chapter are the following:

a) What is the effect of public availability of criminal histories on career decisions?

b) To what extent changing criminal records policy increases the difference between the highly educated and the less educated in levels of criminal activity.

The chapter is organized as follows. The next section reviews the relevant literature on the relationship between education and crime, the effect of criminal records on labor market outcomes and the use of forward looking dynamics in modeling criminal behavior. The contribution of the chapter is then described in the context of this literature. Section 3 describes the NYS, whose microdata is used to estimate the model. Section 4 describes the dynamic model of human capital accumulation and criminal behavior. Section 5 is devoted to estimation issues. There I describe an extension of the estimator proposed in Ackerberg (2001) to allow for endogenous initial conditions that arise in my application. Parameter estimates and model fit are discussed in Section 6 along with a novel validation strategy involving subjective expectations of college completion. Section 7 conducts the main policy experiments of interest. Conclusions follow in Section 8.

3.2 Related Literature and Contribution

Ever since the times of Jeremy Bentham (1789), criminal behavior has been understood as the result of a rational cost-benefit analysis that maximizes pleasure and minimizes pain. Since Becker's (1968) seminal contribution, economists have become seriously involved in the formal study of criminal behavior. The economic approach adopts the rational choice perspective and therefore provides an alternative to mainstream sociological theories of crime.[1]

Education and Crime. Economic research on the relationship between crime and education has been scarce. This is surprising given the often simplistic debate between tougher punishment and better educational opportunities as alternative ways to reduce crime. For notable exceptions see Lochner and Moretti (2004), Lochner (2004) and Witte (1997). Witte (1997) provides a review of the literature on the relationship between education and crime. She highlights the importance of endogeneity of schooling when modeling criminal outcomes. Based mostly on evidence from correlational studies, she concludes that higher educational attainment is not associated with lower levels of criminal activity.[2] She emphasizes the "custodial" role of schools rather than their human capital accumulation role as, in her view, it is the time spent in school

[1] See Hirschi (1986) for an account of the repeated failure of rational choice perspectives in becoming a respected paradigm among sociologists. In particular, note the paragraph in p. 111: "....Sociology rejected rationalistic.....theories and rejects them now. The reasons for rejection remain much the same: Adequate theories of crime must be positive. They must provide motives of (or the causes of) criminal behavior. They cannot assume that crime will occur in the absence of restraint, because absence-of-restraint theories .....are contrary to the scientific assumption that behavior is caused by antecedent events......" It is clear that the approach taken in this chapter, scientific or not, is closer to an absence-of-restraint theory. Note also that while Hirshi (1986) proposes to associate rational choice theories with "situational" crime and reserve social control theories to understand "criminality" or "criminal involvement," this chapter is, if anything, a formalization of the rational choice approach to the latter. That is, the model here explains how a law abiding person may become a criminal using a life-cycle or life-course perspective. His decision is not just a matter of committing crimes when enough criminal opportunities are at hand, conditional on being a criminal already.

[2] See in particular, Tauchen, Witte and Griesinger (1994).

”under supervision” that reduces crime and not the level of educational attainment. Lochner and Moretti (2004) exploit exogenous variation in compulsory school laws and find a significant negative causal effect of education on incarceration probability. Then, they corroborate their findings using microdata from the NLSY 1979. They find that additional education significantly reduces criminal activity and this, in turn, implies lower probabilities of arrest and incarceration. Their results seem to stand in contrast to those surveyed by Witte (1997). In particular they highlight the earnings-enhancing power of additional schooling and the associated increase in the opportunity cost of crime, much in the spirit of classical human capital theory. They conclude with a strong policy recommendation:

”...It is difficult to imagine a better reason to develop policies that prevent high-school drop.”

It is not clear, however, how such a policy should be implemented. Should teachers not fail students in the lowest quantiles of the class distribution when they do not meet minimum absolute levels of academic performance? Should we lower the school quality so these students can succeed by facing easier requirements? The model proposed in this chapter can be used to evaluate explicitly some of these alternatives and to test Witte’s and Lochner and Moretti’s competing explanations for the role of schooling in reducing crime. Lochner (2004) is the first to exploit a classic human capital framework to understand the relationships between education, crime and work. Predictions from his model regarding the form of age-crime and education-crime relationships find support in NLSY79 and FBI data.

The impact of a criminal record. In the last 15 years a specific sub-literature on the impact of criminal records has emerged[3]. With a few notable exceptions the general

[3]Of course, the topic has received the attention of lawyers and criminologists since earlier times. There are several contributions in the area including Lott (1990,1992), Waldfogel (1994 a, b) , Nagin

message in this literature is that criminal records have a negative effect on labor market outcomes. The work of Waldfogel (1994a) is especially important for my purposes. It suggests the existence of an interaction effect by which a given criminal record hits even harder on those with higher education. As explained below, this chapter takes the argument to a next level by exploring the consequences in terms of human capital accumulation and criminal behavior when agents are forward looking and they are aware of this negative interaction between a criminal record and the level of education.

It is important to recognize that this literature is more informative about short run effects. In general these papers do not address the steady state question posed here, which is more aimed to understand the counterfactual behavior of a cohort that would grow up and take decisions facing a completely different set of incentives regulating how transparent reputations are in the labor market.

Dynamic Models of Criminal Behavior. Flinn (1986) pioneered the analysis of criminal careers in the context of explicit dynamic models of behavior. He envisioned a research agenda in which criminal careers could be formalized by means of explicit dynamic, forward looking models of economic (i.e. rational) behavior. At the time of his writing, computational and econometric developments for estimation of dynamic models were at their infancy so he did not actually estimate those behavioral models, but used them to guide the interpretation of the less structured statistical models of criminal careers widely adopted in criminometric research. Davis (1988) considered the implications of heterogeneous discount rates and highlighted the usefulness of explicit dynamic models to analyze the impact of programs that raise offender's income after punishment, such as prisoner schooling and counseling. Davis rightly argued that such programs might reduce recidivism by increasing the opportunity cost associated with further offending, but will certainly raise the level of first offenses as their ex-

and Waldfogel (1993, 1995), Grogger (1995), Bushway (1998), Kling (1999), Kling (2004) Holzer, Raphael and Stoll (2001, 2002a, b, 2003a, b, c) Pager (2002, 2003), Bushway (2004) and Finlay (2007).

pected cost is now lower. As discussed in Section 7, sealing adult criminal records is another example of such a policy intervention. Some recent efforts by Imai and Krishna (2004) and Williams and Sickles (2002, 2003, 2004) take the first steps towards estimation of this type of models exploiting the Philadelphia 1958 Birth Cohort. Imai and Krishna (2004) find significant deterrent effects associated with publicly available criminal histories. Dynamic models in a more macroeconomic/representative agent framework have been developed by Leung (1994) and Imrohoroglu, Merlo and Ruppert (2004) while Burdett, Lagos and Wright (2003, 2004) and Huang, Lain and Wang (2004) have focused on search-theoretic models of crime.

3.2.1 Contribution

At the same time, the model captures the key determinants of a criminal career and their relationship with schooling decisions, overcomes dynamic selection problems and gets at the impact of publicly available criminal histories on labor market outcomes. It does so using observational, non-experimental methods. It deals with deterrence and stigma in an integrated, consistent framework. Policy conclusions may be misleading if they fail to consider dynamic deterrent effects of criminal records policy and only focus on the stigma effects that propel recidivism. This chapter combines features of the models advanced by Lochner (2004), Imai and Krishna (2004), and Eckstein and Wolpin (1999). The chapter is closer in spirit to Lochner (2004) in that an integrated framework is used to understand human capital accumulation, labor supply and criminal activity. It shares, however, some features with Imai and Krishna (2004) and Eckstein and Wolpin (1999) by considering the ”dynamic deterrent” effect of criminal records, exploiting information on grade progression and by pursuing estimation of the full dynamic structural model. There are some key differences, though, that define this contribution. Unlike Imai and Krishna (2004), I explicitly model schooling decisions jointly with criminal and legal activities. This allows me to study questions associated with the relationship between education and crime. Unlike Lochner (2004), I pursue estimation of the full dynamic model instead of testing some of its empirical implications. The modeling of investment in human capital is more linked to the empirics of schooling data as in Keane and Wolpin (1997). The stock of legal and criminal capital are accumulated endogenously through work and criminal activity. The model allows for a ”supervision effect” of schooling in addition to the standard ”earnings-enhancing” human capital effect. Finally, while having an extremely careful modeling and estimation exercise to understand high school drop out decisions, Eckstein and Wolpin (1999) does not aim to understand criminal behavior. I believe the more comprehensive modeling strategy adopted here follows naturally from the large number of criminals who are, in fact, high school dropouts.[4]

In summary, while not an all-purpose model of criminal behavior, the model is quite general and provides a complementary tool to evaluate policy interventions and to compare its predictions with real and natural experiments like those in Pager (2002, 2003) for criminal records policy or Lochner and Moretti (2004) for education policy.

3.3 Data

Given the difficulties in obtaining official criminal records, many studies rely on self-reported measures of criminal activity. While self-reported data are subject to many caveats, they certainly provide a more complete picture than official criminal records regarding the level of criminal activity. This chapter uses the National Youth Survey

[4]In a footnote, Eckstein & Wolpin (1999) note the appropriateness of extending their framework along the direction adopted here. They also point out the significant challenges that such an extension would face, in terms of modeling, computation, data and estimation.

(NYS).[5]

The National Youth Survey began in 1976. At that time 1725 adolescents between the ages of 11 and 17 years old were interviewed along with one of their parents. The survey is a representative sample of the U.S. youth in 1976. Now called the National Youth Survey - Family Study (participants who were once 11-17 are now 39-45) this study has followed these individuals through 29 years to look at their changing attitudes, beliefs and behaviors. They were asked about topics such as career goals, involvement with community and family, attitudes about violence, drugs, and social values. Unfortunately, I only have access to data up to the 7th wave in 1987 when respondents were between 21 and 27. Table 3.1 describes the cohort structure of the sample used for estimation. Self-reported information on selected crime categories is available for all years 1976-1986. This information is consistent because some waves asked about retrospective criminal activity for years when no survey was conducted.[6]

Retrospective information on convictions and incarcerations[7] was collected in the last publicly available wave. Data on schooling are fairly detailed and allow accurate construction of full educational histories including attendance as well as grade and GPA progression. In the first five waves a ”scenario” or ”expectation” question was asked about the chances that the respondent would complete a college degree. Since by 1987 most of the sample is older than 22, we can test the validity of this piece of choice expectation data. Indeed, Table 3.2 shows that these self reports are predictive of actual college completion by respondents. Having a self-report assessing ”Poor” chances of college completion perfectly predicts the lack of a college degree by 1986.

[5]For example Lochner (2004) exploits the NLSY79 which has a crime module in 1980. NLSY97 has several questions on crime and delinquency. These have been used by Lochner (2005) and Paternoster et al. (2003).

[6]The waves fielded in 1986 and 1983 acquired compatible information about 1984-85 and 1981-82, respectively.

[7]I define incarcerations as time spent in detention centers, training facilities, jails and prisons.

Table 3.1: Cohort Structure of the NYS (1976-1986)

1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986
11										
12	12									
13	13	13								
14	14	14	14							
15	15	15	15	15						
16	16	16	16	16	16					
17	17	17	17	17	17	17				
	18	18	18	18	18	18	18			
		19	19	19	19	19	19	19		
			20	20	20	20	20	20	20	
				21	21	21	21	21	21	21
					22	22	22	22	22	22
						23	23	23	23	23
							24	24	24	24
								25	25	25
									26	26
										27

Source: National Youth Survey. Respondent are 11 to 17 in first wave in 1976. First 5 waves go from 1976 to 1980. 6th wave in 1986 collects retrospective information about 1984-85. 7th wave in 1986 collects retrospective information about 1984-85.

Moreover, those who assess their chances as "Good" obtain college degrees at a much higher rate relative to those who only consider their chances as "Fair" (32.7% and 4.4% , respectively). Instead of using these data in the estimation I reserve them to validate the model.[8]

Finally, rich information on labor market experiences is available to construct detailed and consistent work histories. Earnings are available from 1978 onwards. Excluding observations with missing data in some of the variables, I end up with a sample of 3729 person-year observations. Table 3.3 provides some descriptive statistics for the data to be used in the estimation of the model. The crime indicator is broadly

[8]See van der Klaauw (2000) for an effort to combine subjective expectation data with observed choice data in the estimation of dynamic behavioral models.

Table 3.2: The Predictive Power of Self-Reported Expectations of College Completion

	% of respondents who actually completed a college degree by 1986	
	No	Yes
What do you think your chances are for completing a college degree? (Asked from 1976 to 1980)		
"Poor"	100.0	0.0
"Fair"	95.6	4.4
"Good"	67.3	32.7

Note: self reported expectations of college completion come from the first 5 waves of NYS. Actual college completion is observed as of the 7th NYS wave conducted in 1986

defined so that about 27% of observations are criminally active in the sample period. This criminal and delinquent activity is highly concentrated in the late teens and early 20s.[9] The average age in the sample is 18.65 and we observe a fair number of juvenile and adult convictions that provide a key source of identification. About 10% of those attending school fail to successfully complete the current grade and over half of the observations have ”good” grades. The information on grades is self-reported and collected every period. While the allowed answers include 5 categories (Mostly As, Mostly Bs, Mostly Cs, Mostly Ds, Mostly Fs) I choose to collapse them into 2 aggregate categories: Good Grades (Mostly As or Mostly Bs) and Bad Grades (Mostly Cs, Mostly Ds or Mostly Fs).[10] Finally note that the stocks experience and criminal capital are only observed during the sample period. This creates an initial conditions problem. The estimation section details how to overcome this data limitation.

Finally, the lack of geographic identifiers provides some further justification for the

[9]See Figure 3.3 in Appendix C

[10]While I recognize the loss in useful variation resulting from this grouping, the gain in tractability for estimating the model described below is very high.

Table 3.3: National Youth Survey - Descriptive Statistics

Variable	Mean	Std. Dev.
Age (in years)	18.65	3.68
Work (d^W=1)	0.79	0.41
Attend School (d^S=1)	0.54	0.50
Criminally Active (d^C=1)	0.27	0.44
Experience (X)	3.9	2.92
Criminal Capital (CK)	1.4	2.1
School Attainment (ATT)	10.2	2.64
Have a Juvenile Record (CR=1)	0.03	0.17
Have an Adult Record (CR=2)	0.07	0.26
Succesful Grade Transition (PASS=1)	0.91	0.28
GPA=1 (get mostly As or Bs)	0.56	0.50
FS=1 (Incarceration Episode)	0.006	0.08
Earnings ($ in 1986 dollars)	$ 11,956.5	$ 8,530.8
Self-Reported Chances of College Completion		
"Poor"	0.19	0.40
"Fair"	0.35	0.48
"Good"	0.46	0.50

Source: Publicly available NYS waves from 1976 to 1986. Respondents are interview at ages between 11 and 27. The values for experience and criminal capital are stocks accumulated while in the sample window for those who are older than 11 in 1976. GPA was aggregated into 2 categories. GPA=1 if respondent was getting mostly A's or B's in the period, GPA =0 if respondent was gettiong mostly Cs, Ds or Fs in the period.

structural approach adopted here, as the quasi-experimental variation that comes from institutional differences across states cannot be easily exploited.[11]

3.4 Model

Consider a simple forward looking model. Each period agents decide whether to engage in crime, work in legitimate activities and attend school in order to maximize

[11] I have reached an agreement with the directors of the NYS to access their highly confidential data in the future, under strict confidentiality restrictions. Therefore, in future research, this will allow me to combine the structural model proposed here with more quasi-experimental approaches.

expected lifetime utility. I assume school attendance, crime and work are *not* mutually exclusive choices[12]. Agents receive job offers from the legitimate sector. In deciding whether to engage in criminal activities, individuals compare benefits and costs and are fully forward-looking. They take into account the probabilities of arrest and conviction as well as the severity of punishment that will accrue in the future as a consequence of their criminal behavior. Also, individuals consider the fact that they will be building a criminal record and therefore dramatically deteriorating their labor market prospects in the future. In fact, the probability of receiving at least one job offer *and* the wage offer itself will depend on the status of the criminal record. They accumulate criminal capital k_a through learning by doing (i.e. they accumulate criminal capital by being criminally active).[13]

The decision period is annual. Let $a = 11, ..., A$ denote age. Let d_a be a vector of choice indicators at age a.

$$d_a = \left[d_a^S, d_a^W, d_a^C \right]$$

[12]In the simplest case where all these decisions are binary, I then have a set of $K = 2^3 = 8$ choices. There is ample empirical evidence showing that most criminals are employed in some legitimate job. Fagan and Freeman (1999) justify this joint treatment where crime and work are not mutually exclusive choices. See also Grogger (1998), and Reuter, MacMoun and Murphy (1990).

[13]It would be possible to allow for criminal capital to be accumulated with prison tenure. The idea here is that by spending time in jail, individuals learn from other criminals. See Bayer, Pintoff and Pozen (2005). Given the very small number of incarcerations observed in this representative dataset I refrain from seriously pursuing this alternative.

where

$$d_a^S = \begin{cases} 1 & \text{if attend school at age } a \\ 0 & \text{otherwise} \end{cases}$$

$$d_a^W = \begin{cases} 1 & \text{if work at age } a \\ 0 & \text{otherwise} \end{cases}$$

$$d_a^C = \begin{cases} 1 & \text{if criminally active at age } a \\ 0 & \text{otherwise} \end{cases}$$

The indicator for the resulting $2^3 = 8$ mutually exclusive choices is d_a^k for $k = 1, 2......, 8$.

3.4.1 Criminal Environment

Let CR_a denote the criminal history as of age a. This is the most important state variable in the model. The simplest way to proceed is to have a state variable that jointly keeps track of the criminal history that is available to criminal justice system agencies and the history that is available to potential employers.

$$CR_a = \begin{cases} 0 & \text{if completely clean record at age } a \\ 1 & \text{if has a juvenile record} \\ 2 & \text{if has an adult record} \end{cases} \tag{3.1}$$

Note that under the status quo, criminal history is available to employers only when $R_a = 2$, that is, when the record comes from a conviction received in adult court.

I let FS_a be an indicator that it is equal to one if the individual suffers any type of incarceration at age a and zero otherwise.

CJS Probabilities. I focus on the following three events to characterize the possible contacts with the criminal justice system. At any age/period there are three possible outcomes:

1. get arrested, convicted and sent to jail, with probability p_a^J
2. get arrested, convicted and released on probation, with probability p_a^C
3. not get arrested, with probability $1-\left(p_a^J+p_a^C\right)$

Probabilities associated with criminal justice system outcomes are given by an ordered logit model based on the latent index $I^*_{a_{CJS}}$

$$\begin{aligned} I^*_{a_{CJS}} &= \lambda_1^{CJS}(CR_a>0)+\lambda_2^{CJS}FS_a+\lambda_3^{CJS}1\left(d_a^C=1\right) \\ &+\lambda_4^{CJS}ATT_a+\lambda_5^{CJS}a+\lambda_6^{CJS}a^2+\varepsilon_a^{CJS} \end{aligned} \tag{3.2}$$

where ε_{CJS} has logistic distribution and ATT_a stands for educational attainment and a denotes age. The cutoffs, μ_0^{CJS} and μ_1^{CJS} determine the above mentioned probabilities.

$$p_a^J = \Pr\left(I^*_{a_{CJS}}<\mu_0^{CJS}\right) \tag{3.3}$$

$$p_a^C = \Pr\left(\mu_0^{CJS}<I^*_{a_{CJS}}<\mu_1^{CJS}\right) \tag{3.4}$$

These probabilities depend on age to reflect the different treatment received at CJS agencies by individuals of different ages and/or the changing athletic capacity of the individual to commit certain crimes. CJS probabilities are also a function of the criminal record. This is to capture increased surveillance and supervision by probation and police officers. Finally, I allow for wrongful and/or delayed convictions and incarcerations by including the indicator for criminal activity instead of making these transition probabilities conditional on being criminally active.

Criminal Capital. Let CK_a denote criminal capital. It measures experience in criminal activities accumulated endogenously by being criminally active. Allowing for accumulation of criminal capital provides another source of state dependence in criminal behavior. This is ”true” state dependence coming from habit persistence, lock-in effects or positive returns to criminal capital, which make criminal activity more likely once crimes have already been committed in the past. This will capture true state dependence as I will be controlling for unobserved heterogeneity which, as it is well known, can also explain observed systematic persistence in behavior. Finally, criminal capital accumulates deterministically according to

$$CK_{a+1} = CK_a + 1\left(d_a^C = 1\right) \tag{3.5}$$

3.4.2 Schooling

The model allows for a stylized representation of the school environment. The focus is on modeling the transition process for school performance and grade progression. Let ATT_a denote the highest grade attained as of age a; it accumulates stochastically. Its evolution depends on attendance, age, years of education, grade point average (GPA), work status and freedom status.[14]

To accumulate schooling, individuals need to attend school *and* get a passing grade.

$$\pi_{ga} = p_a^{g(a)} = \Pr\left(\text{ get a passing grade} | \, a, g(a), GPA, d_a^W = 1, FS_a \right) \tag{3.6}$$

is a parameter governing the discrete and binary random variable $PASS_i$, so

$$PASS_i(a, g(a)) = \begin{cases} 1 & \text{with probability } \pi_{ga} \\ 0 & \text{with probability } 1 - \pi_{ga} \end{cases} \tag{3.7}$$

for each age a.

[14]There are two parameters that are chosen to reflect institutional features of the educational system. MINSCHOOLAGE = 16 implies that only choices which include school attendance are available up to age 15. MAXED = 18 is set as the maximum number of years of education an individual can have.

If the individual decides to attend school at age a, I parameterize the grade pass probability and the GPA process conditional on attendance $\left(d_a^S = 1\right)$ using logit models with index $I^*_{a_{PASS}}$ and $I^*_{a_{GPA}}$ given by

$$I^*_{a_{PASS}} = \psi_0 + \psi_1 a + \psi_2 ATT_a + \psi_3 1\left(GPA_a = 1\right) + \psi_4 1\left(d_a^W = 1\right) + \psi_5 FS_a + \varepsilon_a^{PASS} \tag{3.8}$$

where ε_{CJS} has logistic distribution and GPA_a is given by[15]

$$GPA_a = \begin{cases} 1 & \text{if GPA = A or B with probability } p^A_{GPA} \\ 0 & \text{if GPA = C, D or F with probability } 1 - p^A_{GPA} \end{cases} \tag{3.9}$$

The index governing the probability for GPA_{a+1} is given by

$$I^*_{a_{GPA}} = \lambda_0^{GPA} + \lambda_1^{GPA} 1\left(GPA_{a-1} = 1\right) + \lambda_2^{GPA} 1\left(d_a^W = 1\right) + \lambda_3^{GPA} FS_a + \varepsilon_a^{PASS} \tag{3.10}$$

where again, ε_{PASS} has logistic distribution.

Both indexes include FS_a as determinant to account for the fact that episodes of incarceration negatively affect academic performance.[16] They also include $1\left(d_a^W = 1\right)$ as there exists now a substantial literature that examines the impact of working while in school on a variety of outcomes, including future labor market opportunities and academic performance.[17] Since the GPA process is highly persistent, $1\left(GPA_{a-1} = 1\right)$ is also included as one of its main determinants.

[15]To keep the state space tractable and to conserve in the number of parameters I limit the GPA variable to only two values as detailed in the Data section.

[16]See Pintoff (2006).

[17]See, for example, Hotz et al. (2002)

3.4.3 Job Offers, Experience & Wages

Let $\lambda\left(CR_a\right)$ denote the job offer probability. Every period, an individual receives at least one job offer with probability

$$\lambda\left(CR_a\right) = \frac{\exp\left(\lambda_0 + \lambda_1 1\left[CR_a = 2\right]\right)}{1 + \exp\left(\lambda_0 + \lambda_1 1\left[CR_a = 2\right]\right)} \tag{3.11}$$

Crucially, I allow the job offer probability to depend on the criminal record. I expect λ_1 to be negative since many employers are reluctant to hire ex-offenders and others are required by law not to do so. Note that these job offers arrive regardless of labor force participation status. Assume the individual observes whether or not an offer has arrived and, in case one arrives, the individual observes its terms, $w(a)$. When estimating the model, the mean in the population distribution of λ_0 can be parameterized in terms of exogenous characteristics observable to employers.

Let X_a denote accumulated experience in legal activities as of age a. I assume no depreciation, so experience evolves according to

$$X_a = X_{a-1} + d_{a-1}^{w} \tag{3.12}$$

Wage offers in legal activities at age a are given by

$$\begin{aligned} w(a) &= w\left(X_a, d_{a-1}^{w}, ATT_a, CR_a, \varepsilon_a^{w}\right) \\ &= \exp[\alpha_0 + \alpha_1 X_a + \alpha_2 X_a^2 + \alpha_3 ATT_a + \alpha_4 ATT_a \times 1(CR_a = 2) \\ &\quad + \alpha_5 1(CR_a = 2) + \alpha_6 HS_a + \varepsilon_a^{w}] \end{aligned} \tag{3.13}$$

where HS_a is an indicator equal to 1 if the individual has received a high school diploma and zero otherwise, to capture a diploma effect on wages associated with finishing high school. ε_a^{w} captures iid shocks to earnings that are normally distributed with zero mean and standard deviation σ_0^{W}

3.4.4 Solution

At any age $\tilde{a}$ the agent maximizes the expected discounted value of remaining lifetime utility

$$\max_{\{d_a\}_{a=\tilde{a}}^{A}} E\left\{\sum_{a=\tilde{a}}^{A} \beta^{a-\tilde{a}} u\left(d_a, s_a\right) | \Omega(\tilde{a})\right\} \tag{3.14}$$

where s_a collects all the state variables at age a and $\Omega(a)$ is the information set at age a. and S. From now on I omit it but it is understood that all expectations taken as of age a are conditional on the information set at that age.

Following Eckstein & Wolpin (1999), I normalize utility to consumption units (i.e. dollars) and parameterize it as linear, additive in consumption and leisure. As shown below, I generalize this utility to account for utility (or disutility) of attending school, committing crimes and being free.

The value of choosing any of the $k = 1, ..., K$ mutually exclusive discrete alternatives is

$$V_k\left(s_a\right) = u_k\left(s_a\right) + \beta E\left[V\left(s_{a+1}\right) | s_a, d_a = k\right] \tag{3.15}$$

$$V\left(s_a\right) = \max_k \left\{V_k\left(s_a\right)\right\} \tag{3.16}$$

Computation of $E\left[V\left(s_{a+1}\right) | s_a, d_a = k\right]$ is burdensome. I use Monte Carlo simulation to integrate over the shocks $(\varepsilon_W, \varepsilon_S, \varepsilon_C, \varepsilon_L)$ and integrate over the possible future states using the discrete transition probabilities for non-deterministic state variables. For example consider $k = 8$, then

$$E\left[V\left(s_{a+1}\right) | d_a = 8, S_a\right] = \sum_{GPA} \sum_{PASS} \sum_{CJS} \left(E\left[V(s_{a+1}) | s_a, d_a = 8,\right]\right) p_{GPA} \pi_{PASS} p_{CJS} \tag{3.17}$$

where $p_{CJS} = p_J, p_C, p_F$ with $p_F = 1 - p_J - p_C$.

The period utility function $u_k(a)$ has base components $\gamma(S_a)$ associated with each activity[18],

$$\begin{aligned} u(d_a, s_a) \quad = \quad & 1\left(d_a^W = 0\right)\left[\gamma_L(s_a) + \varepsilon_a^L\right] \\ & +1\left(d_a^S = 1\right)\left[\gamma_S(s_a) + \varepsilon_a^S\right] \\ & +1\left(d_a^C = 1\right)\left[\gamma_C(s_a) + \varepsilon_a^C\right] \\ & +1\left(FS_a = 1\right)\gamma_F(s_a) + w(a) \end{aligned} \tag{3.18}$$

The program is solved by backwards recursion using Monte Carlo integration and interpolation along the lines of Keane and Wolpin (1994). Note that this is a forward looking model, but I only have data up to age 27. However, most of the action in criminal behavior and human capital accumulation is over at age 27[19] so the available data cover the most relevant period for the purposes of this chapter. ε_a^j for $j = L, S, C$ are normally distributed shocks to the utility of each activity.

Rather than solving the model all the way up to the advanced age in which the individual retires or dies, I approximate the value at some early age $\hat{a} < A$, $V(s_{\hat{a}})$, with the expected present discounted value of remaining lifetime earnings[20]

[18]See Appendix D for exact functional forms.

[19]The crime-age profile is one of the most firmly established empirical facts in the criminological literature.

[20]I assume that job offers arrive at the same rate from $\hat{a}$ onwards and that the effect of the criminal record does not depreciate over time. If offers do not arrive in a given period, the value of leisure is used instead. To minimize extrapolation into age ranges not covered by the data, I set $\hat{a} = 28$. Alternatively, a general parametric function of the state vector at that age (i.e. a "terminal value function") could be specified and estimated to summarize the value of remaining lifetime utility. However, that option has a number of disadvantages in my context. First, allowing for a flexible terminal value function would be expensive in terms of parameters given the several state variables in the model. Second, the identification of those parameters would not be easily separated from the discount factor. Finally, given its ad-hoc nature, such terminal value function would not be suitable for conducting policy experiments, as the ones conducted below, that would depend on it.

3.5 Estimation

To estimate the model, schooling, employment and criminal behavior histories are exploited along with information on contacts with the criminal justice system and wages. Estimation proceeds by simulated maximum likelihood. More details can be found in Appendix A. I propose an extension of the estimator advanced in Ackerberg (2001) to allow for the endogenous initial conditions that arise in my application. Transition probabilities for non-deterministic state variables follow the parametric models described in the model section. The likelihood function is straightforward, once I properly allow for unobserved heterogeneity and the fact that I do not observe some state variables in the first sample period for observations older than 11 in 1976. The building blocks of the likelihood function are the choice probabilities, which are simulated using simple crude frequency. The solution to the DP problem given by $E\left[V\left(s_{a+1}\right)|s_a, d_a = k\right]$ for all a, k and s_a is used as input in computing these probabilities.

$$\begin{aligned}
& \Pr\left(d_a, = k|\, s_a; \theta\right) \qquad (3.19) \\
= \; & \Pr\left(V_k\left(s_a, \theta\right) > V_j\left(s_a, \theta\right) \text{ all j} \neq \text{k}\right) \\
= \; & \Pr\left(\begin{bmatrix} u_k\left(s_a, \theta\right) + \\ \beta E\left[V\left(s_{a+1}, \theta\right)|d_a = k\right] \end{bmatrix} > \begin{bmatrix} u_j\left(s_a\right) + \\ \beta E\left[V\left(S_{a+1}, \theta\right)|d_a = j\right] \end{bmatrix} \text{ for all j} \neq \text{k}\right)
\end{aligned}$$

where $E\left[V\left(s_{a+1}, \theta\right)|s_a, d_a = k\right]$ is, for example, given by (3.17) and θ is the vector of structural parameters that vary over individuals in ways that depend on the specification of unobserved heterogeneity detailed below.[21]

[21] When choosing an alternative involving work the conditional density for the wage is also included in the likelihood.

3.5.1 Unobserved Heterogeneity

The estimation framework used here allows for rich forms of unobserved heterogeneity. Usually this heterogeneity is incorporated non-parametrically in the likelihood function by introducing K types via a finite, discrete distribution[22] and solving the DP problem conditional on each type. Instead, here I adopt a random coefficients framework that allows each individual to have it's own vector of structural parameters by specifying a continuous distribution of types. Many unobserved characteristics can be thought of as being heterogeneous across individuals. Everything from heterogeneous levels of moral restraint to differential levels of patience can explain why some individuals are systematically more prone to committing crimes than others.[23]

I allow for heterogeneity in every preference parameter: the discount factor β, the base utilities from schooling γ_S, crime γ_C, leisure γ_L, and the disutility from prison γ_P.I also allow for heterogeneity in the intercept and the structural parameters of the earnings function. Allowing for unobserved heterogeneity in transition probabilities is costly, because doing so destroys the separability in the likelihood function that would allow estimation of those transitions separately, in a first stage, greatly reducing the size of the parameter space when estimating the model. However, assuming no heterogeneity in transitions would be clearly unrealistic in this setting. Therefore, I allow for heterogeneity in the intercept and slope coefficients of the logit models for GPA evolution λ^{GPA}, and grade successful completion, ψ, as well as in the cutoffs and slopes for the ordered logit model that characterizes contacts with the CJS $\left(\mu_0^{CJS}, \mu_1^{CJS}, \lambda^{CJS}\right)$.

[22]This strategy is now fairly standard in estimation of dynamic models. It has been applied in many articles. See, for example, Wolpin (1984), Van der Klaauw (1996), Eckstein and Wolpin (1999) and Keane and Wolpin (1997, 2001).

[23]See Hauser (2004) for a different application with unobserved heterogeneity in discount factors.

3.5.2 Reducing Computational Burden Using Importance Sampling

I follow the suggestion in Ackerberg (2001) and use importance sampling to reduce the computational burden in the estimation of the dynamic programming model. In particular, assume that unobserved heterogeneity is characterized by a continuous distribution as opposed to a discrete distribution with finite and small number of points of support.[24]

Let α_i be the structural parameter vector for person i. Note that in addition to parameters characterizing preferences and transitions, α_i includes the parameters characterizing the distribution of the i.i.d. shocks to utility and the variance of the wage shock. The Log-likelihood function is given by

$$\log L(\theta) = \log \prod_{i=1}^{N} L_i(\theta) = \sum_{i=1}^{N} \log L_i(\theta) \tag{3.20}$$

The likelihood contribution $L_i(\theta)$ for observations who are 12 years old or more in 1976 is complicated because I do not get to observe some state variables in the first sample period. So, without loss of generality, I sort the observations such that the first $N_{c_{11}}$ are those corresponding to agents who are 11 years old in 1976. The remaining observations belong to any of the other cohorts: $c_{12}, c_{13}, \ldots.., c_{17}$

$$\log L(\theta) = \sum_{i:i\in c_{11}} \log L_i(\theta) + \sum_{i:i\in\{c_{12}, c_{13}, \ldots.., c_{17}\}} \log L_i(\theta) \tag{3.21}$$

For the sake of simplicity, let's focus first on the likelihood contribution for those observations who are 11 years old in 1976. For $i \in c_{11}$

$$\begin{aligned} L_i(\theta) &= \Pr(\text{Observed Sequence of Choices, States and Wages for individual } i) \quad (3.22) \\ &= \Pr(d_i, s_i, w_i | \theta) \end{aligned}$$

[24] Note, however, that the main advantages of this method rely in reduction in computational burden. Whether is it useful to allow for a continuous parametric distribution of unobserved heterogeneity may be more debatable.

To account for unobserved heterogeneity, I integrate out over the distribution of types. I use α, as the new argument for the (now heterogeneous) structural parameters in θ and get

$$L_i(\theta) = \int_\alpha \Pr(d_i, s_i, w_i|\alpha)\, f(\alpha|\theta)\, d\alpha \tag{3.23}$$

where $\alpha \sim N(\gamma, \Sigma)$ and we redefine θ to $\theta = (\gamma, \Sigma)$.

Then, given the markovian nature of the model, the likelihood contribution can be expressed as

$$L_i(\theta) = \int_\alpha \left[\prod_{a=11}^{21} \Pr\left(d_a, d_a^W w_a \middle| s_a; \alpha\right) \prod_{a=11}^{20} \Pr\left(s_{a+1}| s_a, d_a; \alpha\right) \right] f(\alpha|\theta)\, d\alpha$$

The transition probabilities for non-deterministic state variables $\Pr(s_{a+1}|s_a, d_a; \alpha)$ are described in the model section. In particular, note that state transitions cannot be factored out and estimated separately because they are assumed to depend on the unobserved heterogeneity. This is important because, for example, teens from more advantaged unobserved backgrounds may have an easier time improving F grades in school. Similarly, they may face lower probability of conviction upon arrest, conditional on having the same state variables and criminal activity.

As a notational convention, I use $\theta = \left\{\gamma, \Sigma_\eta\right\}$ to denote the vector of parameters I will estimate. Strictly speaking, these are not the structural parameters per-se but the parameters that characterize the distribution of structural parameters across the population in the random coefficient framework. Note that γ will parameterize the mean of the distribution of such structural parameters and Σ_η will parameterize their variance. Note that there will possibly be more than one γ for each of the original structural parameters, because the mean can be specified as a function of exogenous observables Z.[25] I assume the σs in Σ do not depend on Z so there is only one σ to parameterize population variability for each of the original structural parameters.

Since the individual specific parameters are not observable and cannot be estimated, I exploit the random coefficients framework to integrate out this parameter heterogeneity by assuming a distribution for α.

$$\alpha_i = \gamma Z_i + \Gamma_\eta \eta_i \tag{3.24}$$

$$\eta_i \sim N\left(0, I_{K_\alpha}\right) \tag{3.25}$$

where Γ_η is the Cholesky Decomposition of Σ_η so $\Gamma_\eta \Gamma'_\eta = \Sigma_\eta$[26]

Consider u_i a dummy argument for the structural parameters in α_i and use the change of variable

$$u = \gamma Z + \Gamma_\eta \eta \tag{3.26}$$

where

$$u|Z \sim f_{U|Z}\left(u|\gamma, \Sigma, Z\right) = N\left(\gamma Z, \Sigma_\eta\right) \tag{3.27}$$

then

$$L_i\left(\theta, Z_i\right) = \int \Pr\left(d_i, s_i, w_i|u\right) f_{U|Z}\left(u|\theta, Z_i\right) du \tag{3.28}$$

A straightforward approach, would approximate this likelihood contribution by simulating the above integral. While the simulation handles the problem of high di-

[25] In future work I plan to explore the effect of several observables charateristics available in the NYS.

[26] Given the richness of the model and the several parameters to be estimated I assume Σ_η to be diagonal. However, this is not restrictive because some of the Z_i could enter the parameterization of the conditional mean for different structural parameters in the vector α so, unconditionally, the joint distribution of α have non-zero correlations induced by the shared observables in Z_i.

mensionality of u, it does not alter the computational burden associated with the repeated solution to the DP problem. Indeed, given a set $\{\eta_{js}\}_{js=1}^{JS}$, for each new set of parameters (γ, Σ_η) the simulation draws for u would need to be re-computed using (3.26) and with those new draws for u, the JS dynamic programming problems used as input for the choice probabilities would need to be re-solved. Instead, following Ackerberg (2001) it is possible to reduce the computational burden substantially by using importance sampling to eliminate the need to continually solve the DP problems for each new parameter trial.

When using observables Z_i to enrich the population distribution of structural parameters, I define the importance sampling meta-density $g(u)$ as

$$g(u) = \sum_{i=1}^{N} g(u|Z_i) \Pr(Z_i) \tag{3.29}$$

Suppose I draw a person j at random from my sample and then I draw u conditional on the observed heterogeneity of person j, say Z_j. Now, if I draw people at random from my analysis sample, each observation i (and its corresponding vector Z_i) has equal probability of being selected. Therefore $\Pr(Z_i) = \frac{1}{N}$. Also, I can use $f_{U|Z}(u|Z, \gamma_1, \Sigma_1)$ as choice of functional form for $g(u|Z)$ .Then

$$\begin{aligned} g(u) &= \sum_{i=1}^{N} g(u|Z_i) \left(\frac{1}{N}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} g(u|Z_i) \\ &= \frac{1}{N}\sum_{i=1}^{N} f_{U|Z}\left(u|Z_i, \gamma_1, \Sigma_{\eta 1}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} \phi\left(u|Z_i, \gamma_1, \Sigma_{\eta 1}\right) \end{aligned} \tag{3.30}$$

with $\theta_1 = \left(\gamma_1, \Sigma_{\eta 1}\right)$ some arbitrarily given starting values for $\theta = (\gamma, \Sigma_\eta)$ parameterizing the mean and variance of the multivariate normal density ϕ. Then, multiplying and dividing by the importance sampling meta-density $g(u)$, we get

$$L_i(\theta, Z_i) = \int \left[\Pr(d_i, s_i, w_i | u) \frac{f_{U|Z}(u|\theta, Z_i)}{g(u)} \right] g(u)\, du \tag{3.31}$$

which can be simulated using a crude frequency simulator by drawing JS random draws of u from the meta-density $g(u)$ and using those draws to evaluate the expression in brackets inside the integral in (3.31). We get

$$L_i(\theta, Z_i) = \frac{1}{JS} \sum_{js=1}^{JS} \left[\Pr(d_i, s_i, w_i | u_{js}) \frac{f_{U|Z}(u_{js}|\theta, Z_i)}{g(u_{js})} \right] \tag{3.32}$$

$$L_i(\theta, Z_i) = \frac{1}{JS} \sum_{js=1}^{JS} \left[\pi(i|u_{js}) \frac{f_{U|Z}(u_{js}|\theta, Z_i)}{g(u_{js})} \right] \tag{3.33}$$

where for simplicity, the joint probability of individual $i's$ sequence of choices, transitions and wages under structural parameters u_{js}, $\Pr(d_i, s_i, w_i | u_{js})$ is now denoted by $\pi(i|u_{js})$

3.5.3 Initial Conditions

As already mentioned, the task of computing the likelihood contribution for those who are older than 11 in 1976 is further complicated by some unobservable state variables in the first sample period. In particular, I do not know their legal experience (X) and criminal capital (CK) upon entrance into the sample because I do not observe previous work and crime choices, and no retrospective information is collected in the NYS. One option would be to discard the data before age 17 and set that as the baseline age for the model. This would simplify things but would unnecessarily waste much of the data available in the NYS. More importantly, as can be seen in Figure 3.3 in Appendix C, an important part of dynamic behavior regarding criminal activity occurs between 11 and 17. It is preferable to have a model that can account for this well known relationship between age and crime.[27]

In principle, I could use data on those who are 11 years old in 1976. For these I observe the complete behavioral history from 11 to 21 and I am able to estimate a distribution at each age $\widehat{f_a}(X,CK)$. I could use this estimate to integrate out the unobserved state variables for the rest of the sample. However, the relevant distribution is not just the marginal of the unobserved state vector at any age between 12 and 17. Instead, because I am using importance sampling I need the conditional probability of the (partially) unobserved state vector s_a^u, conditional on the unobserved heterogeneity in the structural parameters (α), and the observed state vector s_a^o.[28] Therefore I need to estimate $f_a(X,CK|\alpha,s_a^o)$. As explained below, conditioning on s_a^o prevents straightforward application of Ackerberg's (2001) computationally convenient estimation framework. I therefore provide an estimator that retains the computational advantages of Ackerberg's (2001) estimator in situations like mine, where endogenous initial conditions arise because for most observations the data generating process starts before the available sample window.

To come up with this joint conditional probability I adopt a forward simulation strategy. At each structural parameter draw, after solving the DP problem, and before computing the likelihood, I simulate histories of behavior consistent with that structural parameter draw.

[27]Fortunately, for some observations the sample start at 11 years of age, once stable heterogeneous traits and preferences across the population have been established. In Gottfredson and Hirsch's (1990) general theory of crime, for example, at a fixed level of opportunity, individual differences in criminal offending are attributed to differences in an individual-level attribute they term self-control. Self-control, the capacity to resist the temptation of immediate and easy gratification, is presumed to be the result of effective socialization by an attentive, involved, and conventional care giver. The implication is that self-control must be created. According to Gottfredson and Hirschi, the window for the development of self-control is fairly short. They suggest that it closes by age 8 or 10. See also Nagin & Paternoster (2002).

[28]If not using importance sampling I would still need $f\left(s_{17}^u,\alpha|s_{17}^o\right)$, the joint density of unobserved state variables and unobserved heterogeneity , conditional on the observed state variables.

For concreteness, consider individuals who are 17 years old in 1976 (i.e. $i \in c_{17}$). Only some state variables are not observed in the initial sample period. It is helpful to distinguish these two types of state variables. I partition the vector of first sample period state variables into observable (o) and unobservable (u) components : $s_{17} = \left(s_{17}^{o}, s_{17}^{u}\right)$.

The likelihood contribution for $i \in c_{17}$ is given by

$$L_i\left(\theta, Z_i, s_{17_i}^{o}\right) = \Pr\left(d_i, s_i^o, w_i | \gamma, \Sigma_\eta, Z_i, s_{17}^{o}\right)$$

Again consider u_i a dummy argument for the structural parameters in α_i and use the change of variable

$$u = \gamma Z + \Gamma_\eta \eta \tag{3.34}$$

where

$$u|Z_i \sim f_{U|Z}\left(u|\gamma, \Sigma, Z_i\right) = N\left(\gamma Z_i, \Sigma_\eta\right) \tag{3.35}$$

then

$$L_i\left(\gamma, \Sigma_\eta, Z_i, s_{17_i}^{o}\right) = \int \Pr\left(d_i, s_i^o, w_i | u, s_{17_i}^{o}\right) f\left(u|\theta, Z_i, s_{17_i}^{o}\right) du \tag{3.36}$$

$$\int_u \left[\int_{s_{17}^u} \Pr\left(d_i, s_i^o, w_i | u, s_{17}^{o}, s_{17}^{u}\right) f\left(s_{17}^{u}|u, s_{17}^{o}\right) ds_{17}^{u}\right] f\left(u|\theta, Z_i, s_{17}^{o}\right) du \tag{3.37}$$

Multiplying and dividing by the same importance sampling meta-density $g(u)$, we get

$$\int_u \left\{\left[\int_{s_{17}^u} \Pr\left(d_i, s_i^o, w_i | u, s_{17_i}^{o}, s_{17}^{u}\right) f\left(s_{17}^{u}|u, s_{17_i}^{o}\right) ds_{17}^{u}\right] \frac{f\left(u|\theta, Z_i, s_{17_i}^{o}\right)}{g(u)}\right\} g(u)\, du \tag{3.38}$$

and noting that s_{17}^{u} is actually discrete we get

$$\int_u \left\{\left[\sum_{s_{17}^u} \Pr\left(d_i, s_i^o, w_i | u, s_{17_i}^{o}, s_{17}^{u}\right) p\left(s_{17}^{u}|u, s_{17_i}^{o}\right)\right] \frac{f\left(u|\theta, Z_i, s_{17_i}^{o}\right)}{g(u)}\right\} g(u)\, du \tag{3.39}$$

where

$$\Pr\left(d_i, s_i^o, w_i | u, s_{17}^u, s_{17_i}^o\right) = \Pr\left(\begin{array}{c|c} d_{27}, \ldots . d_{17}, & \\ d_{27}^W w_{27}, \ldots . d_{19}^W w_{19}, & u, s_{17}^u, s_{17_i}^o \\ s_{27}^o, \ldots . s_{18}^o, & \end{array}\right) \tag{3.40}$$

$$\Pr\left(d_i, s_i^o, w_i | u, S_{17}^u, s_{17_i}^o\right) = \prod_{a=17}^{27} \Pr\left(d_a, d_a^W w_a \middle| S_a; u, s_{17}^u\right) \prod_{a=17}^{26} \Pr\left(S_{a+1} \middle| S_a, d_a; u, s_{17}^u\right)$$

Then, letting the probability of the sequence of choices, transitions and outcomes $\Pr\left(d_i, s_i^o, w_i | u_{js}, s_{17_i}^o\right)$ be denoted by $\pi\left(i | u_{js}\right)$ and noting that $\pi\left(i | u_{js}\right)$ already integrates out unobserved first period state variables s_{17}^u using $p\left(s_{17}^u | u, s_{17_i}^o\right)$ we get

$$L_i\left(\theta | Z_i, s_{17_i}^o\right) = \int_u \left\{ \pi\left(i | u_{js}\right) \frac{f\left(u | \theta, Z_i, s_{17_i}^o\right)}{g\left(u\right)} \right\} g\left(u\right) du \tag{3.41}$$

Direct application of Ackerberg's (2001) is not possible because conditioning on $s_{17_i}^o$ shifts the density used to evaluate u in an unknown manner. Below, I show that an estimator based on the embedding of two importance sampling simulators can be used to extend the insights of Ackerberg's (2001) to this context.

We again simulate using JS random draws of u from the importance sampling meta-density $g\left(u\right)$

$$\widehat{L}_i\left(\theta | Z_i, s_{17_i}^o\right) = \frac{1}{JS} \sum_{js=1}^{JS} \left\{ \pi\left(i | u_{js}\right) \frac{f\left(u_{js} | \theta, Z_i, s_{17_i}^o\right)}{g\left(u_{js}\right)} \right\} \tag{3.42}$$

Now, in practice note that the problem of initial conditions arises from the fact that $f\left(u | \theta, Z_i, s_{17_i}^o\right)$ and $p\left(s_{17}^u | u, s_{17_i}^o\right)$ are unknown. To come up with estimates of these I adopt the following steps.

- To get $\widehat{p}\left(s_{17}^{u}|u,s_{17_i}^{o}\right)$

  1. For a given simulated draw of structural parameters u_{js}, after solving the DP problem I simulate histories of agents behaving according to the model under these parameters.

  2. Using those histories I then estimate the joint probability mass function $p\left(s_{17}^{u}s_{17}^{o}|u_{js}\right)$, say $\widehat{p}\left(s_{17}^{u}s_{17}^{o}|u_{js}\right)$, by crude frequency over the simulated paths.

  3. I can then focus on the conditional probability $\widehat{p}\left(s_{17}^{u}|s_{17}^{o},u_{js}\right)$ relevant for each observation depending on its observed state vector at age 17, $s_{17_i}^{o}$, by using
$$\widehat{p}\left(s_{17}^{u}|s_{17}^{o},u_{js}\right)=\frac{\widehat{p}\left(s_{17}^{u}s_{17}^{o}|u_{js}\right)}{\widehat{p}\left(s_{17}^{o}|u_{js}\right)} \tag{3.43}$$
     with
$$\widehat{p}\left(s_{17}^{o}|u\right)=\sum_{S_{17}^{u}}\widehat{p}\left(s_{17}^{u}s_{17}^{o}|u\right) \tag{3.44}$$

  4. I use $\widehat{p}\left(s_{17}^{u}|u,s_{17_i}^{o}\right)$ to integrate out the unobserved (at time of sample entrance) state variables for individuals from cohort $c17$ with $s_{17_i}^{o}$ and structural parameters u.

- To get $\widehat{f}_{U|Z}\left(u|\theta,Z_i,s_{17}^{o}\right)$

  1. Note
$$\begin{aligned} f\left(u|s_{17}^{o},\gamma,\Sigma,Z_i\right) &= \frac{f\left(u,s_{17}^{o}|\theta,Z_i\right)}{f\left(s_{17}^{o}|\theta,Z_i\right)} \\ &= \frac{f\left(s_{17}^{o}|u,\theta,Z_i\right)}{f\left(s_{17}^{o}|\theta,Z_i\right)}f\left(u|\theta,Z_i\right) \\ &= \frac{f\left(s_{17}^{o}|u\right)}{f\left(s_{17}^{o}|\theta,Z_i\right)}f\left(u|\theta,Z_i\right) \\ &= \frac{p\left(s_{17}^{o}|u\right)}{p\left(s_{17}^{o}|\theta,Z_i\right)}f\left(u|\theta,Z_i\right) \end{aligned} \tag{3.45}$$

where we at least got rid of s_{17}^{o} as conditioning factor in the density used to evaluate u. The bad news is that θ appear twice: Not only it shows up in $f(u|\theta, Z_i)$ as before, but also in the denominator of the adjustment factor $\frac{p(s_{17}^{o}|u)}{p(s_{17}^{o}|\theta, Z_i)}$. This brings the computational burden back as $p\left(s_{17}^{o}|\theta, Z_i\right)$ can only be recovered by simulation for each trial θ_r, and new DP solutions would have to be recomputed each time. But this is exactly what we were trying to avoid.

2. Note that the main problem is in the denominator of $\frac{p(s_{17}^{o}|u)}{p(s_{17}^{o}|\theta, Z_i)}$ as $p\left(s_{17}^{o}|u\right)$ can be easily estimated by simulation for given u. If we were able to compute the denominator cheaply, we would be able to avoid repeated DP solutions. To that end, note that

$$\begin{aligned} p(s_{17}^{o}|\theta, Z_i) &= \int_u p(s_{17}^{o}|\theta, Z_i, u) f(u|\theta, Z_i) \, du \\ &= \int_u p(s_{17}^{o}|u) f(u|\theta, Z_i) \, du \end{aligned}$$

because once we condition on u, θ and Z_i are no longer relevant. Then, we can multiply and divide by the same importance sampling meta-density to get

$$p(s_{17}^{o}|\theta, Z_i) = \int_u \left[p(s_{17}^{o}|u) \frac{f(u|\theta, Z_i)}{g(u)} \right] g(u) \, du$$

and we can simulate this integral exploiting the same set of JS simulation histories based on the same JS solutions to the dynamic program as

$$\widehat{p}(s_{17}^{o}|\theta, Z_i) = \frac{1}{JS} \sum_{ls=1}^{JS} \left[p(s_{17}^{o}|u_{ls}) \frac{f(u_{ls}|Z_i)}{g(u_{ls})} \right]$$

which, critically, does not involve additional computational burden given that all the elements inside the brackets need to be computed at some point anyways.

3. So,

$$f\left(u|s_{17}^{o},\theta,Z_i\right)=\left[\frac{p\left(s_{17}^{o}|u\right)}{\frac{1}{JS}\sum_{ls=1}^{JS}\left[p\left(s_{17}^{o}|u_{ls}\right)\frac{f(u_{ls}|\theta,Z_i)}{g(u_{ls})}\right]}\right]f\left(u|\theta,Z_i\right) \quad (3.46)$$

4. So we can estimate $f\left(u|\theta,Z_i,s_{17}^{o}\right)$ by plugging-in the estimator in (3.44) inside the expression in brackets in (3.46). We get,

$$\widehat{f}\left(u|\theta,Z_i,s_{17_i}^{o}\right)=\left[\frac{\widehat{p}\left(s_{17}^{o}|u\right)}{\frac{1}{JS}\sum_{ls=1}^{JS}\left[\widehat{p}\left(s_{17}^{o}|u_{ls}\right)\frac{f(u_{ls}|\theta,Z_i)}{g(u_{ls})}\right]}\right]f\left(u|\theta,Z_i\right) \quad (3.47)$$

Intuitively, here we are adjusting the numerator of the weight upwards, and therefore the weight itself, whenever the particular observation i turns out to have an observed state vector at age 17, $s_{17_i}^{o}$, that is highly consistent with the structural parameter draw u_{js}. In other words, when $s_{17_i}^{o}$ is (on average) more likely to be the observed cumulated behavior under u_{js} than under other structural parameter draw u_{ls} for all l. Also note that $\widehat{p}\left(s_{17}^{o}|u_{ls}\right)$ for $ls=1,....,JS$ have been already computed and stored during the initial stage when a forward simulation was conducted after each DP solution to integrate out unobserved states and compute the $\pi's$.

5. $f\left(u|\theta,Z_i\right)$ is the familiar numerator for the weight, the one we have in the case without initial conditions problem, as in Ackerberg (2001).

Finally, the individual likelihood contribution can be computed embedding importance sampling simulators as,

$$\widehat{L}_i\left(\theta|Z_i,s_{17_i}^{o}\right)=\frac{1}{JS}\sum_{js=1}^{JS}\left\{\pi\left(i|u_{js}\right)\frac{f\left(u_{js}|\theta,Z_i\right)}{g\left(u_{js}\right)}\left(\frac{\widehat{p}\left(s_{17}^{o}|u_{js}\right)}{\frac{1}{JS}\sum_{ls=1}^{JS}\left[\widehat{p}\left(s_{17}^{o}|u_{ls}\right)\frac{f(u_{ls}|\theta,Z_i)}{g(u_{ls})}\right]}\right)\right\} \quad (3.48)$$

These steps are similar for those at age 12, 13 ,14, 15 and 16 in 1976 and for those age 17 but with different S_{17}^{o} . The same forward simulation procedure applies. Note that the simulation task is greatly complicated by the fact that some states are *observed*. Therefore we need to simulate conditional probabilities. This requires a bigger number of simulations.

3.6 Parameter Estimates, Model Fit and Validation

Here I briefly describe the most interesting parameter estimates. Full estimation results are reported in Appendix B. In the current specification without Zs there are 43 parameters in total. 15 parameters are associated with the utility function and 10 are used to describe the earnings equation and the job offer probability. The remaining parameters are used to specify the various transition probabilities. Most of the parameters show the sign expected a priori. The positive coefficient on criminal capital indicates true state dependence coming from either strong habit persistence, lock-in effects and/or high returns to criminal capital. Having a criminal record reduces both the arrival rate of job offers and legitimate earnings. The estimated earnings equation imply a return to education of about 7% and return to experience of about 4%. The interaction of educational attainment and the indicator for the presence of a criminal record is negative implying that the stigma of a criminal record hits harder on those with more education. This latter result is consistent with some of the arguments advanced by Waldfogel (1994).

Regarding parameters in the transition probabilities, having an episode of incarceration in the period strongly reduces the chances of both getting As/Bs and successfully completing a grade. These results are in line with those of Pintoff (2006). There is high persistence in GPA. Having good grades greatly increases the chances of having good grades next period and it is also a significant determinant of successful grade completion.

Figures 3.2 (Work and School Attendance) and 3.3 (Criminal Activity) describing model fit are shown in Appendix C. The figures compare the basic age behavior profiles found in the NYS data with profiles generated by simulation using the estimated model. The model fits the basic age-behavior patterns in the data reasonably well but it tends to slightly overestimate the fraction of individuals working trough their '20s. While patterns of model fit are a useful diagnostic device, their scientific value has recently come under attack, as substantial model pre-testing is undertaken by the econometrician to modify the model in order to fit some basic patterns in the data. As a result, alternative validation strategies have been advanced recently, including the use of social experiments[29] or non-random "holdout samples".[30] Since none of these opportunities are available in my context, I pursue a different strategy which involves data on self reported expectations of college completion. While it is possible to include these data in the likelihood function, it is perhaps more useful to reserve these data to provide a validation test for one of the model's key underlying mechanisms:[31] Everything else constant, increases in the chances of completing a college degree in the future should lead to declines in current criminal activity. This happens because individuals recognize the asymmetric incentives embedded in a open records policy, which induce them to stay away from crime when the prospects for additional human capital accumulation become more favorable.

[29]See Todd & Wolpin (2007)

[30]See Keane & Wolpin (2005). The idea here is to explore how the model estimated with the remaining sample predicts behavior in the "holdout sample". This approach is especially compelling when the holdout subsample undergoes some regime change not present in the estimating sample.

[31]See Fang, Keane, Khawaja, Salm & Silverman (2007) for a related strategy used to validate Khwaja's (2007) model. The basic idea here is to exploit some pieces of data not included in the statistical likelihood function to test the mechanisms underlying the structural model.

3.6.1 Expectations Data

The following question was asked of all NYS respondents in the first five waves.

"What do you think your chances are for completing a college degree?"

Unfortunately, the possible answers to this question were only: 1) Poor, 2) Fair, 3) Good, and 4) I don't know.[32] [33]

The model predicts that increases in the chances of eventual college completion should lead to declines in current crime as individuals who plan to accumulate high levels of education recognize the unfavorable incentives embedded in an open criminal records policy in the event of conviction.

I then consider the following model

$$Crime_{it} = \beta_{0_i} + \beta_1 Fair_{it} + \beta_2 Good_{it} + g\left(Age_{it}\right) + \varepsilon_{it} \tag{3.49}$$

where Fair$_i$ and Good$_i$ are equal to 1 when the respondent reports such chances of college completion and equal 0 otherwise.

Of course, identification of the causal impact of better college prospects on current crime might be threatened by unobserved heterogeneity. Moreover, even if the actual chances of college completion are not correlated with the error term in (3.49) the self-reports we observe might be. In particular, there might be heterogeneity in what different respondents mean by "Poor", "Fair" or "Good" chances of college completion. For example, two respondents might have the same actual chances but one of them

[32] I exclude the few cases that report "I don't know" from the analysis.

[33] Data on expectations of future events provide similar information to current or past observed behavior. My model of human capital accumulation and criminal behavior delivers college completion probabilities. Moreover, those probabilities can be computed conditional on the information set available at any age/time. It is possible in principle, to use the model to compute what those college completion probabilities are, as computed from the perspective of respondents at particular ages and with particular values for the state variables and introduce the expectations data into the likelihood function.

might have a more generous definition of the category "Good" and include himself in it. This spurious variation might be correlated with other unobserved determinants biasing the estimation of the causal effects of interest. To ameliorate these problems I exploit a unique feature of the NYS subjective expectations data: I actually observe up to 5 self-reports of chances of college completion for the same respondents over time. Therefore, I can more reliably exploit the *changes* in these college completion prospects as source of identifying variation when estimating the model in (3.49). Table 3.4 presents both OLS and Fixed Effects estimates for quadratic and non-parametric specifications of $g\left(Age_{it}\right)$.

Table 3.4: Testing a Mechanism of the Dynamic Model: The Effect of Better Prospects for College Completion on Current Crime

	OLS	Fixed Effects	OLS	Fixed Effects
Chances of College Completion ?				
Fair	-0.164***	-0.066**	-0.163***	-0.064**
	[0.033]	[0.032]	[0.033]	[0.032]
Good	-0.179***	-0.035	-0.177***	-0.031
	[0.032]	[0.036]	[0.032]	[0.036]
Age Effects	Quadratic		Non-Parametric	

Robust standard errors in brackets.* significant at 10%; ** significant at 5%; *** significant at 1%. The omitted category is "poor" chances of college completion. Dependent Variable =1 if criminally active in the period, =0 otherwise.

As can be seen in the table, relative to a base category of "Poor" chances of college completion, increasing such chances to "Fair" or "Good" has a substantial negative impact on the probability of being criminally active. The fixed effects estimates are smaller but remain sizable. Moreover, the coefficient on "Fair" is still significant, thus providing some support to one of the key underlying mechanisms in the model.

3.7 Policy Experiments

Below I exploit the estimated model to perform two policy experiments useful in addressing the research questions posed in this chapter. The simulations are steady state simulations: these are the counterfactual histories that we would observe for a given cohort, when exposed to alternative regimes from the very beginning (from age 11 onwards).[34]

I experiment with the choices for sealing and opening criminal histories for adults and juveniles. For example, there seem to be two radical deviations from the status quo:

1. Everything open (that is, to open juvenile records, whereas adult ones are already open).

2. Everything sealed (that is, to seal adult records, whereas juvenile ones are already sealed).

Then I can observe the net effect on crime. Does the dynamic deterrent effect caused by the change in prospective labor market stigma outweigh the recidivism effect? This is policy relevant because it informs about the magnitude of ”entry” effects for younger cohorts that will be exposed to the new regime from the beginning, thus shedding light on the longer run, potentially unintended consequences of these policies.

[34]Simulations do not take into account employer response to the change in criminal records policy. For example, the more limited informational structure embedded in a fully sealed records policy may induce employers to devise alternative ways to get at the valuable information about a prospective employee’s reputation. In particular, they will engage more seriously in statistical discrimination. Those who belong in groups that are more likely to hold a criminal record will still suffer some reduction in earnings and employment probabilities even if records are perfectly sealed. While not addressed in the chapter, accounting for employer response is an straightforward extension in this framework. It involves finding a fixed point between employers’ beliefs and individual behavior when simulating a sealed records policy.

Figure 3.1 shows the crime-age profile corresponding to cohorts behaving under alternative criminal records policy regimes. The status quo corresponds to the baseline simulation using the estimated parameters from Section 3.6. In the status quo, convictions are transparent to the labor market only when received in adult courts.[35] In addition, the figure shows the resulting crime-age profile that corresponds to cohorts facing two radically opposite set of incentives. Under fully open records, not only adult but also juvenile convictions are made transparent to the labor market. Under this regime the fraction criminally active at every age declines substantially, especially during the early teenage years. Two effects might be operating to generate such results. First, dynamic deterrence kicks in from the beginning. Parents might be stricter with their children and warn them more severely about the long run consequences of any misbehavior. Second, as long as this ”stigma-deterrence” prevents youngsters from being criminally active, the state dependence mechanism that operates through criminal capital cannot be set in motion.

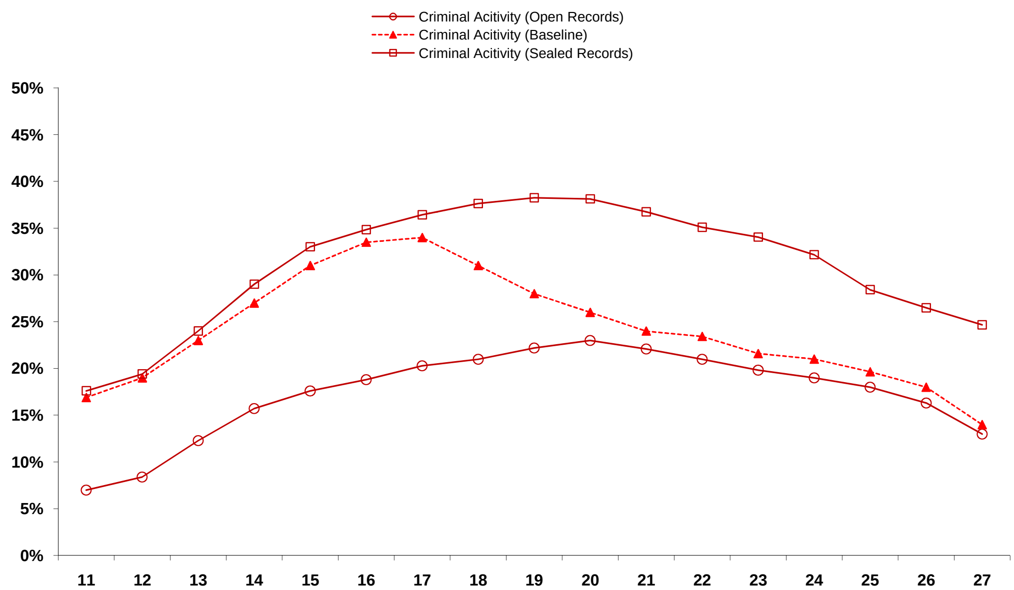
Under Full Sealing, instead, not only juvenile but also adult records now remain sealed and opaque to the labor market. Figure 3.1 shows an increase in the fraction criminally active, especially between the ages of 18 and 27. The increase in the fraction criminally active due to the loss of dynamic deterrence more than outweighs any crime reductions due to better prospects for reintegration into the labor market.

Also, it is interesting to explore to what extent these alternative criminal records policies might induce different correlation patterns between education and crime. As explained before, more education might lead to less crime because

1. education increases the opportunity costs of crime. More educated individuals

[35] In implementing the model I use age 18 as the cutoff to separate juvenile and adult convictions. This is by far the modal age of criminal majority during the period covering the estimating sample. During the ’90s however, there has been a trend towards reducing the age at which offenders start appearing in adult courts.

Figure 3.1: Basic Age-Crime Profiles under Full Opening, Full Sealing and Status-Quo



have more to gain from using their time in legal activities.

2. schools keep children supervised in critical years preventing the triggering of state dependence in criminal activity.

In accordance with the hypothesis advanced earlier in this chapter, an alternative mechanism for the negative causal effect of education on crime is the one operating via the differential deterrent effect of labor market stigma. The degree of educational stratification in criminal activity will be potentially influenced by the idiosyncrasies of those institutions regulating employers' access to the criminal records of their potential employees.

Table 3.5 shows the impact of alternative criminal records policies on the degree of educational stratification in criminal activity. More educated individuals expect to lose more when and if their records are made transparent. This is so because many professional certifications require a clean criminal record and, more generally, because more educated individuals are appointed to positions with higher trust and responsibility requirements.[36] In a sense, more educated individuals have more at stake when it comes to maintaining their reputations.

Given the above arguments, I expect the sealing of adult records to reduce the gap in criminal activity between groups of different education levels. In fact, Table 3.5 shows that the gap in mean criminal capital accumulated by age 27 is somewhat reduced under a full sealing policy. The average person with less than a high school diploma has only 80% more criminal capital than the average person with more than a high school diploma. Conversely, full opening maximizes this distance: the criminal capital of those who are less educated is about 6 times higher than the one for the more educated group. The status quo gap, involving a mixture of open and sealed records is somewhat in between.

Table 3.5: Mean Criminal Capital Accumulated by Age 27 Under Alternative Criminal Records Policies

	Statu-Quo	Full Open Records	Full Sealing
Less than HS	7.55	4.88	7.39
HS or More	1.51	0.83	4.11

3.8 Conclusions

Results indicate that open criminal records policies exert substantial dynamic deterrence via threat of stigmatization in the labor market. Sealing adult records tends to

[36]See Waldfogel (1994)

reduce recidivism but it also raises the level of first-offenses for those individuals that would remain law abiding under the statu-quo. The magnitude of the latter effect is substantial and has important implications for the evaluation of changes in criminal records policy. The net result is an overall counterfactual increase in the fraction of the population criminally active. The strength of this ”stigma-deterrence” seems to be particularly important for those whose unobserved traits make them likely to accumulate higher levels of education. The well known negative correlation between measures of crime and schooling can be explained in part by this process of dynamic sorting. This latter finding has implications for the design of policies that focus on schooling as a crime control device. It suggests that open records are an important complement to such policies. Additional human capital will be more effective in deterring criminal activity when its full potential can be more easily jeopardized by such activity.

3.9 Appendix A: Likelihood Function

Remember I observe choices every period. I only observe wages (if they chose to work) from the 3rd wave onwards. Therefore

$$\Pr(d_i, s_i, w_i|\alpha) = \Pr\left(\begin{array}{c} d_{21}, \ldots .d_{11}, \\ d_{21}^W w_{21}, \ldots .d_{13}^W w_{13}, \\ s_{21}, \ldots .s_{12}, \end{array} \middle| \alpha \right) \tag{3.50}$$

Factorizing the joint probability and using the markovian nature of the model we can simplify the expression to

$$\begin{aligned} \Pr(d_i, s_i, w_i|\alpha) &= \Pr\left(d_{21}, \ldots .d_{11}, d_{21}^W w_{21}, \ldots .d_{13}^W w_{13}, s_{21}, \ldots .s_{12} \middle| \alpha\right) \\ &= \Pr\left(d_{21}, d_{21}^W w_{21} \middle| d_{20}, \ldots .d_{11}, d_{20}^W w_{20}, \ldots .d_{13}^W w_{13}, s_{21}, \ldots .s_{12}; \alpha\right) \\ &\quad \Pr\left(d_{20}, \ldots .d_{11}, d_{20}^W w_{20}, \ldots .d_{13}^W w_{13}, s_{21}, \ldots .s_{12}|\alpha\right) \\ &= \Pr\left(d_{21}, d_{21}^W w_{21} \middle| s_{21}; \alpha\right) \\ &\quad \Pr\left(s_{21}|d_{20}, \ldots .d_{11}, d_{20}^W w_{20}, \ldots .d_{13}^W w_{13}, s_{20}, \ldots .s_{12}; \alpha\right) \\ &\quad \Pr\left(d_{20}, \ldots .d_{11}, d_{20}^W w_{20}, \ldots .d_{13}^W w_{13}, s_{20}, \ldots .s_{12}|\alpha\right) \\ &= \Pr\left(d_{21}, d_{21}^W w_{21} \middle| s_{21}; \alpha\right) \\ &\quad \Pr\left(s_{21}|d_{20}, s_{20}; \alpha\right) \\ &\quad \Pr\left(d_{20}, \ldots .d_{11}, d_{20}^W w_{20}, \ldots .d_{13}^W w_{13}, s_{20}, \ldots .s_{12}|\alpha\right) \end{aligned} \tag{3.51}$$

Repeating this process I obtain

$$\Pr(d_i, s_i, w_i|\alpha) = \prod_{a=11}^{21} \Pr\left(d_a, d_a^W w_a \middle| s_a; \alpha\right) \prod_{a=11}^{20} \Pr\left(s_{a+1} \middle| s_a, d_a; \alpha\right) \tag{3.52}$$

3.9.1 Unobserved state variables in First (sample) Period

$$\Pr\left(d_i, s_i^o, w_i | u, s_{17}^u, s_{17}^o\right) = \Pr\left(\begin{array}{c|c} d_{27}, \ldots . d_{17}, & \\ d_{27}^W w_{27}, \ldots . d_{19}^W w_{19}, & u, s_{17}^u, s_{17}^o \\ s_{27}^o, \ldots . s_{18}^o, & \end{array}\right) \tag{3.53}$$

$$\begin{aligned}
\Pr\left(d_i, s_i^o, w_i | u, s_{17}^u, s_{17}^o\right) &= \Pr\left(d_{27} \ldots d_{17}, d_{27}^W w_{27} \ldots d_{19}^W w_{19}, s_{27}^o \ldots s_{17}^o \middle| u, s_{17}^u, s_{17}^o\right) \qquad (3.54)\\
&= \Pr\left(d_{27}, d_{27}^W w_{27} \middle| d_{26} \ldots d_{17}, d_{26}^W w_{26} \ldots d_{13}^W w_{13}, s_{27}^o \ldots s_{17}^o ; u, s_{17}^u, s_{17}^o\right) \\
&\quad \Pr\left(d_{26}, .., d_{17}, d_{26}^W w_{26}, .., d_{13}^W w_{13}, s_{27}^o, .., s_{17}^o | u, s_{17}^u, s_{17}^o\right) \\
&= \Pr\left(d_{27}, d_{27}^W w_{27} \middle| s_{27}^o, d_{26}, \ldots . d_{17} ; u, s_{17}^u, s_{17}^o\right) \\
&\quad \Pr\left(s_{27}^o | d_{26}, \ldots . d_{17}, s_{26}^o, \ldots . s_{17}^o ; u, s_{17}^u, s_{17}^o\right) \\
&\quad \Pr\left(d_{26}, \ldots . d_{17}, d_{26}^W w_{26}, \ldots . d_{13}^W w_{13}, s_{26}^o, \ldots . s_{17}^o | u, s_{17}^u, s_{17}^o\right) \\
&= \Pr\left(d_{27}, d_{27}^W w_{27} \middle| u, s_{27}\left(s_{27}^o, d_{26}, \ldots ., d_{17}, s_{17}^u, s_{17}^o\right)\right) \\
&\quad \Pr\left(s_{27}^o | d_{26}, s_{26}^o ; u, s_{17}^u\right) \\
&\quad \Pr\left(d_{26}, \ldots . d_{17}, d_{26}^W w_{26}, \ldots . d_{17}^W w_{17}, s_{26}^o, \ldots . s_{17}^o ; u, s_{17}\right)
\end{aligned}$$

Repeating this process I obtain

$$\Pr\left(d_i, s_i^o, w_i | s_{17}^u, s_{17}^o, u\right) = \prod_{a=17}^{27} \Pr\left(d_a, d_a^W w_a \middle| s_a ; s_{17}^u, u\right) \prod_{a=17}^{26} \Pr\left(s_{a+1} \middle| s_a, d_a ; s_{17}^u, u\right) \tag{3.55}$$

3.10 Appendix B: Estimation Results

Table 3.6: Earnings Equation

	Estimate (Mean)	SE (Mean)	Estimate (Std.Dev.)	Description
α_0	8.24	0.015	0.093	constant
α_1	0.04	0.008	0.072	coef on x
α_2	-0.006	0.0008	0.009	coef on x^2
α_3	0.07	0.010	0.071	coef on att
α_4	-0.01	0.011	0.009	coef on cr2*att
α_5	-0.12	0.062	0.055	coef on cr2
α_6	0.05	0.023	0.025	HSdip in wage
σ_{W0}	1.44	0.208	0.157	Std. Dev. of Wage Shock

The first column shows the point estimates for the mean of the structural parameters. 3rd column shows the associated standard deviation in the in the population distribution fo each parameter. Asymptotic Standard Errors were computed using BHHH.

Table 3.7: Job Offer Probability

	Estimate (Mean)	SE (Mean)	Estimate (Std.Dev.)	Description
λ_0	2.34	0.15	0.25	Constant
λ_1	-1.10	0.22	0.28	Coef. on I(CR==2)
Pr(Job Offer)				
CR=1 or CR=0	91%			
CR=2	78%			

The first column shows the point estimates for the mean of the structural parameters. 3rd column shows the associated standard deviation in the in the population distribution fo each parameter. Asymptotic Standard Errors were computed using BHHH.

Table 3.8: Probability of Successful Grade Completion (Logit)

	Estimate (Mean)	SE (Mean)	Estimate (Std.Dev.)	Description
Ψ_0	-4.20	0.03	0.27	constant
Ψ_1	0.67	0.11	0.17	coef on Age
Ψ_2	-0.31	0.37	0.38	coef on Attainment
Ψ_3	1.44	0.02	0.09	coef on Good Student (gpa==B or A)
Ψ_4	0.61	0.10	0.10	coef on (dW=1)
Ψ_5	-2.94	0.04	0.23	coef on (fs=1) < 0

The first column shows the point estimates for the mean of the structural parameters. 3rd column shows the associated standard deviation in the in the population distribution fo each parameter. Asymptotic Standard Errors were computed using BHHH.

Table 3.9: Transition Probability for CJS Outcomes (Ordered Logit)

	Estimate (Mean)	SE (Mean)	Estimate (Std.Dev.)	Description
μ_{CJS_0}	-30.2	0.005	0.50	1st cutoff in CJS OLOGIT
μ_{CJS_1}	-27.0	0.01	0.72	2nd cutoff in CJS OLOGIT
λ_{CJS_1}	-3.11	0.13	0.248	coef on cr
λ_{CJS_2}	-3.54	0.25	0.384	coef on fs
λ_{CJS_3}	-3.72	0.38	0.503	coef on crime
λ_{CJS_4}	0.37	0.12	0.095	coef on att
λ_{CJS_5}	-2.26	0.05	0.194	coef on age
λ_{CJS_6}	0.06	0.03	0.0196	coef on age2

The first column shows the point estimates for the mean of the structural parameters. 3rd column shows the associated standard deviation in the in the population distribution fo each parameter. Asymptotic Standard Errors were computed using BHHH.

Table 3.10: Transition Probability for GPA (Logit)

	Estimate (Mean)	SE (Mean)	Estimate (Std.Dev.)	Description
λ_{GPA_0}	-1.50	0.294	0.28	constant in GPA LOGIT
λ_{GPA_1}	4.22	0.022	0.23	coef. on GPA=1 in GPA LOGIT
λ_{GPA_2}	0.12	0.086	0.08	coef. on d^w=1 in GPA LOGIT
λ_{GPA_3}	-53.07	0.711	2.43	coef. on fs=1 in GPA LOGIT

The first column shows the point estimates for the mean of the structural parameters. 3rd column shows the associated standard deviation in the in the population distribution fo each parameter. Asymptotic Standard Errors were computed using BHHH.

Table 3.11: Utility Function

	Estimate (Mean)	SE (Mean)	Estimate (Std.Dev.)	Parameter Description
γ_{C0}	-6112	7533	30.0	base Utility from Criminal Activity
γ_{C1}	454	2061	3.7	coef. on age^2
γ_{C3}	-27.8	51426	0.8	coef. on ck
γ_{C4}	-1495	3162	8.1	Witte parameter in school
σ_{C0}	582	25.9	2.5	Std. Dev. of Shock to Utility from Criminal Activity
γ_{L}	17534	4297	24.3	base Utility from Leisure
γ_{L1}	3281	2657	9.7	extra utility from leisure when young
σ_{L0}	1403	8.0	7.2	Std. Dev. of Shock to Utility from Leisure
γ_{S0}	10373	2885	20.9	base Utility from Schooling
γ_{S1}	-954	339	2.2	stigma of lagging in school.
γ_{S_C}	5520	2351	11.6	cost of College
γ_{S_GS}	22098	4177	22.4	cost of Grad School
σ_{S0}	474	13	10.5	std. dev. of Shock to Utility from School
γ_{P}	-316466	36507	104.7	base Utility from Freedom (Disutility of Prison)

The first column shows the point estimates for the mean of the structural parameters. 3rd column shows the associated standard deviation in the in the population distribution fo each parameter. Asymptotic Standard Errors Deviations were computed using BHHH.

Table 3.12: Discount Factor

	Estimate (Mean)	SE (Mean)	Estimate (Std.Dev.)	Description
Auxiliar Parameter	2.903	0.002	0.0668	
Beta	0.948			Discount Factor[a]

The first column shows the point estimates for the mean of the the auxiliar parameter. 3rd column shows the associated standard deviation in the population distribution of the auxiliar parameter. Asymptotic Standard Errors were computed using BHHH. In this case, the reported mean and standard deviation refers to that of an auxiliar parameter which is normally distributed in the population. The discount factor is restricted to the (0,1) interval using a logistic transformation on this auxiliar parameter. [a] The induced distribution for the discount factor is skewed to the left, which is a likely feature of the distribution of the level of future orientation in the population. The reported value for the discount factor then corresponds to the median, not the mean of its distribution, which is somewhat lower.

3.11 Appendix C: Model Fit

Figure 3.2: Basic Age-Behavior Patterns NYS Data and Baseline Simulation

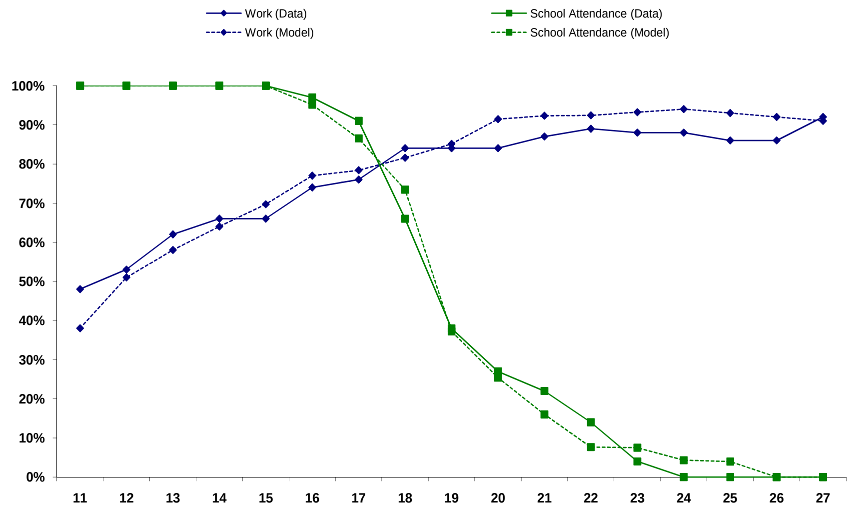


Figure 3.3: Basic Age-Crime Profile NYS Data and Baseline Simulation

3.12 Appendix D: Functional Forms

$$
\begin{aligned}
\gamma_L(S_a) &= \begin{cases} \overline{\gamma}_L & \text{if} \quad d_a^W = 0 \cap d_a^S = 0 \\ \frac{\overline{\gamma}_L}{2} & \text{if} \quad \left(d_a^W = 1 \cap d_a^S = 0\right) \cup \left(d_a^W = 0 \cap d_a^S = 1\right) \\ 0 & \text{if} \quad d_a^W = 1 \cap d_a^S = 1 \end{cases} \\
\overline{\gamma}_L &= \gamma_L + \gamma_{L1}(16 - age)\,1(age < 16) \\
\gamma_S(S_a) &= \gamma_{S_0} 1(age < MINAGESCHOOL) \\
+ & \quad \gamma_{S_0}\left(1 - \frac{1}{1 - \exp(19 - age)}\right) 1(age \geq MINAGESCHOOL) \\
& \quad + \gamma_{S_1} LagStigma_a + \gamma_{S_C} + \gamma_{S_{GS}} \\
\gamma_C(S_a) &= \gamma_{C_0} + \gamma_{C_1} age^2 + \gamma_{C_2} age + \gamma_{C_3} CK_a + \gamma_{C_4} 1\left(d^S = 1\right) \\
\gamma_P(S_a) &= \overline{\gamma}_P 1(FS_a = 1)
\end{aligned}
$$

where $LagStigma_a = age_a - A\_BEG - ATT_a$

3.13 Bibliography

[1] Ackerberg (2001) ”A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation”, mimeo UCLA

[2] Bayer, Pintoff and Pozen (2004), ”Building Criminal Capital behind Bars: Peer effects in Juvenile Corrections, mimeo

[3] Becker, Gary (1968) ”Crime and Punishment: An Economic Approach” The Journal of Political Economy. Vol. 76, No. 2. (March-April) pp. 169-217.

[4] Bentham, Jeremy (1789) ”Introduction to Principles of Morals and Legislation”

[5] Brame and Piquero (2003) ”Selective Attrition and the Age-Crime relationship”, Journal of Quantitative Criminology

[6] Brame and Paternoster (2003) ”Missing Data Problems in Criminological Research: Two Case Studies”, Journal of Quantitaive Criminology

[7] Burdett, Lagos and Wright (2003) ”Crime, Inequality and Unemployment” American Economic Review

[8] Burdett, Lagos and Wright (2004) ”An On-The-Job Search Model of Crime, Inequality and Unemployment” International Economic Review

[9] Bushway, Shawn (1996) ”The impact of a criminal history record on access to legitimate employment”, Ph.D. Dissertation, Carnegie Mellon, Heinze III School of Public Policy

[10] Bushway, Shawn (1998) ”The Impact of an Arrest on the Job Stability of Young White American Men” Journal of Research in Crime and Delinquency 34:454-479

[11] Bushway, Shawn (2004) ”Labor Market Effects of Permitting Employer Access to Criminal History Records” Journal of Contemporary Criminal Justice

[12] Davis, Michael (1988) ”Time and Punishment: An intertemporal Model of Crime”, Journal of Political Economy

[13] Eckstein and Wolpin (1999) ” Why Youths Drop Out from School: the impact of preferences, opportunities and ability” Econometrica

[14] Fagan, Jeff and Freeman, Richard (1999) ”Crime and Work” Crime and Justice: A review of Research 25: 225-290

[15] Finlay, Keith (2007) ”Employer Access to Criminal History Data and the Employment of Young Black Men”, mimeo UC Irvine

[16] Flinn, Christopher (1986) ”Dynamic Models of Criminal Careers” in Criminal Careers and Career Criminals Vol I, Blumstein, Cohen, Roth and Visher, Editors.

[18] Grogger (1995), The Effect of Arrests on the Employment and Earnings of Young Men, Quarterly Journal of Economics

[18] Grogger (1998), Market Wages and Youth Crime, Journal of Labor Economics

[19] Heckman and Singer (1984) ”A method for minimizing the impact of distributional assumptions in econometric models for duration data” Econometrica

[20] Hirschi, Travis (1986) ”On the compatibility of Rational Choice and Social Control Theories of Crime” in The Reasoning Criminal: Rational Choice Perspectives on Offending, Cornish and Clarke, Eds.

[21] Holzer, Raphael and Stoll (2001) ”Will Employers Hire ex-offenders? Employer Preferences, Background Checks and their Determinants, mimeo

[22] Holzer, Raphael and Stoll (2002a) ”How do crime and incarceration affect the employment prospects of less educated young black men”, mimeo

[23] Holzer, Raphael and Stoll (2002b) ”Can employers play a more positive role in prisoner re-entry?”, mimeo

[24] Holzer, Raphael and Stoll (2003a) ”Employment Barriers Facing Ex-Offenders”, mimeo

[25] Holzer, Raphael and Stoll (2003b) ”Employment Demand for Ex-Offenders” Recent evidence from LA, mimeo

[26] Holzer, Raphael and Stoll (2003c) ”The labor market for ex-offenders in Los Angeles: Problems, Challenges and Public Policy, mimeo

[27] Huang, Lain and Wang (2004) ” ”, International Economic Review

[28] Imai and Krishna (2004) ”Employment, Deterrence and Crime in a Dynamic Model”, , International Economic Review

[29] Imrohoroglu, Merlo and Ruppert (2004) ”What accounts for the decline in crime”, International Economic Review

[30] Keane and Wolpin (1994) ”The solution and estimation of discrete choice dynamic programming model by simulation and interpolation: monte carlo evidence”, Review of Economics and Statistics

[31] Keane and Wolpin (1997) ”The Careers Decisions of Young Men”, Journal of Political Economy

[32] Keane and Wolpin (2005) ”Exploring the Usefulness of a Non-Random Holdout Sample for Model Validation: Welfare Effects on Female Behavior” Working Paper

[33] Kling, Jeff (1999) ”The effects of Prison Sentence Length on the Subsequent Employment and earnings of Criminal Defendants”, mimeo

[34] Kling, Jeff (2004) ”Incarceration Length, Employment and Earnings”, Working Paper.

[37] Lochner (2004) ”Education, Work and Crime: A Human Capital Approach”, IER

[37] Lochner (2005) Individual Perceptions of the Criminal Justice System., NBER wp 9474

[37] Lochner and Moretti (2004) ”The Effect of Education on Crime: Evidence from Prison Inmates, Arrests and Self-Reports”, American Economic Review

[38] Lott (1990) ”The effect of conviction on the legitimate incomes of criminals” , Economic Letters

[39] Lott (1992) ”Do we punish high income criminals to heavily” , Economic Inquiry

[40] Manski, Charles (2004) ”Measuring Expectations” Econometrica.

[41] Manski, Charles (1999) ”Analysis of Choice Expectations in Incomplete Scenarios”, Journal of Risk and Uncertainty

[43] Nagin and Waldfogel (1993) ”The Effect of conviction on income over the life cycle” NBER WP 4551

[43] Nagin and Waldfogel (1995) ”The Effects of Criminality and Conviction on the Labor Market Status of Young British Offenders”, IRLE

[44] Needels, Karen (1996), ”Go to jail...”, Journal of Research in Crime and Punishment

[45] Pager, Devah (2002) ”The Mark of a Criminal Record”, Ph.D. Dissertation, University of Wisconsin-Madison

[46] Pager, Devah (2003) ”The Mark of a Criminal Record”, American Journal of Sociology

[47] Persico, Nicola (2002) Racial Profiling, Fairness, and Effectiveness of Policing.” The American Economic Review 92(5), December 2002, pp. 1472-97

[48] Pintoff, Randi (2006), Do Arrest and Incarceration Affect High School Graduation?, mimeo

[49] Reuter, MacMoun and Murphy (1990), Money from Crime, RAND

[50] Rust (1994) ”Structural Estimation of Markov Decision Processes”, in Handbook of Econometrics Vol 4, Chapter 51

[51] Rust and Phelan (1997) ”How Social Security and Medicare affect Social Security in a World of Incomplete Markets” Econometrica

[52] Tauchen, Witte and Griesinger (1994) ”Criminal Deterrence: Revisiting the Issue with a Birth Cohort”, Review of Economics and Statistics

[53] Van der Klauw (1996) ”Female Labour Supply and Marital Status Decisions” Review of Economic Studies

[54] Van der Klauw (2000) ” On the use of expectation data in Estimation of Dynamic Structural Models”, mimeo

[55] Waldfogel (1994) ”The Effect of Criminal Conviction on Income and the Trust Reposed in the Workmen”, Journal of Human Resources

[56] Williams and Sickles (2002) ”An analysis of the crime as work model: Evidence from the Philadelphia Birth Cohort Study” , Journal of Human Resources

[57] Williams and Sickles (2003) ”An intertemporal Model of Rational Criminal Choice” , mimeo, Rice University

[58] Williams and Sickles (2004) ”Turning From Crime: a Dynamic Perspective”, mimeo, Rice University

[59] Witte, Ann Dryden (1997) ”Crime”, Chapter 7 in The Social Benefits of Education, Behrman and Stacey, Eds. The University of Michigan Press, Ann Arbor.

[60] Wolpin (1984) ”An estimable Dynamic Stochastic Model of Fertility and Child Mortality”, Journal of Political Economy

[61] Wolpin (1987) ”Estimating a structural search model: the transition from school to work”, Econometrica

[62] Wolpin (1999) ”Commentary on ”Analysis of Choice Expectations in Incomplete Scenarios” ”, Journal of Risk and Uncertainty