The Transition from School to Jail: Youth Crime and High School Completion Among Black Males[*]

Antonio Merlo[†]   Kenneth I. Wolpin[‡]

January, 2009
Revised, October, 2009

ABSTRACT

In this paper, we study the relationship among schooling, youth employment and youth crime. The framework, a multinomial discrete choice vector autoregression, provides a comprehensive analysis of the dynamic interactions among a youth's schooling, work and crime decisions and arrest and incarceration outcomes. We allow for observable initial conditions, unobserved heterogeneity, measurement error and missing data. We use data from the NLSY97 on black male youths starting from age 14. The estimates indicate important roles both for heterogeneity in initial conditions and for stochastic events that arise during one's youth in determining outcomes as young adults.

[†]University of Pennsylvania, <merloa@econ.upenn.edu>
[‡]University of Pennsylvania, <wolpink@econ.upenn.edu>

I. Introduction

As of June 30, 2007, 9.7 percent of black males between the ages of 20 and 24 were inmates held in state or federal prisons or in local jails (U.S. Department of Justice (2008)). Even more striking is the fact that, in 1995, one-third of black men between the ages of 20 and 29 were either incarcerated, on probation or on parole (Mauer (1995)). While statistics of this kind are often contrasted to those of white males of the same ages, for whom incarceration rates are substantially smaller (e.g., Freeman (1992)), a perhaps more apt contrast is to the majority of black males who do not become involved in the criminal justice system.[1] What is different about the 10 percent of the 20-24 year old black males who were incarcerated as opposed to the 90 percent who were not, or the one-third of 20-29 year old black men who had been incarcerated, on probation or on parole at some point in their lives as opposed to the two-thirds who had not?

Table 1, based on data from the 1997 youth cohort of the National Longitudinal Surveys (NLSY97), shows the relationship between incarceration rates of young adult black males age 19 to 22 and a number of juvenile behaviors and family background characteristics.[2] In terms of prior behaviors, the largest difference in incarceration rates arises between black male youths who did not attend school at age 16, who account for 22.8 percent of the 16 year olds, and those who did. Non-attenders are 4 times more likely to be incarcerated at ages 19 to 22. Working at age 16 while attending school is associated with a lower incarceration rate at ages 19 to 22 and committing a crime at age 14 with a greater incarceration rate, but the differences are small and not precise.[3] With respect to background characteristics, a youth whose mother is a high school dropout is 88 percent more likely to be incarcerated at ages 19 to 22, a youth who was born to a teen mother is 78 percent more likely and a youth who has always lived with both biological parents is 34 percent less likely.

Table 1 also shows the relationship between those same juvenile behaviors and family background characteristics and several other young adult outcomes. Attending school at age

---

[1] As of June 30, 2007, only 1.6 percent of white males between the ages of 20 and 24 were incarcerated (U.S. Department of Justice (2008)).

[2] We describe the data in detail in Section V below.

[3] By precise, we mean that the 95 percent confidence interval shown in table 1 does not include zero.

16 is not only associated with lower incarceration rates, but the 16 year olds that attend school also have a lower crime rate at ages 19 to 22, by 12.2 percentage points, a lower arrest rate, by 15.5 percentage points, and a high school graduation rate that is about 12 times greater. Among the other variables, the largest differences are associated with having committed a crime at age 14. In particular, youths that committed a crime at age 14 are twice as likely to commit a crime at ages 19 to 22 as those that did not, are 10.3 percentage points more likely to be arrested and are 19 times less likely to graduate from high school. The relationship between background characteristics and both young adult crime and arrest rates is generally small and not precisely determined. However, having a mother who did not graduate from high school or who was a teen at the youth's birth is associated with lower youth graduation rates, by 28.4 and 12.5 percentage points.

The relationships among schooling, youth employment and youth crime, as depicted in table 1, have long been of interest to social scientists. In particular, empirical research in economics, psychology and sociology/criminology has focused attention on three main issues, identified with three distinct literatures. The first issue concerns the relationship between education and crime. More recent contributions by economists to this literature include Lochner and Moretti (2004), who estimate the effect of school attainment on participation in criminal activity, Jacob and Lefgren (2003), who estimate the effect of physically being in school on juvenile crime in a particular day, and Hjalmarsson (2008), who estimates the effects of arrest and incarceration on subsequent high school graduation.[4] Using aggregate (state-level) as well as individual-level data, Lochner and Moretti find that school attainment significantly reduces the probability of arrest and incarceration. They find that this effect is due to a reduction in criminal activity, rather than a direct effect on arrest and incarceration probabilities conditional on committing a crime. Jacob and Lefgren combine daily measures of criminal activity and school calendars in a number of local communities across the country, and find that while the level of youth property crime is significantly lower on days when school is in session than on days when it is not, the level of youth violent crime is higher. Using data from the NLSY97, Hjamlarsson finds that arrest and incarceration prior to age

---

[4]Other notable contributions include Lochner (2004), Tauchen et al. (1994), Williams and Sickels (2002) and Witte and Tauchen (1994).

16 significantly reduce the probability of graduating high school by age 19, although the effect of incarceration is relatively more robust than that of arrest.[5]

A second literature focuses on the relationship between working while attending school and delinquency. This issue has been studied primarily by psychologists and criminologists using a variety of data sets from several local and national studies of delinquency prevention. For example, using data from a longitudinal study of two birth cohorts born in Racine, Wisconsin, in 1942 and 1949, Shannon (1982) finds that employment while in school is weakly negatively related to future delinquency. Gottfredson (1985), using data from a national evaluation study that followed samples of youths attending schools located in depressed, predominantly minority, inner-city areas over a period of two years (1981-82), finds that working while in school does not increase delinquency in the following year. On the contrary, using data from the National Youth Survey, Ploeger (1997) finds a positive correlation between youth employment and some forms of delinquency.

The third issue concerns the persistence of youth criminal behavior or recidivism. The literature on this topic is vast, especially in criminology.[6] In a series of articles that are most directly relevant, Nagin and Paternoster (1991), Nagin and Land (1993), Nagin et al. (1995), and Broidy et al. (2003) identify different patterns of criminal behavior over time in several individual-level, longitudinal datasets. They relate these patterns to the presence of unobserved "offender types" among individuals, and assess the relative contributions of unobserved types and state dependence (prior criminal behavior) to explain the data. They find that engaging in criminal activity at an early age increases the propensity to do so in the future. Moreover, individuals differ in their underlying propensity to engage in criminal activity, and this unobserved trait is persistent over time. They find that the distribution of this unobserved trait is summarized well by a small number of discrete types.[7]

---

[5]There is another strand of the literature that focuses on wage effects, either the relationship between youth crime and wages, e.g., Freeman (1996), Grogger (1998) and Gould et al. (2002), or on the relationship between arrest and incarceration and wages, see, e.g., Grogger (1995), Kling (2006) and Waldfogel (1994). We do not review them here because the empirical approach we adopt does not address this issue.

[6]See, e.g., Blumstein et al. (1986), Godttfreddson and Hirschi (1990), and Wilson and Herrnstein (1985).

[7]The importance of accounting for unobserved heterogeneity has also been emphasized in the other two literatures mentioned above (see, e.g., Lochner and Moretti (2004)).

In this paper, we propose an approach to the empirical study of the relationships among schooling, youth employment and youth crime that encompasses the three literatures. Rather than investigating each relationship in isolation, our approach provides a comprehensive analysis of the dynamic interactions among the different choices youths make and their possible exposure to the criminal justice system. The unified empirical framework we consider takes the form of a multinomial discrete choice vector autoregression (VAR). In particular, we specify a dynamic system of equations for an individual's current schooling, work and crime decisions as well as arrest and incarceration outcomes as a function of past decisions and outcomes and unobserved types. We also allow for observable initial conditions, the possibility of measurement error in all the variables, missing data, and correlation among the contemporaneous shocks in all the equations.

In order to motivate the empirical specification and interpret the results, we outline a general dynamic optimization model of youth behavior. The model embeds multiple mechanisms through which unobserved heterogeneity and state dependence (past choices and outcomes) may operate. The model makes clear how the discrete choice VAR can be considered as an approximation to the decision rules of the model and how its parameters are related to the fundamental structure of the behavioral model. Understanding this mapping from the VAR parameters to the structure is critical to a correct interpretation of counterfactual experiments using the VAR estimates.

We estimate the VAR using data from the NLSY97 on black male youths from age 14 to the time they either graduate from high school or reach age 22 without having completed high school. We assess the within-sample fit of the statistical model as well as its out-of-sample performance in predicting the criminal behavior and the arrest and incarceration rates of high school graduates between the ages of 19 and 22. We then use the estimates to quantify the effect of initial conditions (unobserved heterogeneity and criminal activity at age 13) and the persistence of past decisions (state dependence) on early adult outcomes. To assess the importance of persistence, we simulate different hypothetical scenarios for the same individual and compare their outcomes. For example, for given initial conditions, we evaluate the consequences for outcomes at ages 19 to 22 of arresting or incarcerating an individual at age 14 who would have otherwise not being arrested or incarcerated. Borrowing from the

4

terminology of the macro VAR literature, this procedure is similar to calculating "impulse response functions" to unanticipated, temporary outcome shocks. In a similar manner, we also assess the impact of not attending school at age 16 and of working while attending school on early adult behaviors and outcomes.

The VAR estimates indicate important roles for heterogeneity in initial conditions and for state dependence. Among the categories of the initial conditions, we find that high school graduation rates range from as low as 15 percent to as high as 85 percent, crime rates at ages 19 to 22 range from 12 to 35 percent and incarceration rates from essentially zero to 26 percent. We find that committing a crime at age 14 due to a transitory, i.e., once and not for all, shock increases the propensity to commit a crime at ages 19-22 by between 7 and 11 percentage points depending on the youth's initial conditions. This effect, while considerable, is much smaller than the difference in crime rates, 18.7 percent, shown in table 1. Similarly, the effect of not attending school on the likelihood of incarceration, averaged over the estimated distribution of initial conditions, is 3.2 percentage points, while the difference in table 1 is 12.4 percentage points. Although the simple comparisons of means in table 1 greatly overstate the relationship between juvenile (age 14) and young adult (ages 19-22) criminal activity and between school attendance at age 16 and later incarceration, we find in other cases that the difference is small or even reversed. The effect of not attending school at age 16 on the crime rate and on high school graduation are larger as estimated from the VAR than in the raw comparison.

The remainder of the paper is organized as follows. In Section II, we present an illustrative dynamic model of youth crime, youth employment and high school completion which we use in Section III to discuss briefly the relative advantages and limitations of different approaches to the empirical study of youth crime. In Section IV, we describe the VAR specification, in Section V the data and in Section VI the empirical implementation. Our results are contained in Section VII. We conclude with Section VIII.

II. An Illustrative Dynamic Model of Youth Crime, Work and High School Completion

Consider a youth who is currently age 14. At each age $a \geq 14$, the youth potentially chooses among eight mutually exclusive discrete alternatives comprised of all com-

binations of attending school, $s_a = \{0, 1\}$, working in the labor market, $h_a = \{0, 1\}$, and committing crime, $c_a = \{0, 1\}$.[8] At any age, the youth has a history of previous choices $H_a^1 = (H_a^s, H_a^h, H_a^c)$, where $H_a^j = (j_{14}, j_{15}, ..., j_{a-1})$ for $j = s, h, c$. The set of alternatives available to the youth is restricted by whether the youth is incarcerated at age a, in which case the youth can make no choice in that period. We denote arrest at age a by $A_a = \{0, 1\}$ and incarceration at age a by $J_a = \{0, 1\}$. As with the youth's choices, the youth has a history of criminal justice outcomes given by $H_a^2 = (H_a^A, H_a^J)$, where $H_a^j = (j_{14}, j_{15}, ..., j_{a-1})$ for $j = A, J$.

Preferences:

The youth has a per-period utility function that depends on consumption, X_a, on the previously specified choices, s_a, h_a, c_a, on the youth's current incarceration status, J_a, possibly on the histories of choices, H_a^1, as well as the history of the youth's involvement with the criminal justice system, H_a^2, on the youth's age, a, on a set of choice-specific immutable preference endowments as of age 14, $\mu_{14}^1 = (\mu_{14}^s, \mu_{14}^h, \mu_{14}^c)$, and on a set of choice-specific preference shocks at age a, $\epsilon_a = (\epsilon_a^s, \epsilon_a^h, \epsilon_a^c)$. The history of shocks is denoted by $H_a^3 = (H_a^{\epsilon^s}, H_a^{\epsilon^h}, H_a^{\epsilon^c})$, where $H_a^{\epsilon^j} = (\epsilon_{14}^j, \epsilon_{15}^j, ..., \epsilon_{a-1}^j)$ for $j = s, h, c$. The joint distribution of ϵ_a is denoted by $F(\epsilon_a | H_a^3)$, and as shown may depend on the history of shocks. To summarize, the utility function is given by

$$U_a = U(X_a, s_a, h_a, c_a; H_a^1, H_a^2, J_a, a, \mu_{14}^1, \epsilon_a), \tag{1}$$

where

$$\epsilon_a \sim F_\epsilon(\epsilon_a | H_a^3). \tag{2}$$

Budget Constraint:

For simplicity, the youth is assumed to consume his income each period, that is, the youth neither saves nor borrows. The youth (who is not incarcerated) has three sources of income: (i) from parents, w_a^p; (ii) from working in the legitimate sector and earning a wage, w_a^h; and (iii) from working in the illegitimate sector and earning an implicit wage of w_a^c. The transfer received from parents and the wage earned in each sector depend on histories

---

[8] As in the empirical work, the model abstracts from the intensive margin of these decisions. In addition, we do not differentiate between heterogeneous types of crimes.

and the youth's age. They also depend on immutable endowments that reflect parental characteristics (earnings potential and preferences) and on legitimate- and illegitimate-sector skill endowments of the youth at age 14. These are denoted by $\mu_{14}^2 = (\mu_{14}^{w^p}, \mu_{14}^{w^h}, \mu_{14}^{w^c})$. In addition, parental contributions and wages are subject to shocks, denoted by $\eta_a = (\eta_a^p, \eta_a^h, \eta_a^c)$. The history of shocks is denoted by $H_a^4 = (H_a^{\eta^p}, H_a^{\eta^h}, H_a^{\eta^c})$, where $H_a^{\eta^j} = (\eta_{14}^j, \eta_{15}^j, ..., \eta_{a-1}^j)$ for $j = p, h, c$. The joint distribution of η_a, which may be conditional on the history of shocks, is denoted by $F(\eta_a | H_a^4)$. The budget constraint is thus

$$X_a = w_a^p + w_a^h h_a + w_a^c c_a, \tag{3}$$

where

$$w_a^p = w^p(H_a^1, H_a^2, H_a^3, H_a^4, a, \eta_a^p, \mu_{14}^1, \mu_{14}^2), \tag{4}$$

$$w_a^h = w^h(H_a^1, H_a^2, a, \eta_a^h, \mu_{14}^{w^h}), \tag{5}$$

$$w_a^c = w^h(H_a^1, H_a^2, a, \eta_a^c, \mu_{14}^{w^c}), \tag{6}$$

$$\eta_a \sim F_\eta(\eta_a | H_a^4). \tag{7}$$

The parental contribution to the youth's consumption, a decision of the parents, is assumed to depend on the youth's histories of prior choices, criminal justice involvement, preference shocks and shocks to prior income by source, and on the youth's skill endowments. The parental transfer function does not depend on contemporaneous shocks specific to the youth $(\epsilon_a, \eta_a^h, \eta_a^c)$. Thus, the parents base their transfer on everything that is known up to the current realizations of youth-specific shocks; this allows the parent to use a forecast of future behaviors and potential earnings in deciding on how much youth consumption to provide. The parents are, in this sense, modeled as a Stackleberg leader, although the actual decision process is not explicitly considered. The wages in the legitimate and illegitimate sectors depend on the histories of choices and criminal justice involvement and on the sector-specific shocks and endowments. Thus, schooling may affect productivity in both sectors as might work experience in each sector. Skill shocks and endowments are, however, sector-specific.

A youth who is incarcerated at age a, as noted above, makes no choices. The youth's consumption is fixed at an amount X^J during that period.

Criminal Justice System Constraints:

A youth faces a probability of arrest and a probability of incarceration given by

$$\pi_A(A_a | H_a^1, H_a^2, a, \mu_{14}^{wc}, L), \tag{8}$$

$$\pi_J(J_a | H_a^1, H_a^2, a, \mu_{14}^{wc}, L). \tag{9}$$

The probability that a youth is arrested at age a depends on the youth's histories (of criminal activity, schooling, employment and involvement with the criminal justice system), on the youth's age and endowment of illegitimate-sector skill, and on the youth's location (assumed fixed over time), L, reflecting area-specific policing and judicial policies and resources. The probability of being incarcerated is also conditioned on the same histories of the youth, age, the illegitimate-sector skill endowment and location.

Horizon:

To restrict attention to youth crime and high school completion, the decision model is specified only up to the last age an individual is permitted to attend high school, $\bar{a}$, or the age at which high school is completed, whichever comes first. If the individual has passed the high school terminal age ($a = \bar{a} + 1$) without having graduated, the subsequent choice set is restricted to (h_a, c_a) from that age on. Subsequent choices, for example, college attendance, are not explicitly modeled for individuals that graduate before having passed the terminal age. The last decision period that is modeled arises either when a youth has accumulated 11 years of schooling, is age $\bar{a} - 1$ or less, and decides at a to attend school, or when an individual reaches age $\bar{a}$ with 11 years of schooling or less (regardless of the decision at a).

Objective Function:

The youth is assumed to maximize the present discounted value of lifetime utility. Because the model is truncated at either $\bar{a}$ or when the youth has completed high school, to complete the specification of the model, it is necessary to specify terminal value functions. These terminal value functions represent the expected discounted value of remaining lifetime utility when either of these conditions arises. Specifically, the terminal value function if the individual graduates at age a might be modeled as, for example, $V^G(H_a^1, H_a^2, H_a^3, H_a^4, a, \mu_{14}^1, \mu_{14}^2)$. The terminal function if the individual reaches age $\bar{a} + 1$ without having graduated, also a function of histories and endowments, $V^D(H_a^1, H_a^2, H_a^3, H_a^4, \mu_{14}^1, \mu_{14}^2)$, is the solution to the same decision problem as prior to $\bar{a} + 1$, absent the choice of attending school.

<u>Decision Rules:</u>

The maximization problem can be cast as a dynamic programming problem and solved by backwards recursion (see Eckstein and Wolpin (1999) for a similarly structured optimization problem). The solution of the model yields a set of decision rules for the choice among the eight mutually exclusive alternatives (for a non-incarcerated youth) that determine the outcomes at each age. Letting $d_a^k = \{0, 1\}$ for the $k = 1, .., 8$ alternatives and $a = 14, ..., \bar{a}$, the solution may be written as

$$d_a^k = \begin{cases} 1 & \text{if } G^k(H_a^1, H_a^2, H_a^3, H_a^4, a, \mu_{14}^1, \mu_{14}^2, L, \epsilon_a, \eta_a) = 0 \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

where the G^k functions are differences in the value function for alternative k and the maximum of the value functions over all of the 8 alternatives (see Keane and Wolpin (2001)). Notice that when alternative k is chosen, $G^k = 0$ and $G^{k'} < 0$ for all $k' \neq k$. Together with the parental transfer function and the wage functions (4), (5) and (6), and the arrest and incarceration probability functions (8) and (9), these 13 functions define all of the outcomes of the model.

III. Empirical Approaches

The estimation problem arises because the researcher does not observe the shocks or the endowments. There are several possible approaches to the empirical implementation of the kind of dynamic model presented above.

<u>Structural Approach:</u>

The structural approach proceeds by specifying functional forms for the primitives of the model: the utility function, the parental transfer and youth wage functions, the arrest and incarceration probability functions, the terminal graduation value function and the probability distributions for the preference, parental transfer and wage shocks and for the endowments, $F_\mu(\mu_{14}^1, \mu_{14}^2)$. Structural estimation of the model recovers the parameters of the primitives (assuming that they are identified). Given the parameters of the primitives, the parameters of the G^k functions, which are composites of the parameters of the primitives, are also recovered. The advantage of the structural approach is that it enables the researcher to perform a wide range of counterfactual policy experiments, including those that are outside

9

of the sample experience.[9]

Semi-Structural Approach:

The semi-structural approach recovers a subset of the primitives. It does not necessarily require that all of the functional forms and probability distributions be specified. An example would be the estimation of the illegitimate sector wage function. Given exclusion restrictions, that is, the existence of components of the histories that affect the decision to participate in the illegitimate sector but that do not affect the returns in that sector, the illegitimate sector wage function can be semi-parametrically estimated. Estimation of this kind would be used to determine, for example, the effect of education on productivity in the illegitimate sector.[10]

Non-Structural Approach:

The non-structural approach eschews the estimation of the primitive parameters. The goal of the non-structural approach would be to estimate the parameters of the G^k functions. This is the approach followed by the papers we reviewed in the introduction (which essentially estimate different G^k functions in isolation). It is also the approach we follow here. We discuss in detail the methodology that we employ and what is learned from that approach below.

IV. A Discrete Outcome VAR Approximation

Different specifications of the primitive functions will lead to different G^k functions. Rather than specify the primitive functions, we adopt a first-order approximation of the G^k functions in the form of a multinomial discrete choice vector autoregression. To develop the estimation procedure, suppose that the G^k functions are additively separable in histories and contemporaneous shocks and that the shocks are mutually serially uncorrelated (so that, conditional on the histories of observables and permanent unobserved heterogeneity, the histories of shocks do not affect decisions), that is,

$$G^k = \overline{G^k}(H_a^1, H_a^2, a, \mu_{14}^1, \mu_{14}^2) + u_a^k \text{ for } k = 1, ..., 8, \ 14 \leq a \leq \overline{a}. \tag{11}$$

[9]For example, Fella and Gallipoli (2006), Imai and Krishna (2004) and Sickles and Williams (2008) adopt the structural approach.

[10]Examples of articles that follow the semi-structural approach are Grogger (1998) and Williams and Sickles (2002).

The u_a^k's are composites of all of the underlying contemporaneous shocks, ϵ_a and η_a. As noted, an equivalent representation of the decision rule is that the k$^{\text{th}}$ alternative is chosen if and only if $G^k > G^{k'}$ for all $k' \neq k$. As in any multinomial discrete choice problem, one of the G^k functions must be normalized to zero.

The representation given by (11) is obviously not parsimonious. An assumption of joint normality of the u_a^k's would result in a 28-parameter variance-covariance matrix and adding arrest and incarceration as outcomes would increase that total to 55 parameters. Making an extreme value assumption would conserve on covariance parameters, but would still yield a parameter space equal to the number of parameters in each $\overline{G^k}$ function times nine (including the arrest and incarceration probability functions). To reduce the number of parameters in the VAR representation that we estimate, we represent the decision rules by only three, rather than seven, equations.[11]

The VAR in the three discrete choices thus takes the following form, where we account for unobserved heterogeneity, inclusive of location-specific factors (L), by assuming that there are a finite number of types, which we denote by $\tau = m$ for $m = 1, ..., M$:

$$s_a = \begin{cases} 1 & \text{if } Q_a^s \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

$$h_a = \begin{cases} 1 & \text{if } Q_a^h \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

$$c_a = \begin{cases} 1 & \text{if } Q_a^c \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where, for $j = s, h, c$,

$$Q_a^j = \sum_{m=1}^M \mu_{14,m}^j I\left(\tau = m\right) + \alpha_1^j s_{a-1} + \alpha_2^j h_{a-1} + \alpha_3^j c_{a-1} + \alpha_4^j A_{a-1} + \alpha_5^j J_{a-1} + g_j(a) + u_a^j, \tag{15}$$

---

[11]In a static discrete choice framework, a sufficient set of restrictions to yield this representation of the decision rules is that the utility function $U(s_a, h_{a,} c_a, A_a, J_a)$ be additively separable in the choices. In a dynamic setting, these restrictions are not sufficient and this representation is only an approximation to the full set of decision rules.

$I(\tau = m)$ is an indicator variable equal to one if the youth is type m and zero otherwise and $g_j(a)$ is a flexible function of age.

We also incorporate arrest and incarceration probability functions, which, given the behavioral model, are taken to represent approximations to the structural arrest and incarceration probability functions:

$$A_a = \begin{cases} 1 & \text{if } Q_a^A \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

$$J_a = \begin{cases} 1 & \text{if } Q_a^J \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

where Q_a^A and Q_a^J are analogous to the Q_a^j's defined in (15). The five shocks, $(u_a^s, u_a^h, u_a^c, u_a^A, u_a^J)$ $= u_a$, are assumed to be jointly serially uncorrelated, and contemporaneously joint normal with variance-covariance matrix Λ. Each type of youth (m) is defined by a 5-vector endowment at age 14, $\left(\mu_{14,m}^s, \mu_{14,m}^h, \mu_{14,m}^c, \mu_{14,m}^A, \mu_{14,m}^J\right)$. The mth type comprises π_m percent of the population $(\Sigma_{m=1}^M \pi_m = 1)$.

V. Data

The data are from the NLSY97. The NLSY97 consists of approximately 9000 youths age 12 to 16, as of December 31, 1996, who were first interviewed in 1997 and re-interviewed annually since then. We make use of the first eight rounds of the survey. The NLSY97 contains an event history of schooling and employment, as well as detailed information in each survey round about the extent of criminal activity and exposure to the criminal justice system since the previous interview. While the crime related data are extensive, it is not possible to create an event history, that is, to date each crime that a youth committed or to determine whether the youth was apprehended and/or incarcerated for a specific time-dated crime (or set of crimes).

The time period for each annual observation runs from October 1 to September 30. A youth's age (in years) is the age as of October 1 of a given year. The dichotomous variables, attending school at age a, s_a, and being employed at age a, h_a, are obtained from the event history data. The schooling data were hand-edited in order to obtain a consistent profile of attendance and grade completion using information on attendance as of October 1, January 1 and April 1 of each school year, grade attending as of October 1, highest grade completed

as of each interview date and the year of high school graduation (excluding GED's). A youth was considered attending school during the (school) year if we determined that the youth had completed a year of schooling during that year and not attending otherwise. Weekly hours worked, taken from the event history data, were summed over the weeks between October 1 of a given year and September 30 of the following year to obtain the youth's annual hours worked. A youth was considered to be working during the year if their hours worked were at least 780 over the year.[12]

As noted, the crime data is not collected as an event history. Instead, at each interview, the youth is asked about the number of times, if at all, since the last interview that the youth engaged in a series of different criminal activities: stealing something worth less than 50 dollars, stealing something worth 50 dollars or more, other property crimes (e.g., fencing stolen property), selling drugs and assaults. The number of crimes of each type that a youth committed since the date of the last interview was divided by the number of months since the date of the last interview and the resulting figure was distributed uniformly over the months. Summing over the period between October 1 and September 30 gave the number of crimes of each type that the youth committed over the period. The dichotomous variable of whether or not the youth committed a crime at age a, c_a, was determined by whether the preceding sum was positive for any crime.[13]

The arrest data were obtained at each survey round as dated events by month and year.[14] At each survey round, the dates were collected for up to nine separate arrests since

---

[12]This definition of employment essentially excludes youths who are employed only during the Summer from being categorized as working.

[13]By itself, this procedure would potentially lead to an overstatement of the number of periods in which crimes were committed and also introduce spurious correlation over time in the commission of crimes. To avoid this problem, we hand-edited the data in cases where that issue was likely to arise. In defining the dichotomous crime variable, we did not include among the types of crimes having damaged property, which seemed to us generally less significant. About 10 percent of youths in the 1997 survey round who had reported having committed no other crime, reported having engaged in acts of vandalism.

[14]As discussed in Lochner (2007), the arrest data in the NLSY97 are generally consistent with official statistics.

the date of the last interview.[15] Beginning and ending dates (month and year) of each period of incarceration were collected as well. The dichotomous arrest variable, A_a, was defined according to whether the youth of age a had been arrested between October 1 and September 30. Similarly, the dichotomous variable J_a was determined by whether a youth of age a was incarcerated at any time over the October 1-September 30 period.[16]

As might be expected, there is a significant amount of missing data. Missing data is particularly serious in the statistical specification we have adopted. A missing value for any of the variables requires that potentially up to two observations be dropped (at ages a and $a-1$). In particular, we would lose about 20 percent of the observations for each of the dependent variables in the VAR.[17] Further, both because of the sensitive nature of the data as well as the way we processed the data, it is not unreasonable to assume that the dichotomous variables are subject to classification error. Our estimation procedure deals with both of these complications.

VI. Empirical Implementation

To accommodate missing data and measurement error, the 5-equation discrete outcome VAR is estimated using a procedure developed in Keane and Wolpin (2001) and extended in Keane and Sauer (2005).[18] The initial conditions are the youth's type and whether or not the youth committed a crime at age 13, c_{13}. The other initial (i.e., age 13) conditions are assumed not to vary; in particular, it is assumed that $s_{13} = 1$, $h_{13} = 0$, $A_{13} = 0$ and

---

[15]Reported arrests were followed up with questions about whether the youth was charged and convicted for specific types of crime. It is not possible to match, however, those charges and convictions with crimes committed at any prior time because the youth was not asked about the correspondence with previously reported crimes.

[16]The persistence of incarceration would be overstated with this definition, because a youth who had a single spell of incarceration of less than 12 months that overlapped two October 1-September 30 12-month periods would be considered incarcerated in both periods. We hand-edited the data in those cases so that such a youth would only be incarcerated for one period.

[17]Missing observations also arise because, as noted, some youths are observed beginning at age 15 and others at age 16.

[18]In a Monte Carlo exercise with panel data probit specifications similar to the one we estimate, Keane and Sauer find that the estimation procedure performs well.

$J_{13} = 0$.[19] The estimation method is based on the simulation of complete outcome histories from age 14 to age $\bar{a}$ for a set of artificial agents of each unobserved type m and value of c_{13}. A history of outcomes consists of the simulated values of the youth's choices s_a, h_a and c_a and of the youth's involvement in the criminal justice system, A_a and J_a.

For a given set of parameters ($\alpha's$, $\mu's$ and Λ), simulated outcome histories, $\widetilde{O}_a(m, c_{13}) = (s_a(m, c_{13}), h_a(m, c_{13}), c_a(m, c_{13}), A_a(m, c_{13}), J_a(m, c_{13}))$, are obtained for each type m and value of c_{13} by drawing a set of u_{14} values from the joint distribution, solving for $\widetilde{O}_{14}(m, c_{13})$ using equations (12)-(17), drawing a set of values of u_{15}, solving for $\widetilde{O}_{15}(m, c_{13})$ and so on up to the age $\bar{a}$, which we take to be 22. We do this for $n = 1, .., N$ artificial agents, thus yielding the set of simulated histories $\widetilde{O}^n(m, c_{13}) = (\widetilde{O}_{14}^n(m, c_{13}), \widetilde{O}_{15}^n(m, c_{13}), ..., \widetilde{O}_{22}^n(m, c_{13}))$. If we knew a youth's type and value of c_{13}, we could obtain a frequency simulator of any observed outcome vector for a youth (to the lesser of age 22, the age the youth graduates from high school or the age the youth attrites from the sample). Letting O^i be the observed outcome vector for youth i, an unbiased estimator for $\Pr(O^i)$ is simply the fraction of the N simulated histories that are consistent with O^i. Missing observations on any elements of O^i would be treated as being consistent with any entry in the corresponding element of $\widetilde{O}^n(m, c_{13})$. Note that the simulated probabilities rely only on unconditional simulations, that is, on the simulation of entire outcome histories. Thus, to calculate the probability of any observed outcome vector, we do not need to observe lagged outcomes.

Of course, this procedure is impractical as any observed history would be unlikely to be replicated in the simulated histories often enough to get a reasonably precise estimate of $\Pr(O^i)$. Indeed, unless N were huge, most observed outcome histories would have a simulated frequency of zero. The procedure we adopt solves that problem by assuming that all observed outcomes are measured with error, that is, that the discrete outcomes are subject to classification error. With classification error, as seems apt given the data, any

---

[19]In the data, about 4 percent of black youths age 13 report having been arrested (at that age) and 0.6 percent report being incarcerated. In contrast, 24 percent of black youths report having committed a crime at age 13. Although we could have incorporated arrest at age 13 as an initial condition, it would have complicated the estimation without, in our judgment, changing the results in any substantively meaningful way.

observed outcome history is consistent with each of the N simulated histories. If we denote $\Pr(O^i|\widetilde{O}^n(m, c_{13}))$ as the probability that an observed outcome history is generated by the nth simulated outcome history, then using N simulated histories, an unbiased simulator of $\Pr(O^i)$ is

$$\widehat{P}_N(O^i|m, c_{13}) = \frac{1}{N}\sum_{n=1}^{N}\Pr(O^i|\widetilde{O}^n(m, c_{13})). \tag{18}$$

If classification errors are independent across separate outcomes and ages, $\Pr(O^i|\widetilde{O}^n(m, c_{13}))$ is the product of the classification error rates (over the separate outcomes and at each age) that are needed to make O^i consistent with $\widetilde{O}^n(m, c_{13})$. If the youth's type (m) is not observed and c_{13} is observed (without error), then the ith youth's likelihood contribution is $\Sigma_m\widehat{P}_N(O^i|m, c_{13})\pi_{m|c_{13}}$, where $\pi_{m|c_{13}}$ is the probability that a youth is type m given the value of c_{13}. Consistency requires that c_{13} be exogenous conditional on type, that is, that the stochastic component of the process generating c_{13}, say u_{13}, be independent of u_a for $a \geq 14$.[20] If c_{13} is also assumed to be potentially misclassified, then it is necessary to account for the classification error in the estimation (as we do below).

Following Keane and Wolpin (2001), we assume that the classification error is unbiased (and independent across outcomes and ages). To illustrate, consider one of the outcomes at an arbitrary age, say attending school, s_a. Denote s_a^T as the true value and s_a^R as the reported value. Further, let $q_{i|j} = \Pr(s_a^R = i|s_a^T = j)$, $q_i^T = \Pr(s_a^T = i)$ and $q_i^R = \Pr(s_a^R = i)$, $i, j \in \{0, 1\}$. Then unbiasedness requires that $q_{1|1}q_1^T + q_{1|0}q_0^T = q_1^R = q_1^T$, which implies that $\frac{q_{1|0}}{q_{0|1}} = \frac{q_1^T}{1-q_1^T}$. Note that the right hand side of this expression is only a function of the true probability that $s_a = 1$. Given this expression, to satisfy the unbiasedness assumption, it is necessary that the classification error probabilities be linear in the true probability, namely that $q_{1|1} = \theta_s + (1 - \theta_s)q_1^T$, where θ_s, the classification error parameter, is estimable $(0 \leq \theta_s \leq 1)$.[21] Notice that as the probability that $s_a = 1$ approaches one, the classification error rate approaches zero $(q_{1|1} \to 1)$.

If c_{13} is measured with error, which is the assumption we adopt, then we need to weight

---

[20]If they were not independent, an alternative method would be required to correctly account for the stochastic initial condition, e.g., Heckman (1981).

[21]Identification of the classification error rates and VAR parameters is dependent on the normality assumption for u_a (see Hausman et. al., 1998).

$\widehat{P}_N(O^i|m, c_{13})$ for each type and value of c_{13} by $\Pr(\tau = m, c_{13}^T = i|c_{13}^R) = \pi_{m|c_{13}^T} \Pr(c_{13}^T = i|c_{13}^R)$. Using Bayes' rule, $\Pr(c_{13}^T = i|c_{13}^R) = \Pr(c_{13}^R|c_{13}^T = i)\frac{\Pr(c_{13}^T=i)}{\Pr(c_{13}^R)}$. The first term in this product is given by the classification error and the second, given the unbiasedness assumption, is either one if $c_{13}^R = c_{13}^T = i$ or $\frac{\Pr(c_{13}^T=i)}{\Pr(c_{13}^R=i')}$ if $c_{13}^R \neq i$. If c_{13} is unobserved, then, for those observations, we integrate it out using an estimate of $\Pr(c_{13}^T = i)$ from the proportion of youths with $c_{13}^R = i$ (which is equal to the proportion with $c_{13}^T = 1$ given unbiasedness).

VII. Results

<u>Parameter Estimates:</u>

The parameter estimates and standard errors for the VAR given by equations (12)-(17) are reported in appendix table A.1.[22] The estimation sample consists of 1,163 black males accounting for 5,532 person-period observations. There are 757 youths observed beginning at age 14, 239 beginning at age 15 and 167 beginning at age 16. Age effects are modeled as a linear spline with knots at ages 18 and 21. The values of $\widehat{P}_N(O^i|m, c_{13}^T)$ are based on $N = 10,800$ simulated paths for each combination of m and c_{13}^T.

We report the estimates for the model with three types ($M = 3$). However, we also estimated the VAR for one, two and four types as well as for a specification in which the joint distribution over the five $\mu's$ was normal.[23] There is no appropriate formal statistical test on which to base the choice of M. We therefore compared the fit of the alternative specifications using several measures. According to the BIC, $M = 3$ (74 parameters) was best, while according to the AIC, $M = 4$ (81 parameters) was best. We also compared the out-of-sample predictive accuracy of the different specifications based on the outcomes of youths at ages 19-22 who graduated from high school. There was, however, no discernible difference for $M = 2, 3$ or 4, although the specification with $M = 1$ and with the normality assumption were noticeably worse. Interestingly, the value of the BIC criterion was greatest for the normal distribution (80 parameters), while the value of the AIC criterion was greatest for $M = 1$ (60 parameters). Although as noted not all criteria favored $M = 3$, that specification

---

[22]There are 74 parameters. Parameters are generally precisely estimated. We also estimated a VAR with two lags. A likelihood ratio test did not reject the one-lag representation.

[23]In the case of the normal distribution, the mean was allowed to differ by whether or not the youth committed a crime at age 13.

dominated the one with $M = 2$ based on both BIC and AIC, while the specification with $M = 4$ did not dominate either $M = 3$ or $M = 2$ on both. Details are provided in appendix table A.2.[24]

The parameter estimates reflect the underlying structure of the behavioral model, at least as an approximation, but are not directly interpretable. The VAR parameters, the $\alpha's$, the $\mu's$ and those of the age splines, are composites of all of the structural parameters, including parameters of the utility function, the parental transfer function, the legitimate and illegitimate sector wage functions, the arrest and incarceration probability functions and the variance-covariance structure of the shocks. The contemporaneous error correlations (Λ) reflect, in the context of the behavioral model, correlations among composites of the underlying preference, wage, and other shocks. However, given the 12-month time period of the observations, it is likely that these correlations also reflect other components of the structure. For example, as seen in appendix table A.1, the correlation between the shocks to the crime and arrest equations is .512, which likely captures, in part, the structural arrest probability function (that is, that a youth that committed a crime at age a was arrested at age a). Similarly, the contemporaneous correlation between the shocks to attending school and being incarcerated, -.275, likely represents, in part, an incapacitation effect from being incarcerated for at least a part of the year.

Descriptive Statistics and Fit:

In the estimation sample, youths are followed either until they graduate from high school or until age 22, whichever comes first. Observations of youths from the age that they graduate from high school until they reach age 22 are out-of-sample observations (not used in estimation). Table 2 shows the outcome proportions in the estimation sample by age and the corresponding fit of the VAR.[25] As seen in table 2, school attendance rates decline

---

[24]The four independent values of the type probability function, $\pi_{m|c_{13}^T}$, $m = 1, 2, 3$, $c_{13}^T = 0, 1$, are not restricted in estimation. A multinomial logit specification is used to constrain the probabilities to lie between zero and one.

[25]The predicted outcome paths are based on simulations of 10,800 youths for each of the six m and c_{13}^T combinations. Each category is weighted by $\Pr(m, c_{13}^T) = \Pr(m|c_{13}^T) \cdot \Pr(c_{13}^T)$. The first term is $\pi_{m|c_{13}^T}$ and is estimated along with the VAR parameters. The second term, given our assumption of unbiasedness in the classification error, is the proportion of youths in the data reporting having committed a crime (or not) at

from 94 percent at age 14 to 66 percent at age 17, reflecting the large high school dropout rate among black males. The 25 percentage point drop between ages 17 and 18 and the further drop by 30 percentage points between ages 18 and 19 reflect the sample restriction to non-graduates in addition to dropouts.[26] There are three ways in which these large one-year declines are captured in the estimated VAR, through direct age effects, through lagged dependencies (not attending school in a period reduces the probability of attending the next period) that reinforce the age effect and through the compositional change in the types that are represented among those who have not graduated from high school by those ages. Although the first two together account for much of the decline, the compositional change accounts also for a non-trivial part of the decline, 25 percent of the drop between ages 17 and 18 and for 31 percent of the drop between ages 18 and 19. Overall, as seen in the table, deviations between the predicted and actual school attendance rates are small over the entire age range.[27]

The crime rate is essentially constant over the 14 to 17 age range, when the sample is relatively more homogeneous. On average, about 27 percent of black males commit at least one crime at (each of) those ages. Although the crime rate drops at age 18 for the sample of non-graduates, it jumps back up to previous levels at age 19 and (except for age 22) remains there.[28] Between ages 17 and 19, the compostional change accounts for 64 percent of the (predicted) increase in crime. As the table shows, the actual and predicted overall rates at ages 14 to 18 and 19 to 22 are generally close. Arrest rates increase slowly from age 14 to age 17, starting from .098 and increasing to .129. There is a large jump at age 18, to .205, and a further jump at age 19, to .233. As seen, the VAR estimates overstate

age 13. Those youths for whom age 13 crime is missing, primarily those who were ages 15 and 16 at the time of the first survey, are assumed to have the same distribution of age 13 crime as those reporting on c_{13}.

[26] Of those who graduate from high school, 65 percent graduate by age 18 and 95 percent by age 19.

[27] In addition to predicting well the outcome proportions, the VAR also predicts high school graduation rates (cuumulative secondary school attendance) well. The actual and predicted high school graduation rates are 60.5 vs. 62.1 percent.

[28] The drop at age 22 may reflect sample variation given that there are only 53 observations of non-high school graduates at age 22. There is a slight downward trend in the crime rate for the group that will never graduate from high school; the crime rate is 36 percent at age 14 and 29 percent at age 21.

19

the rise in arrest rates to age 17 and understate the jump at age 18, although the level at age 18 is almost identical in the actual data and the model prediction. The compositional change between ages 17 and 18 accounts for 57 percent of the (predicted) jump in the arrest rate and for more than 100 percent of the increase between ages 18 and 19. Incarceration rates also increase with age, with large jumps at ages 18 and 19. The compositional change accounts for 45 percent of the increased incarceration rate between ages 17 and 18 and for 65 percent between the ages of 18 and 19. The trend in the incarceration rate is fit well; the actual and predicted incarceration rates at ages 14 to 18 are .038 and .035 and at ages 19 to 22, .130 and .121.[29] Similarly, the rise in employment to age 17 is captured well, but the employment rate between ages 19 and 22 is overstated, .463 vs. .417. None of the increase in employment rates between the ages of 17 and 19 is due to compositional changes in the sample.

Table 3 assesses the ability of the VAR to fit the crime-related data for the out-of-sample observations of high school graduates at ages 19 to 22.[30] The first column shows the actual outcome proportions for the out-of-sample observations. Comparing tables 2 and 3, crime rates are about half of the level for high school graduates at those ages than for non-high school graduates (.154 vs. .284), arrest rates less than one-third as large (.072 vs. .253) and incarceration rates about one-sixth as large (.019 vs. .130). Clearly, those youths who

---

[29]Outcome proportions over age groups, 14 to 18 and 19 to 22, are the total number of occurrences over the ages in the range divided by the number of observations over those ages. They are thus the weighted averages of the outcome proportions at each age, where the weights are the proportions of observations at each age. Although the number of observations at ages 14 to 18 are fairly similar, within the 19 to 22 age range, ignoring outcome-specific missing values, 45 percent of the observations are at age 19, 30 percent at age 20, 18 percent at age 21 and only 6 percent at age 22.

[30]Given our focus on crime and on high school completion, we do not report the out-of-sample fit at ages 19 to 22 to college attendance or employment. Indeed, we did not extract the college attendance data from the NLSY97, which, given the ages spanned by the cohort, is still incomplete for many observations. By construction, the VAR representation cannot predict the large fraction of youths with exactly 12 years of schooling, and thus will necessarily overstate the fraction of youths who enter college. It appears that this overstatement of college entry causes the model to understate the fraction of high school graduates at those ages who work. The estimated VAR predicts an employment rate of .55 when the actual rate is .63.

graduated from high school, and who are then dropped from the sample, have different outcomes at ages 19-22 than those non-graduates maintained in the sample.

The VAR accounts for those differences because of differences in initial conditions, unobserved types and crime at age 13, differences that tend to be magnified by the autoregressive structure.[31] Youths who graduate from high school have different initial conditions that lead them to attend school and commit fewer crimes at all ages. Path dependencies in those outcomes further augment these differences between graduates and non-graduates. Allowing for unobserved heterogeneity is essential for the out-of-sample predictive accuracy of the VAR specification. As the initial sample ages, the composition of the estimation sample changes dramatically in terms of the type distribution. For example, the proportion of type 3 youths is estimated to be 52 percent at age 17, drops to 40 percent at age 18 (reflecting the elimination of those who graduated from high school by age 18) and to 25 percent at age 19 (reflecting the elimination of those who graduated by age 19).[32]

Given the large differences between the estimation sample and out-of-sample outcome proportions, the out-of-sample predictions of crime rates, arrest rates and incarceration rates at ages 19 to 22, as seen in columns 1 and 2 of table 3, are remarkably good. The predicted crime rate for the out-of-sample observations is only .007 percentage points lower than the actual crime rate, the predicted arrest rate .002 lower than the actual and the predicted incarceration rate .001 lower than the actual. This is in sharp contrast to a model in which there is no unobserved heterogeneity ($M = 1$). In that model, although the difference is about the same for crime rates, the difference for arrest and incarceration rates are .060 and .025.[33] The last two columns of table 3, combining both the non-high school graduate and high school graduate samples, indicate a close correspondence between the actual and predicted outcomes for crime, arrest and incarceration rates for that combined sample.

---

[31] Although the linear age spline has knots at ages 18 and 21, and therefore contributes to the fit at ages 19-22 in a somewhat flexible way, the age trend is restricted to be the same for both the within- and out-of-sample observations.

[32] As we will see below, the type 3's are the most likley to attend school, and thus graduate, and the least likely to engage in crime.

[33] These differences are .049 and .028 for the joint normality specification.

The extent to which the outcome variables suffer from classification error is estimated to be small. The classification error rate, as given by the joint probability of observing a one when the true outcome is a zero plus the joint probability of observing a zero when the true outcome is a one is lowest for incarceration, .004, followed by school attendance, .040, crime, .044, arrest, 063, and work, .111.

The Effect of Initial Conditions on Early Adult Outcomes:

The importance of the two initial conditions, the unobserved type of a youth and whether a youth committed a crime at age 13, in accounting for early adult (age 19-22) outcomes is shown in table 4. The three types appear distinct.[34] Type 3's have the highest high school graduation rate, 85.2 percent for those youths who did not commit a crime at age 13 and 80.1 percent for those that did. These rates exceed those of type 2's by 41 percentage points for youths with $c_{13} = 0$ and by almost 50 percentage points for those with $c_{13} = 1$. The lowest graduation rates are those for type 1's; only one in four youths who did not commit a crime at age 13 graduate from high school and less than one in six of those who did commit a crime graduate. Crime rates, however, are not ordered in the same way. Type 2's are the most crime prone, followed by type 1's and then type 3's. Specifically, at its highest, the age 19-22 crime rate for type 2 youths who committed a crime at age 13 is .354, while the crime rate is only .122 for type 3 youths who did not commit a crime at age 13. Arrest rates, however are highest for type 1's even though they have, as noted, a lower crime rate than type 2 youths, exceeding those of type 2's by about 5 percentage points.[35] Arrest rates are

---

[34]In order to see whether three types were sufficient to capture permanent heterogeneity in our sample, we also estimated a specification in which mother's schooling was added to the VAR as a direct regressor and also as a determinant of type probabilities. Mother's schooling was treated as categorical: high school dropouts, high school graduates plus those with some college and college graduates. A likelihood ratio test did not reject the specification that excluded mother's schooling. We take this as evidence that the three type specifcation does a good job in capturing permanent heterogeneity.

[35]As seen, the number of periods in which type 1's are arrested exceeds the number of periods in which they committed a crime. One reason for the excess of arrests over crimes is because we treat a youth who has committed multiple crimes in a period as the same as a youth who committed only one crime. The former youth may be arrested in more than one future period for those multiple crimes, which will make it look like the youth was arrested more times than the number of times he committed a crime. Another reason is that a youth may in fact be arrested for a crime that he did not commit.

quite low for type 3's, both absolutely and relative to their crime rates. Incarceration rates, like arrest rates, are highest for type 1's and lowest for type 3's. Over one-quarter of type 1's who committed a crime at age 13 are incarcerated at ages 19-22. The same incarceration rate is 12.7 percent for type 2's and 0.2 percent for type 3's. As the table shows, both unobserved heterogeneity, a youth's type, and state dependence, whether the youth committed a crime at age 13, have significant impacts on young adult outcomes.

To obtain a measure of the relative importance of these initial conditions, table 5 shows the effect of changing each of the initial conditions. The first counterfactual makes every youth a type 3 and the second eliminates all crime at age 13. If all youths were type 3's, high school graduation rates would increase by 23.8 percentage points, crime rates between the ages of 19 and 22 would drop by 6.0 percentage points, arrest rates by 11.4 percentage points and incarceration rates by 6.4 percentage points (essentially eliminating all black youth incarceration). The effects of eliminating crime at age 13 are generally more modest. The increase in the high school graduation rates is 3.8 percentage points and the fall in the crime, arrest and incarceration rates are 2.1, 1.9 and 1.2 percentage points.

<u>The Correlates of Initial Conditions:</u>

The experiment of changing a youth's type or their criminal activity at age 13 is perhaps interesting, but would have operational content only if it were known how to accomplish those changes and at what cost. Table 6 shows how observable background characteristics of the youth's parents and the youth's AFQT score are related to initial conditions.[36] There are two reasons to look at these observable characteristics. First, one can see the extent to which observables account for the permanent characteristics captured by type and crime at age 13. Second, to the extent that changing these characteristics would also change the youth's initial conditions, that is, to the extent that these correlations do not reflect either intergenerationally transmitted immutable characteristics or other factors related to these characteristics, one can identify factors that it might prove useful to target in order to reduce youth crime and increase high school completion rates. Table 6 shows the fraction of each type and c_{13} initial-conditions combination by a number of characteristics, where we denote

---

[36]Type probabilities can be assigned to each sample observation (i) by an application of Bayes'rule: $\Pr(m, c_{13}|data_i) = \frac{L(data_i|m,c_{13})\Pr(m,c_{13})}{\Sigma_m \Sigma_{c_{13}} L(data_i|m,c_{13})\Pr(m,c_{13})}$, where L denotes the likelihood.

these fractions by Pmc_{13}, $m = 1, 2, 3$ and $c_{13} = 0, 1$.

As seen in table 6, the proportion of youths who are type 3's is monotonically increasing (with one exception) in mother's schooling, regardless of whether or not they commit a crime at age 13, and the proportion who are type 1's or 2's monotonically decreasing. For example, the proportion of youth who are type 3 (summing over c_{13}) is .432 if the youth's mother is a high school dropout and .690 if the youth's mother is a college graduate. Although these differences by mother's schooling are not substantively small (and are statistically significant), mother's schooling does not explain a large part of the variance in initial conditions. For example, the R^2 from a regression of $P30$ on the entire set of individual schooling level dummies of mother's schooling is only .041. And, while the mother schooling dummies are statistically significant, the R^2's from the same regression on the rest of the type and c_{13} combinations range from only .012 (for $P21$) to .056 (for $P10$).[37]

To quantify the size of the effect, consider the extreme scenario of making all mothers college graduates and assume that such a change would lead to the initial conditions distribution of that group shown in the table. Combined with the figures in table 4, one can compute the effect of that change on age 19 to 22 outcomes. In particular, the crime rate would be predicted to fall from .206 to .176 and the incarceration rate from .064 to .043. Although these changes are not insubstantial (though perhaps surprisingly small), achieving them in this way is outside the realm of feasibility.

Table 6 also shows that not having been born to a teen mother increases the probability of being a type 3 youth (for both values of c_{13}) and reduces the probability of the other initial conditions categories. Having lived with both biological parents also increases the probability of being a type 3 youth, but only for $c_{13} = 0$, and reduces the probability of being of type 1 and 2 (for both values of c_{13}). However, the differences in the initial conditions distributions for either background variable is fairly small as measured by their consequences for crime, arrests and incarcerations. For example, if all youths were born to non-teen mothers, and if that induced the initial conditions distribution associated with non-teen mothers, the predicted incarceration rate would fall from .064 to .059.

---

[37]Part of the variance in the iniital conditions proportions (Pmc_{13}) is due to measurement error, that is to the fact that these are sample estimates of the dichotomous true initial conditions.

The next two rows in table 6 contrast the initial conditions probabilities for two combinations of the three variables depicting worst and best case scenarios: (1) a youth whose mother is a high school dropout, who was born to a teen mother and who did not always live with both biological parents vs. (2) a youth whose mother is a college graduate, who was not born to a teen mother and who always lived with both biological parents. These are extremes in the sample, with only 4.8 percent of the sample falling into the worst case and 5.2 percent into the best case. Even in the best case scenario, the proportion of type 3 youths who do not commit a crime at age 13 is still below 55 percent.

Finally, the table shows the relationship between the (black-specific) quartiles of the AFQT score and initial conditions probabilities.[38] The difference between the lowest and highest quartile is similar to that for mother's schooling. Regression R^2's range from .025 to .079 over the initial conditions categories. Hence, even if we were able to move AFQT scores of black youths so that all achieve the fourth quartile level of performance, say by increasing school quality, and if we assume that by doing so we would also change the type and c_{13} probability distribution to that quartile's distribution, the proportion of type 3's would still be only about 70% of the population.[39]

How Persistent are "Transitory" Decisions on Early Adult Outcomes?

To understand the mechanism by which youth behaviors, such as attending school at age 16, working at age 16 while attending school and committing a crime at age 14, are related to early adult outcomes, we perform three counterfactual experiments that are based on individual-level "impulse response functions" derived (by simulation) from the VAR specification. The experiments demonstrate the extent to which these relationships stem from transitory shocks that set a youth on a decision path that is more likely to be aberrant

[38]The AFQT (Armed Forces Qualifying Test) score is a combination of several subtests of the ASVAB (Armed Services Vocational Aptitude Battery) based on a formula similar to the AFQT score generated by the DOD and found in the NLSY79. The AFQT score for the NLSY97 was created by the CHRR staff at The Ohio State University.

[39]We also looked at whether the youth lived in an urban area at age 12 and the youth's region of residence at age 12. The percentage of type 3 youths residing in an urban area at age 12 was about 9 percentage points lower than the percentage in a non-urban area. In addition, the largest proportion of type 3 youths resided in the West.

as opposed to a youth's permanent propensity to engage in aberrant behavior. In the first experiment (table 7), we simulate the impact of a youth committing a crime at age 14 on his likelihood of committing a crime and of being incarcerated at ages 19 to 22 (separately for each of the three types). We compare those outcomes to the case where the youth did not commit a crime at age 14, holding constant all combinations of whether or not the youth was arrested, incarcerated, attended school or worked at age 14.[40] We repeat this experiment by simulating the impact of the youth, having committed a crime at age 14, also being arrested at age 14 and then also being incarcerated.[41] In the second and third experiments, we use the same methodology to consider the impact on the likelihood of committing a crime and of being incarcerated at ages 19 to 22 of a youth working at age 16 while attending school as opposed to not working while attending school (table 8) and of a youth not attending school at age 16 as opposed to attending at age 16 (table 9).[42]

In table 7, the first row (for each type) shows the difference in the likelihood of committing a crime and of being incarcerated at ages 19-22 for a youth who does not commit a crime at age 14 and for a youth who, having committed a crime at age 14, is neither arrested or incarcerated. As seen, the crime rate for a type 1 black youth would be 7.5 percentage points higher at ages 19-22 for the youth had he committed a crime at age 14 (absent arrest and incarceration) than for the youth had he not committed a crime at age 14. Correspondingly, the probability that the youth would be incarcerated at ages 19-22 would be 9.6 percentage points higher.[43] The difference in the crime rate of a type 2 youth is slightly higher, 10.5

---

[40]All of the counterfactuals are based on simulating 86,400 outcome paths for each of the six initial condition categories. Letting the outcome vector at age 14 be $(c_{14}, A_{14}, J_{14}, s_{14}, h_{14})$, this counterfactual compares the later outcomes for youths with $(0, 0, 0, s_{14}, h_{14})$ to youths with $(1, 0, 0, s_{14}, h_{14})$ integrating over the four possible combinations of (s_{14}, h_{14}).

[41]These counterfactuals are based on comparisons of youths with the outcome vector at age 14 given by $(0, 0, 0, s_{14}, h_{14})$ to youths with the vector $(1, 1, 0, s_{14}, h_{14})$ and $(1, 1, 1, s_{14}, h_4)$, each taken over the four possible combinations of (s_{14}, h_{14}).

[42]The counterfactual in table 8 compares youths with age 14 outcome vectors $(c_{14}, A_{14}, J_{14}, 1, 1)$ to $(c_{14}, A_{14}, J_{14}, 1, 0)$ over all combinations of (c_{14}, A_{14}, J_{14}). Similarly, the counterfactual in table 9 compares youths with age 14 outcome vectors $(c_{14}, A_{14}, J_{14}, 0, 0)$ to $(c_{14}, A_{14}, J_{14}, 1, 0)$ over all combinations of (c_{14}, A_{14}, J_{14}).

[43]The increase in the incarceration rate may exceed the increase in the crime rate for several reasons.

26

percentage points, while the difference in the incarceration rate is somewhat lower, 5.5 percentage points. Recall that type 1 and type 2 youths have the highest crime rates over the three groups. However, the impact of having committed a crime at age 14 on the propensity to engage in crime at ages 19-22 is significant even for the less crime-prone type (type 3). Having committed a crime at age 14, a type 3 youth is 7.2 percentage points more likely to commit a crime at ages 19-22, comparable to the more crime-prone type 1 youth. However, the effect on the likelihood of incarceration is essentially zero for type 3.[44]

Next, consider the case of a youth who, having committed a crime at age 14, is arrested but not incarcerated. As the second row (for each type) shows, arrest without incarceration does not change the age 19-22 propensity to engage in crime relative to that of not being arrested (see row 1) for a type 1 youth, but increases it by 1.2 percentage points for type 2 and by 1.6 percentage points for a type 3 youths. Suppose, however, that the youth in addition to being arrested is also incarcerated. In that case, a type 1 youth would have an increased propensity to commit a crime at ages 19-22 that is 7.8 percentage points higher than a youth who did not commit a crime at age 14. This increase is only slightly greater than the increase of 7.5 percentage points if the youth had been arrested but not incarcerated or if the youth had not been arrested. The difference in the propensity given incarceation is only slightly larger than that given arrest, 11.8 vs. 11.7 percentage points.

The result that the increase in crime at ages 19-22 is greater when, having committed a crime at age 14, the youth is arrested without being incarcerated, and still greater when arrested and incarcerated, should not be interpreted as reflecting the lack of a deterrent effect for black youths in the usual sense, that is, as a response to a change in the *probability*

First, an individual may be incarcerated for multiple years for a single crime. Second, we do not account for the commission of multiple crimes in a given year. Thus, for example, an individual who commits two crimes at age 19 may be arrested and incarcerated at age 19 for one crime and then arrested and incarcerated at age 20 for the other crime.

[44]One possible explanation for the negligible effect on incarceration is that type 3 youths may commit less serious crimes, which we do not account for in the estimation. Although not reported, as with the likelihood of incarceration, the effect on the probability of arrest is very small for type 3. Thus, another possibility is that youths of these types are more skilled criminals.

of arrest or incarceration.[45] What is true of a single youth who, having committed a crime at age 14 (randomly) experiences an arrest or a spell of incarceration relative to one who does not, would not necessarily be true if there was a general increase in the probability of arrest or incarceration faced by all youths. This fallacy of composition arises because a change in the probability of arrest or incarceration, as emphasized previously, would change the parameters of the VAR, which depend on the entire set of structural parameters of the behavioral model.

Arrest at age 14 also increases the likelihood of being incarcerated at ages 19 to 22 relative to not being arrested, by 1.9 percentage points for a type 1 youth and by 1.7 percentage points for a type 2 youth. This increase in the likelihood of incarceration perhaps reflects an increased capability by the police to identify someone as having committed a crime if he had been arrested previously.[46] As also seen in the table, for type 1 youths, incarceration at age 14 increases the likelihood of incarceration at ages 19-22 by 1.1 percentage points relative to the case where the youth was arrested but not incarcerated (12.6 vs. 11.5 percentage points) and by 0.6 percentage points for type 2 youths.[47]

Table 8 compares outcomes at ages 19-22 for a youth who works at age 16 while in school as opposed to one who does not. As seen, a youth, having experienced a shock (or set of shocks) that induces him to work while attending school at age 16, has a lower propensity to commit a crime at ages 19 to 22 ranging from 3.7 percentage points for a type 3 youth to 6.4 percentage points for a type 2 youth. An age 16 type 1 youth who works while attending school also has a 7.7 percentage point lower likelihood of being incarcerated at ages 19-22 relative to a youth who attends school without working. The effect is 4.1 percentage points for a type 2 youth and essentially zero for a type 3.

The effect on crime and incarceration at ages 19 to 22 of not attending school at age 16 are shown in table 9. For all types, not attending school at age 16 increases the propensity

---

[45]The modern empirical literature on crime and deterrence began with Ehrlich (1973).

[46]In addition, although, as noted, we do not account explicitly for the type of crime, possibly a youth who is arrested at age 14 commits more serious crimes at later ages than a youth who is not arrested.

[47]Note that even with over 80,000 simulated paths, the number of times a type 3 youth was arrested and incarcerated at age 14 was too small to obtain a reliable comparison.

to commit a crime significantly, from a low of 7.2 percentage points for a type 1 youth to a high of 14.8 percentage points for a type 3 youth. The likelihood of incarceration also increases non-trivially for a type 1 youth, by 8.1 percentage points, and for a type 2 youth, by 5.6 percentage points. As was true for the other experiments, there is no effect on the probability of incarceration for type 3 youths.

The Transition from School to Jail:

Table 1 showed that youths who do not attend school at age 16 have incarceration rates that are 4 times greater at ages 19 to 22, crime rates that are two-thirds greater, arrests rates that are 2.5 times greater and a high school graduation rate that is almost 70 percentage points lower. Table 10 uses the same simulated data as for the counterfactual experiments presented above to determine the extent to which controlling for classification error, for the inter-related nature of youth decisions/outcomes and for unobserved heterogeneity affect those comparisons. The first row of table 10 repeats the first row of table 1. The second row shows the fit of the VAR to the data, that is, it takes the data simulated from the VAR, adds classification error based on the estimates, and shows the same differences in means by school attendance at age 16. As seen, the predictions are close in terms of the point estimates and, with the exception of the high school graduation rate, fall well within the 95 percent confidence intervals of the differences in means in the actual data.[48]

The third row takes the simulated data with classification error, estimates linear probability models for each outcome at ages 19 to 22 on whether or not the youth was in school at age 16 and on the other age 16 decisions/outcomes (whether or not the youth committed a crime at age 16, was arrested at age 16, was incarcerated at age 16 and worked at age 16) and reports the coefficient on the age 16 school attendance variable.[49] The next row reports the same regression coefficient after adding dummy variables for the youth's type.

---

[48]Even with respect to the high school graduation rate, the predicted point estimate is only 5 percentage points too high.

[49]We also estimated a five-equation restricted specification in which these decision/outcomes were not inter-related through their histories. Each equation included types, the age spline and the own lagged variable. The specification also allowed for contemporaneous error correlations. A likelihood ratio test rejected the specification.

As the table shows, accounting for the other decision/outcomes that arise along with the school attendance decision at age 16 reduces the impact of not attending school at age 16 on all of the young adult outcomes. Heterogeneity further reduces the impact of not attending school, generally by a large amount and to a greater extent than including the other age 16 outcomes. Overall, the difference in the likelihood of being incarcerated at ages 19-22 falls from 11.5 percentage points (row 2, the prediction from the VAR estimates) to 8.4 percentage points when the other age 16 variables are included, and finally to only 2.8 percentage points when the type dummies are also included. Similarly, the difference in the probability of committing a crime falls from 11.6 to 6.6 percentage points and in the probability of arrest from 16.4 to 3.6 percentage points. The difference in the high school graduation rate is also attenuated, by 20 percentage points.

The next three rows perform the same regressions accounting for the effect of classification error, that is, by using the simulated data without adding on the classification error. As seen, effects are larger when classification error is accounted for. In particular, the effect of school attendance on the likelihood of incarceration controlling for both other age 16 outcomes and for heterogeneity increases from 2.8 to 3.4 percentage points when classification error is taken into account. Similarly, the effect on the probability of committing a crime increases from 6.6 to 9.7 percentage points, the effect on the likelihood of arrest increases from 3.6 to 6.2 and the effect on the probability of graduating from high school increases (in absolute value) from 52.9 to 64.4 percentage points.

The last row in table 10 provides the "true" estimates. These estimates allow not only for classification error, for other inter-related outcomes and for (additive) heterogeneity, but also for the heterogeneous effects that arise from the VAR specification. They are obtained directly from the counterfactuals performed in table 9 for crime and incarceration probabilities and from the same counterfactuals (not reported) for arrest and high school graduation probabilities. The effect of not attending school at age 16 on the outcomes are simply the weighted averages of the effects for each of the three types, where the weights are the estimated sample type proportions. As seen, these estimates imply that not attending school at age 16 increases the likelihood of being incarcerated at ages 19 to 22 by 3.2 percentage points on average, the probability of committing a crime by 12.9 percentage points and of an

arrest by 6.1 percentage points. Not attending school at age 16 also reduces the probability of graduating from high school by 75.4 percentage points on average. Thus, the degree to which the simple comparison of outcomes for those youths that do or do not attend school at age 16 would be misleading varies across outcomes; the difference between the simple comparison and the "true" estimate is large for incarceration (3.2 vs. 11.5) and for arrest (6.1 vs. 16.4), but small for crime (12.9 vs.11.6) and for high school graduation (75.4 vs. 73.1).

VIII. Conclusions

Using data on black male youths from the NLSY97 we have estimated a five-variate discrete-outcome vector autoregression. The five variables, school attendance, employment, criminal activity, arrest and incarceration, were viewed as representing inter-related youth decisions/outcomes. The VAR approach is especially well suited to provide a quantitative assessment of the impact of stochastic occurrences during one's youth on later outcomes and to distinguish those impacts from permanent behavioral differences among youths. What, for example, are the consequence for young adult outcomes of not dropping out of school at age 16 because, for example, a youth happens to have an inspiring teacher that year? Or, what are the consequences of stealing something at age 14 because a random opportunity arose in which the likelihood of being caught appeared to be very low? What happens if that youth is nevertheless apprehended, and what if the youth is incarcerated as well? Or, what are the consequences if a 16 year old youth happens to come upon an appealing after-school job?

In estimating the effects of stochastically generated path dependencies, we accounted for permanent unobserved heterogeneity in the youths' traits and environment and for measurement error in the data. The estimation revealed that there were three distinct types in the NLSY97 sample of black male youths. One type has a high school graduation rate of around 20 percent, another around 40 percent and the third around 85 percent. The two lower graduation types engage in crime and are arrested more often when young adults. Among those two types, the lowest graduation type is incarcerated at over twice the rate.

We found that stochastic occurrences can also be important in determining outcomes as young adults and can vary considerably by permanent underlying propensities. To take

one example, for the two lower graduation types, comprising 47 percent of black youths, the impact of not attending school at age 16, which is tantamount to never completing high school, is to increase the probability of being incarcerated as a young adult (ages 19-22) by 8.1 and 5.6 percentage points. These represent increases in the probability of incarceration for these two types of 39 percent and 58 percent. On the other hand, not attending school at age 16 does not affect the incarceration probability for the high graduation type, although the probability of committing a crime at ages 19-22 increases markedly. Either youths of this type, who ordinarily have a high school graduation rate around 85 percent, are "smart" criminals or commit crimes that are unlikely to be punished by incarceration (or both).

There is a growing literature that appears to indicate that the behavior of young adults is to a large extent determined by permanent characteristics (cognitive and non-cognitive skills) developed during childhood.[50] The results in this paper illustrate, however, that stochastic events also may play a significant role, setting youths off on a path that leads them to engage in socially undesirable activities as young adults.

---

[50]See Keane and Wolpin (1997) and Cunha and Heckman (2007) for example.

<center>REFERENCES</center>

[1] Blumstein, Alfred, Cohen, Jacqueline, Roth, Jeffrey A. and Christy A. Visher, Editors (1986). *Criminal Careers and "Career Criminals", Vols I and II*. Washington, DC: National Academy Press.

[2] Broidy, Lisa M., Nagin, Daniel S., Tremblay, Richard E., Bates, John, Brame, Bobby, Dodge, Kenneth A., Fergusson, David, Horwood, John L., Loeber, Rolf, Laird, Robert, Lynam, Donald R., Moffitt, Terrie E., Pettit, Gregory S. and Frank Vitaro (2003). "Developmental Trajectories of Childhood Disruptive Behaviors and Adolescent Delinquency: A Six-Site, Cross-National Study." *Developmental Psychology*, 39, 222-245.

[3] Cunha, Flavio and James Heckman (2007). "The Technology of Skill Formation." *American Economic Review, Papers and Proceedings*, 97, 31-47.

[4] Eckstein, Zvi and Kenneth I. Wolpin (1999). "Why Youth Drop out of High School: The Impact of Preferences, Opportunities and Abilities." *Econometrica*, 67, 1295-1339.

[5] Ehrlich, Isaac (1973). "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." *Journal of Political Economy*, 81, 521-65.

[6] Fella, Giulio and Giovanni Gallipoli (2005). "Education and Crime over the Life Cycle." Working Paper, University of British Columbia.

[7] Freeman, Richard B. (1992). "Crime and the Employment of Disadvantaged Youth," in George Peterson and Wayne Vroman, eds., *Urban Labor Markets and Job Opportunity*. Washington, DC: Urban Institute.

[8] Freeman, Richard B. (1996). "Why Do So Many Young American Men Commit Crimes and What Might We Do About It?" *Journal of Economic Perspectives*, 10, 25-42.

[9] Gottfredson, Denise C. (1985). "Youth Employment, Crime, and Schooling: A Longitudinal Study of a National Sample." *Developmental Psychology*, 21, 419-432.

[10] Gottfredson, Michael and Travis Hirschi (1990). *A General Theory of Crime*. Stanford, CA: Stanford University Press.

<center>33</center>

[11] Gould, Eric D., Weinberg, Bruce, A. and David B. Mustard (2002). "Crime Rates and Local Labor Market Opportunities in the United States: 1977-1997." *Review of Economics and Statistics*, 84, 45-61.

[12] Grogger, Jeffrey (1995). "The Effect of Arrests on the Employment and Earnings of Young Men." *Quarterly Journal of Economics*, 110, 51-71.

[13] Grogger, Jeffrey (1998). "Market Wages and Youth Crime." *Journal of Labor Economics*, 16, 756-791.

[14] Hausman, Jerry A., J. Abrevaya and F.M. Scott-Morton (1998). "Misclassification of the Dependent Variable in the Discrete Response Setting." *Journal of Econometrics*, 87, 239-269.

[15] Heckman, James J. (1981). "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in Charles F. Manski and Daniel McFadden, eds., *The Structural Analysis of Discrete Data*. Cambridge, MA: MIT Press.

[16] Hjalmarsson, Randi (2008). "Criminal Justice Involvement and High School Completion." *Journal of Urban Economics*, 63, 613-630.

[17] Imai, Susumu and Kala Krishna (2004). "Employment, Deterrence, and Crime in a Dynamic Model." *International Economic Review*, 45, 845-872.

[18] Jacob, Brian A. and Lars Lefgren (2003). "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration and Juvenile Crime." *American Economic Review*, 93, 1560-1577.

[19] Keane, Michael P. and Robert M. Sauer (2005). "A Computationally Practical Simulation Estimation Algorithm for Dynamic Panel Data Models with Unobserved Endogenous State Variables." Working Paper, University of Bristol.

[20] Keane, Michael P. and Kenneth I. Wolpin (1997). "The Career Decisions of Young Men." *Journal of Political Economy*, 105, 473-522.

[21] Keane, Michael P. and Kenneth I. Wolpin (2001). "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review*, 42, 1051-1103.

[22] Kling, Jeffrey R. (2006). "Incarceration Length, Employment, and Earnings." *American Economic Review*, 96, 863-876.

[23] Lochner, Lance (2004). "Education, Work and Crime: A Human Capital Approach." *International Economic Review*, 45, 811-843.

[24] Lochner, Lance (2007). "Individual Perceptions of the Criminal Justice System." *American Economic Review*, 97, 444-460.

[25] Lochner, Lance and Enrico Moretti (2004). "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review*, 94, 155-189.

[26] Mauer, Marc (1985). *Young Black Americans and the Criminal Justice System: Five Years Later.* Washington D.C.: The Sentencing Project.

[27] Nagin, Daniel S., Farrington, David P. and Terrie E. Moffitt (1995). "Life-Course Trajectories of Different Types of Offenders." *Criminology*, 33, 111-139.

[28] Nagin, Daniel S. and Kenneth C. Land (1993). "Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric, Mixed Poisson Model." *Criminology*, 31, 327-362.

[29] Nagin, Daniel S. and Raymond Paternoster (1991). "On the Relationship of Past to Future Participation in Delinquency." *Criminology*, 29, 163-189.

[30] Ploeger, Matthew (1997). "Youth Employment and Delinquency: Reconsidering a Problematic Relationship." *Criminology*, 35, 659-675.

[31] Shannon, Lyle W. (1982). *Assessing the Relationship of Adult Criminal Careers to Juvenile Careers: A Summary.* Washington, DC: U.S. Department of Justice, National Institute of Justice.

[32] Sickles, Robin C. and Jenny Williams (2008). "Turning from Crime: A Dynamic Perspective." *Journal of Econometrics*, forthcoming.

[33] Tauchen, Helen, Witte, Ann Dryden and Harriet Griesinger (1994). "Criminal Deterrence: Revisiting the Issue with a Birth Cohort." *Review of Economics and Statistics*, 76, 399-412.

[34] U.S. Department of Justice (2008). *Prison and Jail Inmates at Midyear 2007.* Washington, DC: Bureau of Justice Statistics.

[35] Waldfogel, Joel (1994). "The Effect of Criminal Conviction on Income and the Trust 'Reposed in the Workmen'." *Journal of Human Resources*, 29, 62-81.

[36] Williams, Jenny and Robin C. Sickles (2002). "An Analysis of the Crime as Work Model: Evidence from the 1958 Philadelphia Birth Cohort Study." *Journal of Human Resources*, 37, 479-509.

[37] Wilson, James Q. and Richard J. Herrnstein (1985). *Crime and Human Nature.* New York, NY: Simon & Schuster.

[38] Witte, Ann Dryden and Helen Tauchen (1994). "Work and Crime: An Exploration Using Panel Data." *Public Finance*, 49, S155-S167.

Table 1
The Relationship of Youth Behaviors and Family Background to Young Adult Incarceration, Crime, and Arrests Rates and High School Completion[a]

	Percent	Percent Incarcerated Ages 19-22	Percent Commit Crime Ages 19-22	Percent Arrested Ages 19-22	Percent HS Graduate
Attend School at Age 16					
No	22.8	16.4	30.6	27.1	6.1
Yes	77.2	4.0	18.4	11.6	74.3
Difference	-	12.4	12.2	15.5	-68.2
95% Conf. Int.	-	(6.8,18.0)	(5.6,18.8)	(9.5,21.5)	(-72.2,-64.2)
Work and Attend School at Age 16					
No	80.6	4.2	16.4	12.2	73.0
Yes	19.4	3.5	18.0	7.6	79.9
Difference	-	0.7	-1.6	4.6	-6.9
95% Conf. Int.	-	(-2.5,3.8)	(-9.0,5.9)	(-0.1,9.2)	(-13.8,0.0)
Commit Crime at Age 14					
Yes	26.6	4.1	35.8	19.7	44.3
No	73.4	3.9	17.1	9.4	63.3
Difference	-	0.2	18.7	10.3	-19.0
95% Conf. Int.		(-3.8,4.1)	(9.6,27.9)	(3.5,6.6)	(-27.2,-10.8)
Mother High School Dropout					
Yes	25.7	9.6	20.6	16.4	37..8
No	74.3	5.1	20.2	13.1	66.2
Difference	-	4.5	0.4	3.3	-28.4
95% Conf. Int.	-	(0.1,8.9)	(-5.2,6.0)	(-1.5,8.1)	(-35.0,-21.8)
Teen Mother					
Yes	42.3	8.2	20.2	14.2	53.5
No	57.7	4.6	20.4	14.3	65.0
Difference	-	3.6	-0.2	-0.1	-12.5
95% Conf. Int.	-	(0.0,7.2)	(-5.4,5.0)	(-4.3,4.1)	(-18.7,-6.4)
Always Live in Nuclear Family					
No	34.4	8.2	23.6	18.7	54.7
Yes	65.6	5.4	19.1	12.7	59.7
Difference	-	2.8	4.5	6.0	-5.0
95% Conf. Int.	-	(-0.8,6.4)	(-0.8,9.8)	(1.5,10.5)	(-11.0,1.0)

Table 2
Model Fit: Estimation Sample

Age	In School Actual	In School Predicted	Crime Actual	Crime Predicted	Arrest Actual	Arrest Predicted	Incarceration Actual	Incarceration Predicted	Work Actual	Work Predicted
14	.939	.940	.266	.281	.098	.075	.015	.009	.026	.038
15	.867	.859	.287	.281	.101	.112	.026	.024	.083	.097
16	.772	.760	.273	.267	.112	.136	.036	.037	.188	.177
17	.660	.646	.256	.249	.129	.150	.047	.048	.295	.297
18	.412	.397	.238	.261	.205	.206	.075	.079	.370	.430
19	.118	.123	.291	.293	.233	.246	.122	.113	.420	.467
20	.033	.027	.296	.307	.273	.248	.111	.120	.400	.444
21	.014	.011	.285	.266	.283	.246	.177	.137	.415	.465
22	.000	.005	.174	.199	.214	.212	.136	.151	.478	.510
14-18	.742	.751	.266	.268	.124	.129	.038	.035	.190	.187
19-22	.067	.070	.284	.287	.253	.245	.130	.121	.417	.463

Table 3
Model Fit, Ages 19-22 : Out-of-Sample and Combined Samples

	Out-of-Sample		Combined Samples	
	Actual	Predicted	Actual	Predicted
Crime	.154	.147	.206	.206
Arrest	.072	.070	.145	.144
Incarceration	.019	.018	.064	.062

Table 4
Selected Outcomes by Initial Conditions: Model Predictions

	Type 1		Type 2		Type 3	
	$c_{13}=0$	$c_{13}=1$	$c_{13}=0$	$c_{13}=1$	$c_{13}=0$	$c_{13}=1$
HSG	.250	.150	.443	.320	.852	.801
$c_{19\text{-}22}$	.182	.238	.271	.354	.122	.177
$A_{19\text{-}22}$	.280	.343	.224	.297	.022	.033
$J_{19\text{-}22}$	.185	.262	.084	.127	.002	.002
Percent	12.4	5.2	20.3	9.1	42.3	10.7

$c_{13} = 1$ if committed a crime at age 13, = 0 otherwise.

$$X_{1922} = \sum_{i} \sum_{a=19}^{22} X_{ia} / \sum_{a=19}^{22} N_a ,$$

Where X = c: crime,
      = A: arrest,
      = J: incarcerated.

    N_a = number of simulations
      = 10,800

Table 5
Effect on Selected Outcomes of Changing Initial Condition

	All Type 3	All $c_{13}=0$
Δ HSG	.238	.038
Δ $c_{19\text{-}22}$	-.060	-.021
Δ $A_{19\text{-}22}$	-.114	-.019
Δ $J_{19\text{-}22}$	-.064	-.012

Table 6
Selected Correlates of Initial Conditions

	P30	P31	P20	P21	P10	P11	Proportion of Sample
Overall Proportion	.423	.107	.203	.091	.124	.052	1.00
Mother's Schooling							
< 12	.341	.091	.228	.098	.174	.069	25.7
12	.441	.087	.212	.087	.122	.049	42.4
13-15	.475	.137	.182	.086	.086	.034	21.5
16 +	.523	.167	.136	.070	.071	.033	10.4
Teen Mother							
Yes	.383	.093	.226	.090	.153	.055	.423
No	.464	.114	.188	.086	.101	.046	.577
Always Lived With Both Bio Parents							
No	.378	.112	.210	.102	.135	.062	.656
Yes	.445	.102	.202	.084	.120	.046	.344
Mother's Schooling <12, Teen Mother, Not Always Lived with Both Bio Parents	.245	.081	.238	.105	.227	.105	.048
Mother's Schooling >=16, Non-Teen Mother, Always Lived with Both Bio Parents	.548	.184	.132	.066	.054	.015	.052
AFQT							
1st Quartile	.333	.068	.239	.118	.166	.075	.25
2nd Quartile	.392	.118	.225	.089	.131	.044	.25
3rd Quartile	.502	.109	.180	.070	.102	.037	.25
4th Quartile	.558	.159	.154	.055	.052	.022	.25

Table 7
The Effect of Crime, Arrests and Incarceration at Age 14 on Crime and Incarceration Rates at Age 19-22

Type	c_{14}	A_{14}	J_{14}	$\Delta c_{19\text{-}22}$	$\Delta J_{19\text{-}22}$
	1	0	0	.075	.096
1	1	1	0	.075	.115
	1	1	1	.078	.126
	1	0	0	.105	.055
2	1	1	0	.117	.072
	1	1	1	.118	.078
	1	0	0	.072	.001
3	1	1	0	.088	.002
	1	1	1	-	-

$\Delta x_{19\text{-}22} = x_{19\text{-}22} (c_{14}, A_{14}, J_{14}) - x_{19\text{-}2}(0, 0, 0); x=c,J.$

"-" indicates too few simulated observations in (c_{14}, A_{14}, J_{14}).

Table 8
The Effect of Working at Age 16 While Attending School on Crime and Incarceration Rates at Ages 19-22

Type	$\Delta c_{19\text{-}22}$	$\Delta J_{19\text{-}22}$
1	.-.045	-.077
2	-.064	-.041
3	-.037	-.001

$\Delta X_{19\text{-}22} = X_{19\text{-}22} (s_{16}=1, h_{16}=1) - X_{19\text{-}22} (s_{16}=1, h_{16}=0); X=c,J$

$s_{16} = 1$ if in school at age 16, $h_{16} = 1$ if work at age 16

$s_{16} = 0, h_{16} = 0$ otherwise

Table 9
The Effect of Not Attending School at Age 16 on Crime and Incarceration Rates at Ages 19-22

Type	$\Delta c_{19\text{-}22}$	$\Delta J_{19\text{-}22}$
1	.072	.081
2	.127	.056
3	.148	.002

$\Delta X_{19\text{-}22} = X_{19\text{-}22} (s_{16}=0) - X_{19\text{-}22} (s_{16}=1)$; X=c,J

$s_{16} = 1$ if in school at age 16

$s_{16} = 0$ otherwise

Table 10
The Importance of Heterogeneity and Classification Error in Estimating the Effect of School Attendance at
Age 16 on Young Adult Outcomes

$(s_{16} = 0) - (s_{16} = 1)$	Change in Percent Incarcerated Ages 19-22	Change in Percent Commit Crime Ages 19-22	Change in Percent Arrested Ages 19-22	Change in Percent HS Graduate
Data	12.4 (6.8, 18.0)	12.2 (5.6, 18.8)	15.5 (9.5, 21.5)	-68.2 (-64.2, -72.2)
Model Prediction – Not Accounting for Classification Error				
No Heterogeneity - Exc. Other Outcomes	11.5	11.6	16.4	-73.1
No Heterogeneity - Inc. Other Outcomes	8.4	9.3	13.2	-68.5
Heterogeneity - Inc. Other Outcomes	2.8	6.6	3.6	-52.9
Model Prediction – Accounting for Classification Error				
No Heterogeneity Exc. Other Outcomes	12.3	14.4	19.0	-79.6
No Heterogeneity - Inc. Other Outcomes	8.9	12.0	15.2	-76.0
Heterogeneity Inc. Other Outcomes	3.4	9.7	6.2	-64.4
Weighted Average of Heterogeneous Effects	3.2	12.9	6.1	-75.4

VAR Parameter Estimates
(Asymptotic Standard Errors in Parentheses)

	s_a	c_a	A_a	J_a	h_a
Type 3	-0.658	0.645	-2.968	-3.898	-5.538
	(0.032)	(0.014)	(0.013)	(0.023)	(0.016)
Type1-Type3	-1.272	-0.026	1.197	1.538	0.027
	(0.065)	(0.009)	(0.064)	(0.030)	(0.040)
Type2-Type3	-0.887	-0.152	0.0872	1.002	-0.271
	(0.030)	(0.014)	(0.051)	(0.040)	(0.031)
s_{a-1}	6.402	-0.495	-0.232	-0.209	0.053
	(0.037)	(0.006)	(0.013)	(0.009)	(0.011)
c_{a-1}	-0.321	2.160	1.042	0.579	0.172
	(0.027)	(0.029)	(0.042)	(0.039)	(0.007)
A_{a-1}	-0.286	-0.030	1.246	1.037	0.068
	(0.038)	(0.009)	(0.041)	(0.063)	(0.133)
J_{a-1}	-0.386	-0.314	-0.408	1.147	-1.812
	(0.081)	(0.023)	(0.163)	(0.067)	(0.346)
h_{a-1}	0.188	-0.193	-0.311	-0.391	3.012
	(0.033)	(0.015)	(0.050)	(0.099)	(0.101)
Age					
Spline 1	-0.240	-0.105	0.046	0.032	0.265
	(0.002)	(0.001)	(0.003)	(0.001)	(0.001)
Spline 2	-0.535	-0.007	-0.042	0.021	-0.345
	(0.112)	(0.018)	(0.010)	(2.322)	(0.032)
Spline 3	-0.149	-0.234	-0.057	0.119	0.272
	(1.582)	(0.012)	(0.024)	(0.035)	(0.229)

Correlation Matrix Parameters

	s_a	c_a	A_a	J_a	h_a
s_a	1.000				
c_a	0.094	1.000			
	(0.009)				
A_a	-0.272	0.512	1.000		
	(0.005)	(0.007)			
J_a	-0.275	0.236	0.818	1.000	
	(0.014)	(0.006)	(0.005)		
h_a	-0.248	-0.092	0.116	-0.024	1.000
	(0.022)	(0.011)	(0.008)	(0.005)	

Table A.1 continued

<table>
<tr><td colspan="2" align="center">Type Probability Parameters</td></tr>
<tr><td>Type 1</td><td></td></tr>
<tr><td> Constant</td><td align="center">-1.2251</td></tr>
<tr><td></td><td align="center">(0.228)</td></tr>
<tr><td> c_{13}</td><td align="center">0.504</td></tr>
<tr><td></td><td align="center">(0.442)</td></tr>
<tr><td>Type 2</td><td></td></tr>
<tr><td> Constant</td><td align="center">-0.732</td></tr>
<tr><td></td><td align="center">(0.242)</td></tr>
<tr><td> c_{13}</td><td align="center">0.569</td></tr>
<tr><td></td><td align="center">(0.426)</td></tr>
<tr><td colspan="2" align="center">Classification Error Parameters</td></tr>
<tr><td> s_a</td><td align="center">1.526</td></tr>
<tr><td></td><td align="center">(0.116)</td></tr>
<tr><td> c_a</td><td align="center">1.697</td></tr>
<tr><td></td><td align="center">(0.128)</td></tr>
<tr><td> A_a</td><td align="center">0.706</td></tr>
<tr><td></td><td align="center">(0.149)</td></tr>
<tr><td> J_a</td><td align="center">3.089</td></tr>
<tr><td></td><td align="center">(0.503)</td></tr>
<tr><td> h_a</td><td align="center">0.882</td></tr>
<tr><td></td><td align="center">(0.094)</td></tr>
</table>

lnL=-8,550

Table A.2
Assessments of Alternative Degrees of Unobserved Heterogeneity

	No Heterogeneity	Two Types	Three Types	Four Types	Normal Distribution
Log Likelihood	-8,654	-8,587	-8,550	-8,540	-8,628
No. of Parameters	60	67	74	81	80
AIC	17,428	17,312	17,258	17,241	17,417
BIC	17,825	17,755	17,748	17,778	17,947