Employment, Deterrence and Crime in a Dynamic Model*

Susumu Imai†
Concordia University and CIREQ

Kala Krishna
The Pennsylvania State University and NBER

January 31, 2004

Abstract

Using maximum likelihood techniques and monthly panel data we solve and estimate an explicitly dynamic model of criminal behavior where current criminal activity adversely affects future labor market outcomes. Since individuals look forward to the consquences of their behavior, the threat of future adverse effects in the labor market when caught committing crimes reduces the incentive to commit them.

We show that the effect of such forward looking behavior is strong in the data. Hence, policies which weaken it will be less effective in fighting crime. This suggests that prevention is more powerful than redemption since anticipated redemption allows criminals to look forward to negating the consequences of their crimes.

*We are grateful to Steve Collins of the FBI for many useful conversations. We are also grateful to Ricardo Cavalcanti, Chris Ferrall, Dan Houser, Steve Levitt, Bob Marshall, Antonio Merlo, Tatiana Michailova, Boris Molls, Anne Piehl, Maria Pisu, Mark Roberts, John Rust, Robin Sickles, Norm Swanson, Helen Tauchen, Tor Winston and Anne Witte and anonymous referees for useful comments.

†Address all correspondence to Susumu Imai at: Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, QC. H3G 1M8 or e-mail: simai@alcor.concordia.ca

1 Introduction

Public policy concerning crime and punishment is indisputably of great importance in the U.S.. Criminal behavior involves choices which impact on the future, and on the payoffs from choices made in the future. Hence, explicitly dynamic models of individual behavior need to be estimated. Dynamic models incorporate both state dependence and forward-looking behavior. The former deals with how past arrests affect current criminal choice. The latter deals with how current criminal choice is affected by the future consequences of today's criminal choices. Forward-looking behavior looks at effects directly through utility, as well as indirectly through labor market consequences. We call deterrence through future labor market consequences "dynamic deterrence".

Consider the effect of rehabilitation in prison. While this would reduce crime upon release by raising the payoff of not committing crimes due to state dependence, it would tend to increase crimes early on because of forward-looking behavior. Much of the work by economists has been static and/or reduced form in nature, see for example, the seminal theoretical work by Becker (1968), and recent empirical work by Tauchen et. al. (1994), Witte and Tauchen (1994) and Grogger (1995). While such work is undoubtedly useful, it may give different policy insights than a dynamic model would, as well as being less amenable to counter-factual policy experiments than a more structural approach.

There is evidence that dynamic aspects are important in understanding criminal behavior. For example, since juvenile records are sealed at age 18 and juvenile courts sanctions are much milder than those in adult courts, there is reason to expect crime to be higher below age 18. Levitt (1997) shows that states where juvenile punishments are relatively mild compared to adult ones see a sharper drop off in the age arrest profile after 18 than states where juvenile punishments are relatively harsh. This is consistent with anticipatory behavior on the part of individuals.

The only paper we are aware of that begins to take a dynamic structural approach is that of Williams and Sickles (1997). However, their focus is on how differences across individuals in the extent of initial social capital translate into different behavior and hence different paths of social capital and career choices. This explains how criminals and non-criminals can face similar wages yet make different choices. They estimate a model of continuous choice of hours of criminal activities using the Euler equation GMM approach. In contrast, we use a maximum likelihood approach and emphasize the choice of whether to commit a crime or not and the consequences on future employment outcomes. We also include both criminal choice during high school and beyond into our estimation.[1]

[1] The Euler equation GMM estimation technique used by Williams and Sickles (1997) works well with continuous data, and not so well with the discrete crime data they use.This is because they have to infer the number of hours allocated for criminal activities on the basis of arrests. In addition, they do not focus on dynamic deterrence or counter-factual experiments.

Lochner (1999) emphasizes the role of human capital accumulation on criminal behavior. There have also been simulation studies of criminal behavior based on calibrated dynamic models of representative agents. Among them are Flinn (1986), Leung (1994), Bearse (1997) and Imrohoroglu, Merlo and Rupert (2000). However, there has been little effort devoted to actually estimating a dynamic model.

We model the choice of committing a crime to be a function of variables like wages and employment today as well as variables like future wages and employment, which are affected by the outcome today. While we explicitly solve the dynamic choice problem faced by agents in the process of estimation, we do not consider any general equilibrium effects like those modeled by Burdett, Lagos and Wright (2003), Huang, Laing and Wang (2003) and Imrohoroglu, Merlo and Rupert (2000). We leave such work for the future.

We allow for unobserved heterogeneity insomuch as there are 4 types of individuals. Type probability assignment is estimated to maximize the likelihood function. We find that forward-looking behavior regarding labor market outcomes is important in deterring crime. Although wages are not strongly affected by criminal history in our data, the probability of unemployment depends positively on past arrest records and this drives the result. The paper proceeds as follows. The data are described in Section 2. The model specification is discussed in Section 3, and estimation results are presented in Section 4. Section 5 presents some simulation exercises including policy experiments. Section 6 contains some concluding remarks. The details of the model and its solution algorithms are presented in the Appendix.

2 Data

Our data comes from the 1958 Philadelphia Birth Cohort Study developed by Figlio et. al. (1994). The cohort consists of all individuals who were born in 1958 and who resided in Philadelphia from age 10 to 18. The entire cohort was stratified on the basis of gender, socioeconomic status when growing up, race, and the number of juvenile offenses. A stratified random sample of 577 men and 201 women was taken from the cohort. Detailed information on all their juvenile and adult records was collected. Then, in 1988 the sample from the cohort was resurveyed. The data provides, among other things, detailed juvenile as well as adult arrest records[2], basic demographic information, and employment and schooling records from age 14 to 26. The dataset contains information over a relatively long time period. Moreover, it is drawn from the general youth population of Philadelphia. This is in contrast to many datasets on crime, which only focus on delinquents.

We use only the male sample because males are significantly more criminally active. We could link the various records and recover the necessary variables for our analysis for only 440 individuals in the male sample. We

[2]We ignore arrests for certain offense categories, in particular, arrests for driving while intoxicated, drunkenness, disorderly conduct, vagrancy, cruelty to animals, selling fireworks, fortune telling, violation of cigarette tax act, scavenging, Sunday law violation (except sale of liquor), traffic and motor vehicle violations.

dropped all agents going to college from the sample. 66 individuals out of 440, i.e. less than 20 percent of the sample went to college. After we removed the individuals who did not have a proper birth record, we were left with 364 individuals. Since individuals who only went to trade schools and other similar institutions attended them sporadically, we treated them as if they did not attend school.

The demographic information included variables such as sex, race, date of birth, church membership, and the socioeconomic status of the individual. Juvenile arrest records from age 14 were compiled from rap sheet and police investigation reports provided by the Juvenile Aid Division of the Philadelphia Police Department. Adult arrest records up to age 26 came from the Municipal and Common Pleas Courts of Philadelphia. Data on education, employment, health, and some self reported variables on criminal activities, etc. were collected in a 1988 follow-up survey interview.

Instead of converting the data into an annual panel, as is usually done, we construct a monthly panel of arrests and employment activities, thereby obtaining a more detailed panel history. We think this difference is important. If we use annual data, almost everybody works positive hours. But with monthly data, we observe both short and long unemployment spells. A limitation of the data is that only the starting and the ending wage in a job are available. We interpolate the wage in between linearly.

In our data, only employment spells of 6 months or more, and only unemployment spells of 2 months or more are recorded. Hence, if we just estimate the employment dynamics directly from the data, our results will be biased. Using the steps below, we try to recover the missing employment and unemployment spells through the model of employment dynamics.

1) Missing data after an employment spell must contain an immediate unemployment spell of less than 2 months. Had the unemployment spell not been immediate, it would have been recorded as employment. Had it been longer than or equal to 2 months, it would have been recorded as an unemployment spell. Similarly, missing data after an unemployment spell must contain an immediate employment spell of less than 6 months. If the blank data after an employment spell is of one period, then we infer it has to be an unemployment spell. If it is of 2 periods, it must be an unemployment spell followed by the employment spell of a month. If a blank data after an unemployment spell is of one month, it must be an employment spell. If it is more than one month, we cannot say.

2) In all other cases, we use the probability of employment in the entering state to run the employment/unemployment probabilities forward. To be consistent with the data, the augmented employment spells are restricted not to exceed 6 months and unemployment spells are restricted not to exceed 2 months.

Some sample statistics are shown in Table 1. Note that only roughly half of the sample is white, and more than a third were gang members before they were 18 years old. This suggests that the sample is not nationally representative.

Figures 1 to 4 depict how arrests, unemployment and wages are related to age. From Figure 1, we see that the arrest rate of males peaks at 17.[3] In Figure 2, we plot the age arrest profiles for individuals with different past criminal records. Note that these profiles shift up with past arrests. This is consistent with the fact that repeat offenders account for a large proportion of arrests. Figure 3 depicts mean unemployment, which falls with age. In Figure 4, we plot the mean and median age wage profiles. The mean wage profile is far above the median. This is typical of wage data, since wages are known to have many outliers.

3 Model Specification

The model consists of the following structural elements: the choice set, state space, and preferences. The choice set of an individual is simply whether or not to commit a crime. However, we allow for the fact that when the individual commits a crime and gets caught, it can affect his high school graduation, employment, wages, etc., which in turn are inputs in his choice of whether or not to commit a crime. While we do not explicitly incorporate choice of high school graduation and employment, we allow for these effects in a re-

[3]It is interesting that in the original work of Quetelet more than 15 years ago, the age crime profile peaked at age 24 or thereabouts. This is reported in Leung (1994).

duced form manner. Past criminal history is allowed to affect the probability of high school graduation, employment as well as the distribution of wage draws. We also allow for the possibility that agents differ in ways that are inherently unobservable. We allow for 4 types of agents, which arise from combinations of two crime and two unemployment types. Parameters for the different types could differ. We chose to allow for crime and unemployment types instead of just crime types. Had we not done so, the coefficient of past arrests on the probability of unemployment in our structural model would tend to be biased upwards if high crime types also tend to be high unemployment types.[4] This would make dynamic deterrence play an excessively large role.

Let $\mathbf{s}_t \in S_t$ be the state space vector for period t, where t is the age of the individual (in months). Since the panel data starts at age 14, an individual who is 14 years old is normalized to have $t = 0$. This state space expands as an agent reaches maturity to reflect the additional choices made by an adult as opposed to a child.[5] Before age 16, $\mathbf{s}_t = (t, n_t, i_{h,t})$, where $i_{h,t} = 1$ if in the data the individual attends high school in period t and 0 otherwise. In addition, n_t is the criminal record of the individual in period t, while t is the age of the individual (in months). At or after the age 16, the individual starts working, and the state space vector is augmented by labor market

[4] In fact, this is exactly what we find. The probability of being a low unemployment type is 0.6143 for a non-criminal type, and is 0.3128 for a criminal type.

[5] Throughout, we keep the model deliberately simple both because of computational reasons and data limitations.

information. Hence, $\mathbf{s}_t = (t, n_t, i_{h,t}, i_{u,t}, W_t)$ between the ages of 16 and 18, where $i_{u,t} = 1$ if the individual is not employed in period t and 0 otherwise and W_t is the real wage rate. From age 18 onwards, the state space is further augmented to be $\mathbf{s}_t = (t, n_t, i_{h,t}, i_{hg}, i_{u,t}, W_t)$ to reflect whether or not the individual graduates from high school. We set $i_{hg} = 1$ if he graduates from high school, and 0 otherwise.

Past arrest records depreciate at rate δ_t in period t. This allows arrests in the distant past, or when committed as juveniles to be treated differently from recent arrests committed by adults. We expect δ_t to be less than unity since empirical evidence (see Grogger (1995), Kling (1999)) suggests that past arrests have only temporary effects on labor market outcomes. It also allows for the fact that juvenile records are sealed at adulthood. If the agent commits a crime and gets caught, then his criminal record is augmented by unity. That is, $n_{t+1} = \delta_t n_t + 1$. Otherwise, $n_{t+1} = \delta_t n_t$. We assume the depreciation rate to be constant at δ except at age 18, i.e., at $t = 48$, $\delta_t = \widehat{\delta}$. We expect $\widehat{\delta}$ to be less than δ since juvenile records are sealed at adulthood.

High school age unemployment is not exogenous and the probability of being unemployed at age 16, i.e. $t = 24$, has the following logit form:

$$P_{u,24} = \exp(\phi_{u,24})/[1 + \exp(\phi_{u,24})] \tag{1}$$

where

$$\phi_{u,24} = h_0 + h_1 n_{24} \tag{2}$$

Note that we allow criminal records to affect the probability of unemployment.

After the first month of age 16, i.e., for $t > 24$, the individual experiences job transitions. The probability of staying unemployed depends both on his past criminal history and the employment status. We let,

$$P_{u,t+1} = \exp(\phi_{u,t+1})/[1 + \exp(\phi_{u,t+1})] \tag{3}$$

where

$$\begin{aligned} \phi_{u,t+1} &= b_{00}I(age < 18) + b_{01}I(age \geq 18) + b_1(t + 1 - 24) + b_2 i_{hg} + b_3 n_{t+1} \\ &\quad + [b_{40}I(age < 18) + b_{41}I(age \geq 18)]i_{u,t}. \end{aligned} \tag{4}$$

$I(age < 18)$ is an indicator function which equals 1 when the agent is below 18, and 0 otherwise. All the other indicator functions are analogously defined. The above specification allows a jump in unemployment probabilities and in the persistence of unemployment at 18. Also, $i_{hg} = 0$ denotes that the agent did not graduate from high school and $i_{hg} = 1$ denotes that he did.

Whether an individual attends school or not is taken to be exogenous. However, at age 18, we assume the individual either graduates or does not graduate from high school. He graduates from high school with probability

$$P_{hg} = \exp(\phi_{hg})/[1 + \exp(\phi_{hg})] \tag{5}$$

where $\phi_{hg} = g_0 + g_1 n_{48}$.[6] The starting wage of the individual, in his first month of employment, which occurs at some $t > 24$, follows the log normal

[6] This separates high school graduation which is observable and hence impacts employment opportunities, and high school attendance which affects criminal choices at the time

distribution

$$\ln(W_t) \;\sim\; N(\mu_b(n_t), \sigma_b) \tag{6}$$

where

$$\mu_b(n_t) = \mu_{b0} + \mu_{b1} n_t. \tag{7}$$

Furthermore, the wage growth for the individual on a job is assumed to be log normally distributed such that

$$\ln(W_t) - \ln(W_{t-1}) \;\sim\; N(\mu_{gt}(.), \sigma_g), \tag{8}$$

where

$$\begin{aligned} \mu_{gt}(.) \quad = \quad & \kappa_1 I(16 < age \leq 19) + \kappa_2 I(20 < age \leq 23) + \kappa_3 I(24 < age) \\ & + [\kappa_4 I(16 < age \leq 23)\,(t-36) + \kappa_5 I(24 < age)\,(t-120)] \\ & + \kappa_6 n_t. \end{aligned} \tag{9}$$

This form allows for changes in intercepts at age 20, 24 and change in slope at age 24.

Now we turn to preferences. The utility obtained from not committing a crime is interpreted as the static single period payoff from not being arrested. This depends on factors such as age, unemployment status, wages as well as intangibles arising from being a high school graduate and past criminal

but which is not verifiable in the future and hence, does not impact employment opportunities. We do not allow past attendance to affect the probability of graduation from high school since doing so would involve calculating the dynamic programming problem for different attendance records.

records. We assume that the deterministic part of the per period utility from not committing a crime takes the following form.

$$
\begin{aligned}
u_N(\mathbf{s}_t) \;=\;& c_{01}I(age < 18) + c_{02}I(age \geq 18) \\
& + [c_{03}I(17 \leq age < 18) + c_{04}I(17 \leq age)](t - 36) + c_1 i_{h,t} \\
& + [c_{u1}i_{u,t} + [c_{m1}I_m + c_{h1}I_h]\,(1 - i_{u,t})]I(age < 18) \\
& + [c_{u2}i_{u,t} + [c_{m2}I_m + c_{h2}I_h]\,(1 - i_{u,t})]I(age \geq 18) \\
& + c_5 i_{hg} + c_6(n_t)^{\alpha} \qquad (10)
\end{aligned}
$$

where $I_j,\ j = m, h$ are the indicator functions for medium and high wage groups given they are employed.[7] We drop the low wage dummy to avert the dummy variable trap. There are also indicator functions for being below age 18, $I(age < 18)$, as well as being between 17 and 18, and more than 18. Our formulation allows for differential slopes for u_N as a function of t between age 17 and 18, and 18 onwards, while setting the slope below age 17 to be zero. It also permits a jump in u_N at age 18. We introduce this differentiation both to reflect the differences in treatment of juveniles and adults under the law and to allow us to fit the age arrest profiles which peak at age 17. In general, the criminal justice system treats individuals under and over age 18 quite differently. α allows for convexity or concavity in the effect of criminal history.

[7] Low wage group individuals are those with real wages below \$5. Medium wage group individuals are those with real wages between \$5 and \$8. High wage group individuals are those with real wages greater than or equal to \$8.

As is well known from the discrete choice econometric literature, in a static model of criminal choice, we cannot separately identify the utility of not committing a crime and the utility of committing a crime just on the basis of data on criminal choice. On the other hand, in a dynamic setting where individuals look forward to the consequences of their actions, the future utilities of both committing and not committing a crime enter into their calculations. Through this, we can identify the coefficients on the utilities obtained from the two alternatives. However, it is unreasonable to expect identification of u_C and u_N to be tight since it comes from a relatively complicated model.

We parameterize the deterministic utility of committing a crime very simply as follows:

$$u_C(\mathbf{s}_t) = \left[d_{u1} I(age < 18) + d_{u2} I(age \geq 18)\right] i_{u,t} + d_1 \sqrt{n_t}. \tag{11}$$

We interpret u_C as the direct benefit of committing a crime. Our parameterization has the deterministic utility of committing a crime depending only on the past criminal record when the agent is employed, while it allows for different deterministic utilities before and after age 18 when the agent is unemployed. Notice that the constant term and the age coefficient are not included due to the identification problem mentioned above. We allow the parameters of u_N and u_C in equation (10) and (11) to differ between the two crime types. In addition, we allow the parameters for the unemployment probabilities to differ between the two unemployment types.

The agent's objective is to maximize the expected present value of lifetime utility. To close the model, we assume that in the terminal period, $T = 228$, at age 33,[8] he receives a payoff of $V_T(n_T)$ which depends on his past arrests. The criminal history in the terminal period is summarized by the index n_T. The final period value function is approximated to take a simple linear form: $V_T(n_T) = \gamma n_T$.

The only choice we model is whether to commit a crime or not. The value of not committing a crime is

$$V_{Nt}(\mathbf{s}_t) = u_N(\mathbf{s}_t) + \beta E\left[V_{t+1}(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct} = 0\right] + \epsilon_{Nt} \tag{12}$$

where $i_{Ct} = 0$ denotes that the agent was not arrested in t, and $i_{Ct} = 1$ denotes that he was. $\mathbf{s}_t \in S_t$ is the state space vector at time t. The value function tomorrow depends on the state variable tomorrow but the generating process of the state variable depends on the state variable today and its augmentation via arrests. Hence we condition on s_t and on whether an arrest occurs. Furthermore, as there is randomness in the state variables we take expectations. $u_{Nt}(\mathbf{s}_t)$ is the one period utility of not committing a crime, ϵ_{Nt} is the utility shock of not committing a crime. Of course, in the event of not committing a crime, $n_{t+1} = \delta n_t$.

[8] We chose this somewhat early age, because we only have data until age 26. Since the parameters estimated are based on data from age 14 to 26, it makes little sense to solve the model too far out.

The value of committing a crime is

$$
\begin{aligned}
V_{Ct}(\mathbf{s}_t) &= u_C(\mathbf{s}_t) + P_A \beta E[V_{t+1}(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct} = 1] \\
&\quad + [1 - P_A]\,[u_N(s_t) + \beta E V_{t+1}(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct} = 0] + \epsilon_{Ct} \qquad (13)
\end{aligned}
$$

where ϵ_{Ct} is the utility shock of committing a crime. ϵ_{Nt}, ϵ_{Ct} are assumed to be i.i.d. and to follow an extreme valued distribution. P_A is the probability of getting caught after committing a crime and $u_C(\mathbf{s}_t)$ is the one period utility of committing a crime. The slightly unconventional form of V_{Ct} comes from our interpretation of $u_c(s_t)$ in equation (13) as the direct benefits of committing a crime. When not caught, the criminal can enjoy a "normal life", i.e. obtain $u_N(\mathbf{s}_t)$. However, when caught, he forgoes this payoff for a single period, and this is the cost of being caught. We can do no more since we do not have data on sentencing. Since we only have data on arrests and not crimes, the probability of catching the offender is not identified, and we therefore set it to be $P_A = 0.16$. Clearance rates are the ratio of arrests to reported crimes. In 1991 they ranges from about 67% for murder and non-negligent manslaughter and 13.5% for burglary.[9] We chose 16% as a reasonable average.[10]

An agent enters the period with his known state variables, s_t. Draws of the utility shocks $\epsilon_t = (\epsilon_{Nt}, \epsilon_{Ct})$ occur, and the decision to commit or

[9] These numbers are based upon Table 1 in Ehrlich (1996).

[10] This specification ignores the possible endogeneity of the probability of getting caught which could vary with the seriousness of the crime. Those aspects are pointed out by Tauchen et. al. (1994) and others. Lochner (2000) estimates the manner in which beliefs about being apprehended are affected by past arrests and other information.

not to commit a crime is made. If the agent is arrested, after committing a crime, then his state variables for the next period are drawn from the relevant distributions conditional on his increased arrest record. If he is not caught or does not commit a crime, then the state variables are drawn from the distribution conditional on the depreciated arrest record. Hence, the value function is defined as follows:

$$V_t(\mathbf{s}_t) = Max\{V_{Nt}(\mathbf{s}_t), V_{Ct}(\mathbf{s}_t)\}. \tag{14}$$

In the section on Bellman equations in the Appendix, we elucidate on the value functions of the individuals at various ages.

We now turn to the construction of the likelihood function. We observe whether or not an agent is arrested, employment and wage draws, high school graduation. The parameters of the likelihood function are chosen to maximize the probability of generating the actual data. Equation (14) drives the construction of the arrest increment of the likelihood function. Since our data is on arrests, and not on crimes committed, there are only two possibilities. Either the agent commits a crime and gets arrested, or he does not get arrested. The latter outcome includes his not committing a crime and committing a crime and not getting arrested. The probability an agent commits a crime equals the probability that $\epsilon_{Ct} - \epsilon_{Nt} > \overline{V}_{Nt}(\mathbf{s}_t) - \overline{V}_{Ct}(\mathbf{s}_t)$, where $\overline{V}_{Nt}(\mathbf{s}_t)$ and $\overline{V}_{Ct}(\mathbf{s}_t)$ are the deterministic components of the values of committing and not committing a crime. To obtain these, we need to solve the dynamic programming problem at each t. Details of how this is done are found in the Appendix. The probability of arrest is the product of P_A and the probability of committing a crime.

In addition to the arrest increments, the likelihood of individual i at period t has employment, starting wage, wage growth and high school graduation increments. The employment and the high school graduation increments of the likelihood take a simple logit form described in equations (3) and (5), respectively. The starting wage and the wage growth increments of the likelihood take a normal form as described in equations (6) and (8), respectively. Details of how these increments are constructed can be found in the Appendix. The likelihood function for a particular type is a product of the likelihood increments for that type of each agent in each time period. The likelihood is the weighted sum of the likelihood function of each type, and the weights are the probability of each type. Finally, the parameters are chosen to maximize the likelihood.

The unit of time in our paper is months. Since no individuals have multiple arrests in our data, we can abstract from multiple crimes and assume that the individuals only have two choices, either to commit a crime or not to do so. Incarceration is interpreted as being unemployed and at the same time, unable to commit any crimes.[11]

[11] Because of this, the effect of past crimes on unemployment will be biased upwards and the effect on crimes committed by the unemployed biased downwards. Since in our results, past crimes raise unemployment and past unemployment raises crimes, the former might change signs after the bias is removed, but the latter would not. However, in any case, we believe the bias to be small since the probability of incarceration in the data is small (see Fig. 5).

We also include some unobserved heterogeneities. We take a minimalist stand and assume 2 criminal types and 2 unemployment types. We have crime type 1 and 2 and unemployment type 1 and 2. The agent's type is modeled as a random effect, and the probability of an agent's type is estimated so as to maximize the likelihood function. Crime type 2 and unemployment type 2 turn out to be the high crime/high unemployment types. As in other estimation exercises such as Keane and Wolpin (1997) or Eckstein and Wolpin (1999), we do not include any observed heterogeneity. As a check, we later look at the regression relationship between the unobserved heterogeneities and the observed differences in individual characteristics and conclude that the unobserved heterogeneities estimated from the data are loosely related to observed differences in individual characteristics.

In general, solving and estimating such dynamic discrete choice models, is computationally demanding. Recall that in order to solve for the Bellman equation described in more detail in the Appendix, we needed to solve for the expected values, $E[V_{t+1}(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct}]$. To derive these expected value functions, we needed to integrate over the shocks ϵ_{Nt} and ϵ_{Ct} and over the wage and employment shocks. This integration had to be done for each point in the state space, $\mathbf{s}_t$, at each period t. On top of this, the above Dynamic Programming problem had to be solved once at each likelihood evaluation, when we assume no heterogeneity, and several times when we introduce some unobserved heterogeneities. As a result, the programming and computation were non trivial. Details on model estimation are to be found in the Appendix.[12]

4 Estimation Results

Parameter estimates are presented in Table 2. We discuss only those that are significant. Many parameters are not significant, which is not unusual in such dynamic models. We allow the parameters for the two crime types to differ in $u_N(.)$ and $u_C(.)$. Their estimates are shown in the beginning of table 2. However, they are not precisely estimated.

The discount factor β is close to 1. The monthly depreciation rate $(1-\delta)$ is about 2% per month, which amounts to an annual depreciation rate of about 21%. Our estimates are consistent with past work such as Grogger (1995), Kling (1999), who have pointed out that the effect of past criminal history on current variables such as employment and wages is temporary. The depreciation rate of the juvenile criminal record at age 18, or $(1-\widehat{\delta})$ is about 26%. This, combined with the annual depreciation rate of 21% implies that juvenile crime records have a relatively small effect on the adult behavior.

The probability of high school graduation is also allowed to differ in the intercept according to criminal type, i.e. g_0 can differ between crime types, though g_1 is common for both types. Note that criminal history has a sig-

[12]The FORTRAN programs used to implement the estimation is available upon request.

nificant negative effect on high school graduation.

The initial unemployment probability at age 16 is allowed to differ across unemployment types. Not surprisingly, the high unemployment type (type 2) has a significantly positive intercept.

Unemployment probabilities are also allowed to differ across employment types. In particular, the intercept before and after 18 as well as the effect of age is allowed to differ. Type 1 is estimated to be the low unemployment type since the intercepts are smaller, and unemployment falls with age at a greater rate for type 1. High school graduation reduces the probability of unemployment and criminal history raises it. Unemployment is also persistent as revealed by b_{40} and b_{41} being positive. Moreover, persistence is greater after age 18.

The wage growth equation is common across all types. This is dictated by the deficiency of the wage data as explained previously. Notable is that the criminal history has a small but significantly negative effect on wage growth. Similarly, criminal history has a small positive effect on the starting wage, though it is insignificant. This might have to do with criminals requiring a higher wage to consider employment.

The non-criminal type (type 1) has higher probability of being of a low unemployment type (type 1) as evidenced by $\pi_{11} = 0.6143 > \pi_{21} = 0.3128$. π_{11} is the conditional probability of being a low unemployment type given that the agent is the low crime type. π_{21} is the conditional probability of being a low unemployment type given that the agent is the high crime type. This suggests correlation between crime and unemployment.

Our results show that the criminally at risk type is about 33% of the sample. Moreover, these types clearly affect behavior, as is evident in the difference in their crime, unemployment and wage profiles depicted in Figures $5-7$. In Table 3, we report the results of a logit type regression that relates the odds of the individual being of crime type 1 with several observed characteristics. That is, we estimated the following equation.

$$\ln[P_{C1,i}/(1-P_{C1,i})] = \beta_0 + \beta_1 X_i + \nu_i. \tag{15}$$

where $P_{C1,i}$ is the probability individual i is a low crime type. One minus the number in the last column of table 3 provides an idea of the significance of the coefficient. For example, race with a coefficient of 0.167 is significant only at the 83rd percentile. The most significant of these coefficients are the ones on parents arrested, gang member, race. Being a gang member of having parents who have been arrested reduce the probability of being a non-criminal type, as might be expected. Whites have a slightly higher probability of being the non-criminal type. This suggests that unobserved heterogeneity is loosely correlated to observed heterogeneities, but much remains unexplained.

5 Simulation Exercises

This section has two distinct components. The first deals with how well the data and the simulated model track each other. The second deals with the effects of some policy experiments.

5.1 Generated and Actual Data

In Figures 1-8, we compare the simulation results with the data. The model fits well with regard to the overall age arrest profile (Figure 1) and the age unemployment profile (Figure 3). It fits the age *median* wage profile much better than the age *mean* wage profile (Figure 4). This is quite natural since outliers affect the mean wage more than the median, and our assumption of a normal distribution for log wages limits outliers. Figure 8 depicts the simulated age arrest profiles with different past criminal records. Notice that the simulated arrest profiles for individuals with more past arrest records lie above those with fewer ones. This corresponds to the actual profiles, as depicted in Figure 2. That is, the profiles indicate that repeat offenders commit more crimes than others. Thus, greater criminal activity by repeat offenders comes naturally from our setup.

Figure 5 plots the simulated age arrest profiles of the four types separately. Notice that there are large differences in arrest rates among the types. In particular, the arrest rate of the at-risk youths (criminal types with both low and high unemployment) seem to be 2 to 4 times as high as that of the others. This is also consistent with repeat offenders committing most crimes. Also, notice from Figure 5, that it is the arrest rate of the criminal types (criminal type 2) that shows a rapid decline after age 18.[13] This is consistent with the

[13]Note however that crimes committed need not track arrests as older more criminally

arrest rate decreasing with age after 18. Figures $6-7$ plot the simulated unemployment and wage profiles for the 4 different types. The types do look different: when it comes to crime, the crime types matter, and when it comes to labor market outcomes, the unemployment types matter. Just as the difference in age arrest profiles is greatest for the criminal types in Figure 5, the difference between the age unemployment and age wage profiles are largest for the two unemployment types in Figures 6 and 7.

5.2 Counter-factuals

Next, we conduct some counter-factual simulations to better understand some policy issues of interest.

5.2.1 Eliminating the Effect of Arrest Records on Labor Market Outcomes.

First, we look at what the outcomes would have been if criminal history did not affect employment outcomes, and individuals did not expect them to do so. This involves setting the coefficients for the past criminal history in the unemployment probability, (b_3 in equation (4)) and in the wage equations (μ_{b1} in equation (7) and κ_6 in equation (9)) to be zero. In addition, b_3, μ_{b1} and κ_6 are also set to zero in the dynamic program.

In Figure 9, the line labeled "ratio (anticipated)" plots the ratio of the age arrest profiles in this scenario relative to the simulations using estimated parameters. Figures 10 and 11 plot the analogous age unemployment and the

active agents may become better at avoiding arrest.

age wage profile ratios. We can see that this exercise tends to reduce the age unemployment profile. This makes sense since unemployment probability is reduced with the elimination of past arrest records. The age arrest profile ratio exceeds unity in this counter-factual. When past arrests have no effect on unemployment, and this is anticipated, dynamic deterrence is nonexistent and agents commit more crimes and hence get arrested more often. The age wage profile ratio falls below unity initially. This is because arrests increase the starting wage, i.e., $b_{13} > 0$. However, since wage growth is higher when arrests are lower, i.e. $\kappa_6 < 0$ the ratio is increasing with age.

In this experiment, two things are happening. Forward-looking behavior and state dependence as related to the labor market are being removed. By having arrests not affect future labor market outcomes while agents think they do, we could artificially separate out the role of state dependence though labor market effects in this experiment.[14] Namely, we only set b_3, μ_{b1} and κ_6 to be zero in simulating the data, but not in the dynamic program. The ratio of the simulations when state dependence alone is removed, and the outcome from simulation using the estimated parameters is referred to as the ratio (unanticipated) in Figures $9-11$. The ratio (anticipated) refers to what occurs when b_3, μ_{b1} and κ_6 are also set to be zero in the dynamic program. Note that state dependence has small effects on the age arrest profile ratio as depicted in Figure 9. It remains close to unity throughout.

[14]Note that this is a pure counter-factual, since it assumes irrationality on the part of the individual.

This suggests that state dependence alone is not affecting crimes committed (which are proportional to arrests) by much. In Figure 10, which depicts the ratio of age unemployment profiles, note that the ratio (anticipated) lies above the ratio (unanticipated). This occurs through the effect on graduation from high school. When agents anticipate that there will be no labor market effects of their criminal behavior, they commit more crimes, which reduces the probability of graduation from high school. This in turn, raises the probability of unemployment. Figure 11 depicts the ratio of age wage profiles. Since greater unemployment probabilities pull down wages through reduced wage growth, the age wage profile ratio (anticipated) lies below the ratio (unanticipated).

5.2.2 Policies that Change Unemployment

As is well known, crime rates have fallen in the past decade but show signs of leveling off. An important question in the policy arena is the extent to which this is due to the booming economy of the period. To get a partial handle on this, consider another policy experiment, where, given the current state variable, we reduce the one period ahead unemployment probability after the first month of age 18, by 5%. That is,

$$P_{u,t+1} = 0.95 \times \exp(\phi_{u,t+1})/[1 + \exp(\phi_{u,t+1})] \tag{16}$$

The results on unemployment and arrest rate ratios are plotted in Figures $12-13$, respectively. Because of the persistence of unemployment, the unemployment ratios (unanticipated) as well as (anticipated) are reduced by more than 5%. When the reduction is unanticipated, the employed commit fewer crimes, so the induced effect on the arrest ratio pulls it below unity.[15] In contrast to this, if the policy is anticipated, the reduction in unemployment transition probability increases crime. Again this is due to the strong dynamic deterrence effect since it is the prospect of future unemployment that deters crime. This exercise thus makes it hard to argue that the boom in the $90's$ alone can be seen as responsible for the reduction in crime. However, to the extent that this boom, due to its length and depth managed to bring those at the very bottom into the labor force, our approach may be under-estimating the effect of crime reduction. Bringing such agents into the labor force, thereby providing a dynamic deterrence effect where none existed before, could well reduce crime.[16]

One way to obtain larger reductions in the arrest ratio is to consider the effects of an anticipated boom followed by a bust. Given the current state variable, we reduce the one period ahead unemployment probability by 5%, after the first month of age 20, for 2 years.[17] After this the unemployment transition probability is assumed to increase by 5%, compared to the original one for 2 years. The effects on crime ratios are depicted in Figure 14. Behav-

[15] Feedback from arrest to unemployment explains why the unemployment ratio (unanticipated) lies below the anticipated one.

[16] The effect on wages is small, less than 1.2%. Over time, the fall in unemployment tends to raise wages in both the (unanticipated) as well as (anticipated) scenarios.

[17] We choose the age of 20 so that the change in the behavior of adults as well as juveniles anticipating this boom and then slump can be illustrated. It makes no difference to the earlier simulations if the same age (of 20) is used there.

ior is affected even before the onset of the reduction of the unemployment transition probability because individuals anticipate the reduction in future threat of unemployment. Before age 20, expectations of a good labor market make crime and hence arrests rise. As a result, the arrest rate ratio rises above unity. Once the boom begins, expectations of a slump reduce crime, and hence, the arrest rate ratio drops below unity. As the slump occurs, expectations of normal times raise crime and the arrest ratio. The policy maker, failing to understand the deterrence aspect of the unemployment effect, could erroneously conclude that low unemployment is the cure for crime. However, a permanent reduction in unemployment raises crime! Note that if this anticipated boom-slump were the reason for the observed decline in crime seen in the 90's, we should expect an increase once the slump comes.

5.2.3 Other Policies: Enforcement, Erasing Juvenile Records and Depreciation of Criminal History

What about the effect of greater enforcement? This policy, joint with harsher sentencing, has been the standard approach to combating crime. In our next experiment, we increase the anticipated probability of being caught by 10%. As shown in Figure 15, the effect is to raise the arrest ratio for the young and reduce it for adults. This occurs as the young face weaker penalties as their criminal history depreciates at age 18, and they inter-temporally substitute towards crime.[18]

[18] As expected, the increase in arrests translates into greater unemployment.

We also look at the effect of not sealing juvenile records, i.e., making $\widehat{\delta} = \delta$.[19] The arrest ratio is depicted in Figure 16. As expected, the young commit fewer crimes, realizing that their criminal record is more permanent. On the other hand, adults commit more crimes, as they cannot get away from their juvenile records.

Finally we look at what happens if we increase both δ and $\widehat{\delta}$ by 0.1%. This corresponds to decreasing the depreciation rate of past criminal histories. There are large differences between countries in the extent to which an individual's past haunts him. In Japan for example, a criminal record is relatively permanent. In the U.S. on the other hand, criminal history is much easier to disguise. In fact, only in recent years have there been laws such as Megan's Law, on informing neighbors of sex offenders who move in. The results are shown in Figure 17. We notice that even a small decrease in depreciation rate generates a large decrease in the crime rate. The work of Glaeser, Sacerdote and Scheinkman (1996), or Williams and Sickles (2000) on social human capital suggests that such effects could be important even though they do not directly look at depreciation as we do. Casual observations across different countries and regions reinforce this conclusion. Crime tends to be lower in countries where people live in closely-knit communities. In these communities, even though there may be few legal consequences of

[19] Of course, if juvenile records were completely eradicated, and there were no other effects such as differences in criminal and other human capital among juvenile offenders and others, then $\widehat{\delta}$ should be zero. This is why making $\widehat{\delta} = \delta$ only roughly corresponds to the opening of the juvenile records.

offenses, past misconduct of members is not forgotten. The long memory of community members works as a strong deterrent against crime. There are also other aspects of social effects. Calvo-Armengol and Zenou (2003) analyze interactions within the criminal networks and Silverman (2003) models reputation effect and "street culture".

In sum, our results emphasize the role of future unemployment as an important factor holding people back from committing crimes. Even though much attention has been paid to the relationship between labor market outcomes and crime, we think this aspect has been neglected. When researchers consider the effect of unemployment and wages on crime, they mainly focus on the direct state dependence effect on criminal behavior. Instruments and other methods are used to avoid endogeneity problems due to state dependence or heterogeneity. However, correcting endogeneity in this manner does not give all the structural parameters of interest, and hence only incompletely addresses the effect of government policy since expectational effects cannot be incorporated. Our results agree with many past results insofar as unemployment and wages have small direct effects on crime. What is new in our work is that despite such small direct effects, government employment and wage policies could change criminal behavior significantly, mainly through changing peoples' anticipations about their future.

6 Concluding Remarks

What are the implications of our work for the conduct of public policy towards crime? Our structural dynamic approach provides a unified understanding of a number of findings in the traditional literature. Kahan (1995) claims that effective anti-crime policies are those that change people's anticipation of future punishments. We agree and argue that these future punishments seem to come from the labor market! There have been several papers showing that early intervention programs such as the Job Corps, The Perry Preschool Program, The Syracuse University Family Development Plan and the Quantum Opportunity Program are very effective in reducing crime[20], see Lochner (1999) for a summary of such results. This is exactly what would be expected from our model, since anticipated later intervention allows criminals to look forward to negating the consequences of their actions. Early intervention has no such adverse effect. This suggests that early prevention is more effective than redemption.

As is the case with all structural estimation results, we need to interpret the above results with caution, and more work needs to be done to assess the robustness of the results with respect to various alternative model specifications. For example, we assumed that the individuals only choose between committing a crime and not committing a crime and we treated all crimes

[20]The Perry Preschool Program for disadvantaged minority children reduced arrests through age 27 by 50%.

as being the same. Obviously, they are not. Not only does the criminal justice system pursue offenders of different crimes with different intensities, and punish them with different degrees of severity, but society treats different types of offenders very differently. Hence, both state dependence and deterrence should be different depending on the types of crimes committed. Such issues could be addressed in the future, with better datasets.

7 References

Bearse, P. M.,"On the Age Distribution of Arrests and Crime." mimeo, University of Tennessee, 1997.

Becker G. S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 73 (March-April 1968), $169-217$.

Burdett, K., R. Lagos and R. Wright, "An On-the-Job Search Model of Crime, Inequality and Unemployment." PIER Working Paper No. $03-030$, Department of Economics, University of Pennsylvania, September 2004.

Calvo-Armengol, A. and Y. Zenou, "Social Networks and Crime Decisions: The Role of Social Structure in Facilitating Delinquent Behavior", mimeo, University of Southhampton, 2003.

Eckstein, Z. and K. I.,Wolpin, "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica* 67

(November 1999), 1295 – 1339.

Ehrlich, I., "Crime, Punishment, and the Market for Offenses," *Journal of Economic Perspectives* (Winter 1996) 10, 43 – 67.

Figlio, R. M., P. E. Tracy and M. E. Wolfgang, "Delinquency in a Birth Cohort II: Philadelphia, 1958 – 1988, " Sellin Center for Studies in Criminology and Law, Wharton School, University of Pennsylvania, Inter-University Consortium for Political and Social Research, Ann-Arbor, Michigan, 1994.

Flinn, C., "Dynamic Models of Criminal Careers," in A. Blumstein et. al., Ed. *Criminal Careers and "Career Criminals,"* (Washington D.C.: National Academy Press, 1986).

Freeman, R. B., "Why Do So Many Young American Men Commit Crime and What Might We Do About It?," *Journal of Economic Perspectives* 10 (Winter 1996) 25 – 42.

Glaeser, E. L., B. Sacerdote and J. A. Scheinkman, "Crime and Social Interaction," *Quarterly Journal of Economics* (May 1996), 507 – 548.

Grogger, J., "The Effects of Arrests on the Employment and Earnings of Young Men," *Quarterly Journal of Economics* 110 (February 1995) 51 – 71.

Grogger, J., “Market Wages and Youth Crime,” *Journal of Labor Economics* 16 (October 1998), 756 – 791.

Huang, C. C., D. Laing and P. Wang, “Crime and Poverty: A Search Theoretic Approach”, mimeo, Vanderbilt University, 2003.

Imrohoroglu, A., A. Merlo and P. Rupert, “What Accounts for the Decline in Crime?”, mimeo, University of Pennsylvania, 2001.

Judd K., *Numerical Methods in Economics* (Cambridge:MIT Press, 1998).

Kahan, D. M. “Social Meaning and the Economic Analysis of Crime,” *Journal of Legal Studies* 27 (June 1998), 609 – 622.

Keane, M. P. and K. I. Wolpin, “The Career Decisions of Young Men,” *Journal of Political Economy* 105 (June 1997) 473 – 522.

Kling, J. R., “The Effect of Prison Sentence Length on the Subsequent Employment and Earnings of Criminal Defendants,” mimeo, Princeton University, 2002.

Leung, S. F., (1994) “An Economic Analysis of the Age-Crime Profile,” *Journal of Economic Dynamics and Control* 18 (March 1994), 481 – 497.

Levitt, S., (1996) “The Effect of Prison Population on Crime Rates: Evidence from Prison Overcrowding Litigation,” *Quarterly Journal of Economics* 111 (May 1996), 319 – 351.

Levitt, S., "Juvenile Crime and Punishment," *Journal of Political Economy*, 106, (December 1998) 1156 – 1185.

Lochner, L.,"Education, Work, and Crime: Theory and Evidence," Working Paper No. 465, Department of Economics, University of Rochester, September 1999.

Lochner, L., "Individual Perceptions of the Criminal Justice System," NBER working paper No. 9474, (February 2002).

Rust, J., "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica* 55 (September 1997), 999 – 1033.

Rust, J., "Using Randomization to Break the Curse of Dimensionality," *Econometrica* 65 (May 1997) 487 – 516.

Silverman, D., "Street Crime and Street Culture," mimeo, University of Michigan, 2003.

Tauchen, H., A. D. Witte and H. Griesinger, "Criminal Deterrence: Revisiting the Issue with a Birth Cohort," *Review of Economics and Statistics* 76 (August 1994), 399 – 412.

Williams, J. and R. C. Sickles, "An Inter-temporal Model of Rational Criminal Choice," mimeo, Rice University, 1997.

Williams, J. and R. C. Sickles, "An Analysis of the Crime as Work Model:

Evidence from the 1958 Philadelphia Birth Cohort Study", Journal of Human Resources 37 (Summer 2002), 479 – 569

A. D. Witte and H. Tauchen, "Work and Crime: An Exploration Using Panel Data," *Public Finance* 49 (Supplement 1994), 155 – 167

8 Appendix

8.1 Bellman Equations

The value function of the individual from age 14 until age 16 of not committing a crime is

$$V_{Nt}(\mathbf{s}_t) = u_N(\mathbf{s}_t) + \beta E[V_{t+1}(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct} = 0] + \epsilon_{Nt}. \tag{A1}$$

where

$$\mathbf{s}_t = (t, n_t, i_{h,t}), n_{t+1} = \delta n_t. \tag{A2}$$

The value function of committing a crime is

$$\begin{aligned} V_{Ct}(\mathbf{s}_t) &= u_C(\mathbf{s}_t) + P_A \beta E[V_{t+1}(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct} = 1] \\ &\quad + [1 - P_A] \left\{ u_N(s_t) + \beta E \left[V_{t+1}(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct} = 0 \right] \right\} + \epsilon_{Ct} \end{aligned} \tag{A3}$$

From here on, we will only elucidate on the value function of not committing crimes. The value function of committing a crime is defined analogously to that shown above.

The value function of not committing a crime at or after age 16 but before age 18 incorporates the possibility of employment:

$$\begin{aligned} V_{Nt}(\mathbf{s}_t) &= u_N(\mathbf{s}_t) \\ &\quad + [1 - P_{u,t+1}] \beta E[V_{t+1}(t, \delta n_t, i_{u,t+1} = 0, W_{t+1}, i_{h,t+1})|\mathbf{s}_t, i_{Ct} = 0] \\ &\quad + P_{u,t+1} \beta E[V_{t+1}(t, \delta n_t, i_{u,t+1} = 1, W_{t+1} = 0, i_{h,t+1})|\mathbf{s}_t, i_{Ct} = 0] \\ &\quad + \epsilon_{Nt}. \end{aligned} \tag{A4}$$

That is, it is the utility of not committing a crime today and having an arrest record of δn_t tomorrow. The probability of unemployment in the next period is $P_{u,t+1}$, and this is incorporated in the expression above.

The value function of the individual at the first month of age 18 of not committing a crime is has 4 elements, which consists of the continuation payoffs from the 4 combinations of graduating or not, and being employed or not. That is,

$$
\begin{aligned}
V_{Nt}(\mathbf{s}_t) \quad = \quad & u_{Nt}(\mathbf{s}_t) \\
& +P_{hg}(1-P_{u,t+1}) \\
& E\left[V_{t+1}(t, \widehat{\delta} n_t, i_{u,t+1}=0, W_{t+1}, i_{hg}=1)|\mathbf{s}_t, i_{Ct}=0\right] \\
& +P_{hg}P_{u,t+1} \\
& E\left[V_{t+1}(t, \widehat{\delta} n_t, i_{u,t+1}=1, W_{t+1}=0, i_{hg}=1)|\mathbf{s}_t, i_{Ct}=0\right] \\
& +(1-P_{hg})(1-P_{u,t+1}) \\
& E\left[V_{t+1}(t, \widehat{\delta} n_t, i_{u,t+1}=0, W_{t+1}, i_{hg}=0)|\mathbf{s}_t, i_{Ct}=0\right] \\
& +(1-P_{hg})P_{u,t+1} \\
& E\left[V_{t+1}(t, \widehat{\delta} n_t, i_{u,t+1}=1, W_{t+1}=0, i_{hg}=0)|\mathbf{s}_t, i_{Ct}=0\right] \\
& +\epsilon_{Nt}.
\end{aligned} \tag{A5}
$$

After the first month of the age 18, the individual has either graduated from high school or not. Hence, the value of the individual not committing any crime is just a combination of the payoffs from being employed or not. That is,

$$\begin{aligned}
V_{Nt}(\mathbf{s}_t) &= u_N(\mathbf{s}_t) \\
&+[1-P_{u,t+1}]\beta E[V_{t+1}(t,\widehat{\delta}n_t, i_{u,t+1}=0, W_{t+1}, i_{hg})|\mathbf{s}_t, i_{Ct}=0] \\
&+P_{u,t+1}\beta E[V_{t+1}(t,\widehat{\delta}n_t, i_{u,t+1}=1, W_{t+1}=0, i_{hg})|\mathbf{s}_t, i_{Ct}=0] \\
&+\epsilon_{Nt}. \qquad \text{(A6)}
\end{aligned}$$

8.2 The Solution Algorithm and the Log Likelihood[21]

The probability an agent commits a crime equals the probability that $\epsilon_{Ct} - \epsilon_{Nt} > \overline{V}_{Nt}(\mathbf{s}_t) - \overline{V}_{Ct}(\mathbf{s}_t)$, where $\overline{V}_{Nt}(\mathbf{s}_t)$ and $\overline{V}_{Ct}(\mathbf{s}_t)$ are the deterministic components of the values of committing and not committing the crime. From equation (**??**), $\overline{V}_{Nt}(\mathbf{s}_t) = u_N(\mathbf{s}_t) + \beta E\left[V_{t+1}(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct}=0\right]$, and similarly for $\overline{V}_{Ct}(\mathbf{s}_t)$. Hence, we need to derive the expected value functions $E[V(\mathbf{s}_{t+1})|\mathbf{s}_t, i_{Ct}]$ at each DP solution step. To do this we need to integrate the value function with respect to the taste shock $(\epsilon_{Nt}, \epsilon_{Ct})$, the wage, high school graduation and employment shocks when needed. We follow the steps described below.

1) Integration with respect to the taste shock: Rust (1987) suggests a method that allows for the analytical integration of the value function when we assume that the shocks ϵ_{Nt}, ϵ_{Ct} have *i.i.d.* extreme values distributions. Note that this is for given values of all other shocks. In this event he

[21] The programs are available on request.

points out that the expected value function in period t has the following expression:

$$E_{\{\epsilon\}}[V_t(\mathbf{s}_t)] = log[exp(\overline{V}_{Nt}(\mathbf{s}_t)) + exp(\overline{V}_{Ct}(\mathbf{s}_t))]. \tag{A7}$$

$\overline{V}_{Nt}(\mathbf{s}_t)$ and $\overline{V}_{Ct}(\mathbf{s}_t)$ are the deterministic values of not committing a crime and committing a crime at period t and state vector $\mathbf{s}_t$. This eliminates the need to numerically integrate the value function with respect to the taste shocks ϵ_{Nt}, ϵ_{Ct}. In addition for the extreme values distribution, the probability of committing a crime and getting caught for individual i given s_{it} is

$$P(i_C = 1|\mathbf{s}_{it}) = P_A \frac{\exp(\overline{V}_{Ct}(\mathbf{s}_{it}))}{\exp(\overline{V}_{Nt}(\mathbf{s}_{it})) + \exp(\overline{V}_{Ct}(\mathbf{s}_{it}))} \tag{A8}$$

where P_A is the probability of getting caught.

2) Integration with respect to the wages and employment: The expected value function at period t is

$$E_{\{W,i_u,\epsilon\}}[V(\mathbf{s}_t)|\mathbf{s}_{t-1}, i_{Ct-1}] = E_{\{W,i_u\}}(\log[exp(\overline{V}_{Nt}(\mathbf{s}_t))+exp(\overline{V}_{Ct}(\mathbf{s}_t))]|\mathbf{s}_{t-1}, i_{Ct-1}). \tag{A9}$$

which follows from $(A7)$. Expanding further,

$$
\begin{aligned}
& E_{\{W,i_u\}}(\log[exp(\overline{V}_{Nt}(\mathbf{s}_t)) + exp(\overline{V}_{Ct}(\mathbf{s}_t))]|\mathbf{s}_{t-1}, i_{Ct-1}) \\
= \; & (1 - P_{u,t-1}) \int \log[exp(\overline{V}_{Nt}(.,.,i_{u,t} = 0, W_t)) \\
& + exp(\overline{V}_{Ct}(.,.,i_{u,t} = 0, W_t))] f(W_t|\mathbf{s}_{t-1}, i_{Ct-1}) dW_t \\
& + P_{u,t-1} \log[exp(\overline{V}_{Nt}(.,.,i_{u,t} = 1, W_t = 0)) \\
& + exp(\overline{V}_{Ct}(.,.,i_{u,t} = 1, W_t = 0))] \qquad \text{(A10)}
\end{aligned}
$$

where $f(W_t|.)$ in $(A10)$ is defined to be the distribution of wages parameterized in equations (6) to (9). We approximate this integral by taking finite grid points over the wage distribution and evaluate the density-weighted sum of the value function as the integral (See Rust 1998).

Since we assume that past criminal records depreciate at rate $(1-\delta_t)$, the past criminal history variable n_t can take values other than integers. Since we cannot evaluate the expected value function at so many state space points of n_t, we solve for the expected value function at finite q Chebychev grid points $(n_1, ..., n_q)$ and then interpolate them using the Chebychev Polynomial Least Squares Interpolation (for details, see Judd (1998)).

Recall that we allow for unobserved heterogeneities as well. When individuals are of different types, then the above calculations need to be done for each type. Finally, the above calculation will depend on the parameter θ as well. The likelihood increment for individual of type j in period t who has period t state variable s_{it}, period $t-1$ state variable s_{it-1}, whose period t criminal choice is i_{Ct} and period $t-1$ criminal choice is i_{Ct-1} is[22]

$$L_{it}(\theta_j) = L_{itC}(\theta_j)L_{itE}(\theta_j)L_{itHS}(\theta_j) \tag{A11}$$

where θ_j is the parameter vector of type j. Furthermore,

$$\begin{aligned}
L_{itC}(\theta_j) &= [P(i_{Ct}=1|\mathbf{s}_{it},\theta_j)]^{i_{Ct}}[1-P(i_{Ct}=1|s_{it},\theta_j)]^{1-i_{Ct}}, \\
L_{itE}(\theta_j) &= I(age<16)+I(age\geq 16)[P_{u,t}(s_{i,t-1},i_{Ct-1},\theta_j)^{i_{u,t}}] \\
&\quad \{[1-P_{u,t}(s_{i,t-1},i_{Ct-1},\theta_j)]f(W_t|\mathbf{s}_{it-1},i_{Ct-1},\theta_j)\}^{1-i_{u,t}} \\
L_{itHS}(\theta_j) &= I(age\neq 18) \\
&\quad +I(age=18)P_{hg}(s_{i,48},\theta_j)^{i_{hg}}\left[1-P_{hg}(s_{i,48},\theta_j)\right]^{1-i_{hg}}. \qquad \text{(A12)}
\end{aligned}$$

where $L_{itC}(\theta_j)$ is the crime increment of the likelihood function. If the individual commits the crime, then the likelihood is given by the first term, and if he does not, by the second term. $L_{itE}(\theta_j)$ is the employment and wage increment of the likelihood, where $P_{u,t}(s_{i,t-1},i_{Ct-1},\theta_j)$ is the unemployment probability of type j individual at period t given θ_j . If he is below 16, it equals unity. If he is above 16, and is unemployed, the likelihood increment is $P_{u,t}(s_{i,t-1},i_{Ct-1},\theta_j)$. If he is employed, then the likelihood increment is the product of the probability of employment and the wage density. $L_{itHS}(\theta_j)$ is the high school graduation increment of the likelihood, which is defined similarly. $P_{hg}(s_{i,48},\theta_j)$ is the high school graduation probability of type j individual.

[22]Recall that there are four types. s_{it-1}, i_{Ct-1} are relevant for current labor market outcomes.

The likelihood increment for individual i is the product of the likelihood increments for all quarters and types so that

$$\begin{aligned} L_i(\theta) &= \pi_1\pi_{11}\prod_{t=1}^{T}[L_{it}(\theta_1)] + \pi_2\pi_{21}\prod_{t=1}^{T}[L_{it}(\theta_2)] \\ &\quad +\pi_1\pi_{12}\prod_{t=1}^{T}[L_{it}(\theta_3)] + \pi_2\pi_{22}\prod_{t=1}^{T}[L_{it}(\theta_4)] \end{aligned} \tag{A13}$$

where θ is the vector of parameters for all types. Also π_j is the probability of the individual being of crime type j, while π_{jl} is the conditional probability of the individual being of unemployment type l, given he is of crime type j.

The total log likelihood is

$$l(\theta) = \sum_{i=1}^{N} log[L_i(\theta)]. \tag{A14}$$

9 Tables and Graphs

Table 1: Sample Statistics

% Whites	52.2
% Father present in childhood home	78.3
% Father unemployed during respondents childhood	14.8
% Mother present in childhood home	97.0
% Mother worked during respondents childhood	56.9
% High socioeconomic status	49.2
% Grew up in a not loving household	6.04
% Gang member before 18 years old	36.3
No. of friends arrested; average	1.53
% Parents arrested	2.47
% Protestant	42.3
% Catholic	31.3
% Jewish	1.92
% Other religion	1.92
% No religious beliefs	12.6
% Unknown on religion	9.89
% Of high school graduates	44.2
% Who obtained the high school equivalency degree	17.9
% Who are none of the above two	37.9

Table 2: Parameter Estimates[23]

Utility of not committing a crime

Crime type 1						
	Before 18			After 18		
Constant	c_{01}^1	9.134	(9.453)	c_{02}^1	12.997	(11.13)
High school attend.	c_1^1	-5.858	(4.357)			
Not working	c_{u1}^1	8.542	(7.071)	c_{u2}^1	-3.756	(4.287)
Medium wage	c_{m1}^1	-5.042	(5.529)	c_{m2}^1	-2.302	(2.016)
High wage	c_{h1}^1	-2.230	(7.023)	c_{h2}^1	0.6592	(1.495)
State dependence	c_{61}^1	-0.7507	(0.6338)	c_{62}^2	-5.944	(6.508)
State dependence	α_1^1	0.3475	(0.8063)	α_2^1	0.03222	(0.06382)
Final period value				γ^1	-10.737	(144.2)
Age (after 17)	c_{04}^1	0.02511	(0.09305)			
Crime type 2						
	Before 18			After 18		
Constant	c_{01}^2	-6.085	(11.15)	c_{02}^2	-1.069	(12.78)
High school attend.	c_1^2	3.185	(4.759)			
Not working	c_{u1}^2	-3.166	(9.331)	c_{u2}^2	-2.614	(3.161)
Medium wage	c_{m1}^2	3.737	(10.76)	c_{m2}^2	0.6199	(1.464)
High wage	c_{h1}^2	-7.348	(12.25)	c_{h2}^2	-1.150	(1.465)
State dependence	c_{61}^2	-0.6751	(2.967)	c_{62}^2	-0.6318	(0.7100)
State dependence	α_1^2	0.2017	(1.322)	α_2^2	0.4602	(0.9926)
Final period value				γ^2	-14.554	100.9
Age (after 17)	c_{04}^2	0.06721	(0.06500)			
All types						
Age (after 17 before 18)	c_{03}	0.1479	(0.4096)			
High school grad.	c_5	-0.6380	(0.9061)			
Crime type 1 prob.	π_1	0.6729*	(0.1297)			

Utility of committing a crime

	Before 18			After 18		
Not working (Crime type 1)	d_{u1}^1	12.613	(6.894)	d_{u2}^1	1.061	(3.649)
Not working (Crime type 2)	d_{u1}^2	0.009310	(10.05)	d_{u2}^2	-1.947	(3.405)
All ages						
State dependence	d_1	2.227	(4.584)			

[23]Standard errors are in parenthesis. * means that the estimate is significantly different from zero at the 95% confidence level..

Other Parameters

Discount Factor	β	0.9902*	(0.01619)
Depreciation	δ	0.9790*	(0.001228)
Depreciation at 18	$\widehat{\delta}$	0.7394*	(0.05294)
Crime type 1 prob.	π_1	0.6729*	(0.1297)

$$
\begin{aligned}
u_N(\mathbf{s}_t) &= c_{01}I(age < 18) + c_{02}I(age \geq 18) \\
&+[c_{03}I(17 \leq age < 18) + c_{04}I(17 \leq age)](t-36) + c_1 i_{h,t} \\
&+[c_{u1}i_{u,t} + [c_{m1}I_m + c_{h1}I_h](1 - i_{u,t})]I(age < 18) \\
&+[c_{u2}i_{u,t} + [c_{m2}I_m + c_{h2}I_h](1 - i_{u,t})]I(age \geq 18) \\
&+c_5 i_{hg} + c_6(n_t)^{\alpha}
\end{aligned}
$$

$$
u_C(\mathbf{s}_t) = [d_{u1}I(age < 18) + d_{u2}I(age \geq 18)]\, i_{u,t} + d_1\sqrt{n_t}.
$$

High school graduation parameters

Crime type1	g_0^1	0.2783	(0.2034)
Crime type2	g_0^2	0.06626	(0.6299)
Criminal history	g_1	-0.4056*	(0.1333)

$$
\begin{aligned}
P_{hg} &= \exp(\phi_{hg})/[1 + \exp(\phi_{hg})] \\
\phi_{hg} &= g_0 + g_1 n_{48}.
\end{aligned}
$$

Initial unemployment probability

Employment type 1	h_0^1	-0.3121	(0.6710)
Employment type 2	h_0^2	2.0441*	(0.9866)
Criminal history	h_1	-0.006883	(0.4514)

$$
\begin{aligned}
P_{u,24} &= \exp(\phi_{u,24})/[1 + \exp(\phi_{u,24})] \\
\phi_{u,24} &= h_0 + h_1 n_{24}.
\end{aligned}
$$

Unemployment probability

Employment type 1				Employment type 2			
Before 18	b^1_{00}	-2.803*	(0.1758)	Before 18	b^2_{00}	-1.263*	(0.1411)
After 18	b^1_{01}	-2.821*	(0.2280)	After 18	b^2_{01}	-1.788*	(0.2036)
Age	b^1_1	-0.06022*	(0.01137)	Age	b^2_1	-0.03630*	(0.01090)
All types							
High school grad.	b_2	-0.4309*	(0.05949)				
Criminal history	b_3	0.1524*	(0.03027)				
Before 18	b_{40}	3.949*	(0.1777)				
After 18	b_{41}	5.402*	(0.04812)				

$$
\begin{aligned}
P_{u,t+1} &= \exp(\phi_{u,t+1})/[1+\exp(\phi_{u,t+1})] \\
\phi_{u,t+1} &= b_{00}I(age < 18) + b_{01}I(age \geq 18) + b_1(t+1-24) + b_2 i_{hg} + b_3 n_{t+1} \\
&\quad +[b_{40}I(age < 18) + b_{41}I(age \geq 18)]i_{u,t}.
\end{aligned}
$$

Wage growth

16-19 dummy	κ_1	0.01801*	(7.576E-4)
20-23 dummy	κ_2	0.001836	(0.001011)
24-26 dummy	κ_3	0.002710*	(6.735E-4)
16-23, age	κ_4	5.963E-5	(3.472E-5)
24-26, age	κ_5	0.001558*	(2.468E-5)
Criminal history	κ_6	-0.006970*	(2.885E-4)
Std. error	σ_g	0.07030*	(4.769E-4)

$$\log(W_t) - \log(W_{t-1}) \sim N(\mu_{gt}(.), \sigma_g),$$

$$\begin{aligned}\mu_{gt}(.) &= \kappa_1 I(16 < age \leq 19) + \kappa_2 I(20 < age \leq 23) + \kappa_3 I(24 < age) \\ &+[\kappa_4 I(16 < age \leq 23)\,(t-36) + \kappa_5 I(24 < age)\,(t-120)] + \kappa_6 n_t.\end{aligned}$$

Starting wage

Const.	μ_{b0}	1.779*	(0.04520)
Criminal history	μ_{b1}	0.04406	(0.04426)
Std. error	σ_b	0.5853*	(0.007830)

$$\begin{aligned}\log(W_t) &\sim N(\mu_b(n_t), \sigma_b) \\ \mu_b(n_t) &= \mu_{b0} + \mu_{b1} n_t.\end{aligned}$$

Probability of being employment type l conditional on being crime type j (π_{jl})

Crime type 1	π_{11}	0.6143*	(0.08810)
Crime type 2	π_{21}	0.3128*	(0.09885)

$$\pi_{j2} = 1 - \pi_{j1},\ j = 1, 2$$

$$\begin{aligned}L_i(\theta) &= \pi_1 \pi_{11} \prod_{t=1}^{T} [L_{it1}(\theta_1)] + \pi_2 \pi_{21} \prod_{t=1}^{T} [L_{it2}(\theta_2)] \\ &+\pi_1 \pi_{12} \prod_{t=1}^{T} [L_{it3}(\theta_3)] + \pi_2 \pi_{22} \prod_{t=1}^{T} [L_{it4}(\theta_4)]\end{aligned}$$

Table 3: Regression Results of Observable Characteristics on Posterior Probability of Crime Type 1

Variable	Estimates	Std. Error	t-Statistic	P>|t|
Constant	-0.4152233	.0810065	-5.13	0.000
Race	.0365201	.0263435	1.39	0.167
Father at home	.0210769	.0276889	0.76	0.447
Father unemployed	.0152203	.0304982	0.50	0.618
Mother at home	-.0358439	.065344	-0.55	0.584
Mother worked	.0253842	.022418	1.13	0.258
Socioeconomic status	-.0133342	.0219386	-0.61	0.544
Loving household	.0505304	.0454547	1.11	0.267
Gang member	-.0313391	.0239391	-1.31	0.191
No. friends arrested	-.0033121	.0083974	-0.39	0.694
Parents arrested	-.1116975	.0702102	-1.59	0.113
Rel: Protestant	-.0275576	.0290023	-0.95	0.343
Rel: Catholic	.0103268	.0309944	0.33	0.739
Rel: Jewish	-.0031731	.0816086	-0.04	0.969
Rel: other, none	.0484281	.0822513	0.59	0.556

R-Squared: .0524

Adjusted R-Squared: .0140

Figures

Figure 1: Age Arrest Profiles

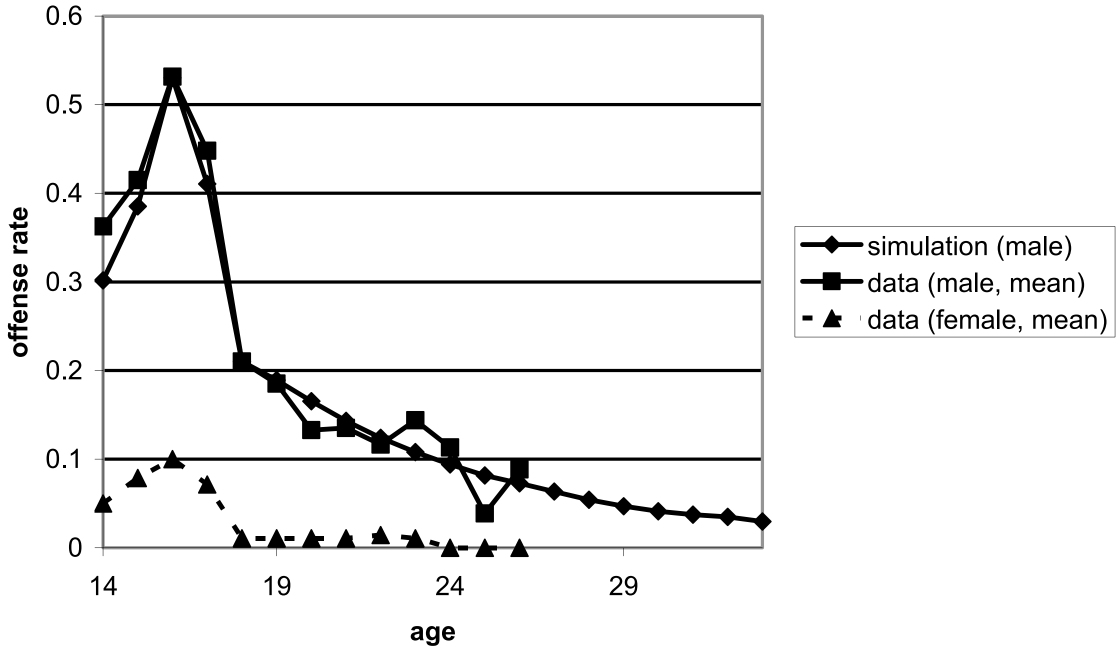

Figure 2: Age Arrest Profiles with Different Past Criminal Records

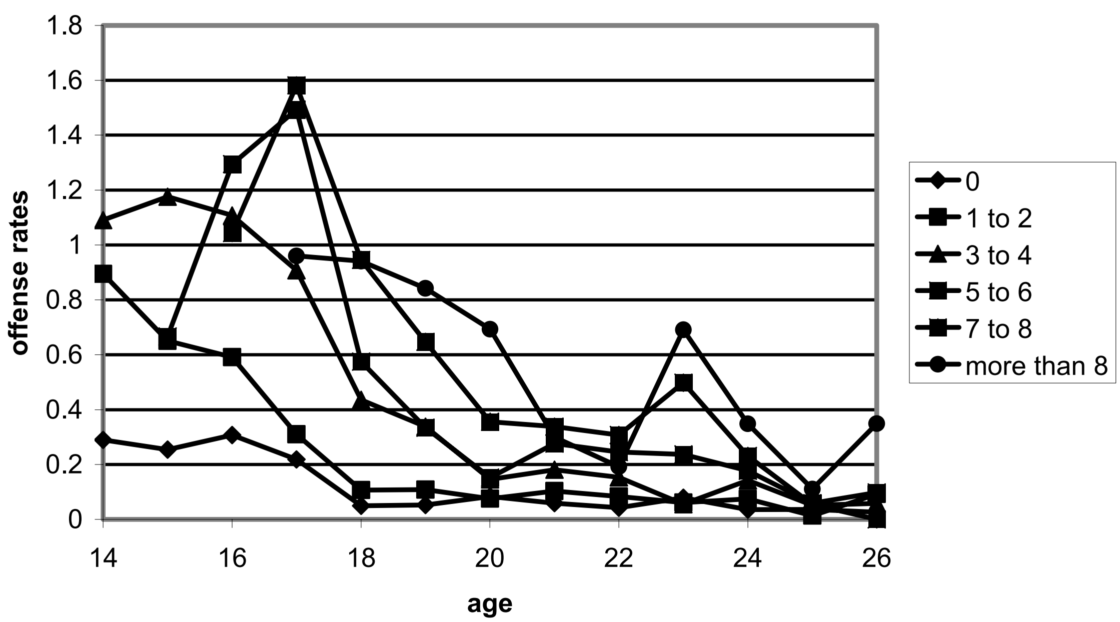

Figure 3: Age Unemployment Profiles

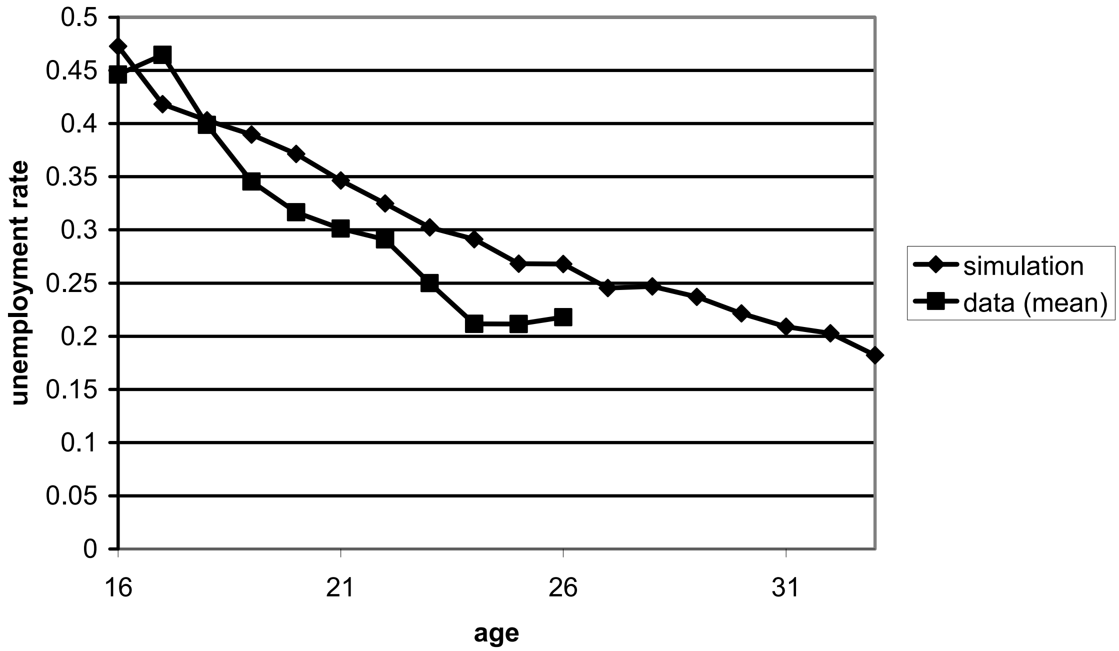

Figure 4: Age Wage Profiles

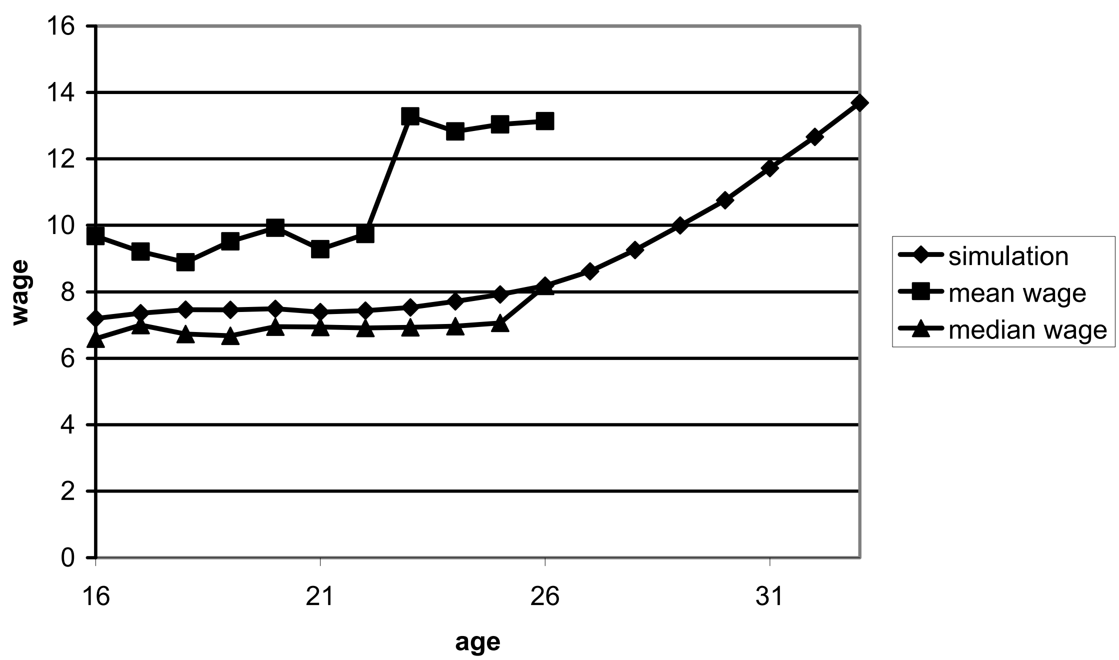

Figure 5: Simulated Age Arrest Profiles for Various Types

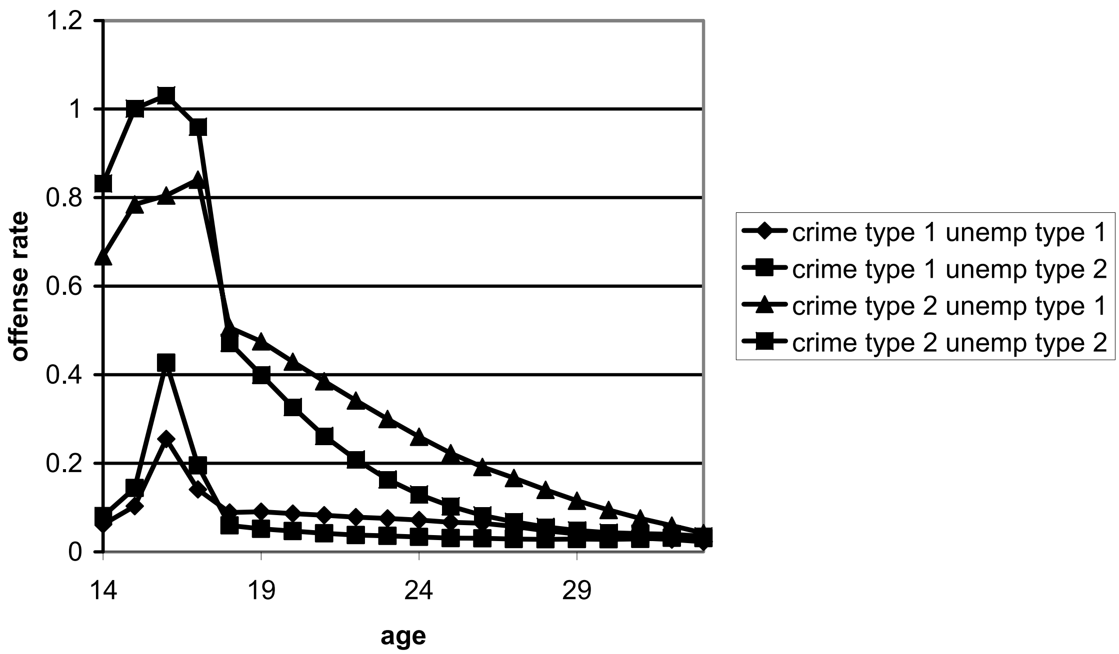

Figure 6: Simulated Age Unemployment Profiles of Various Types

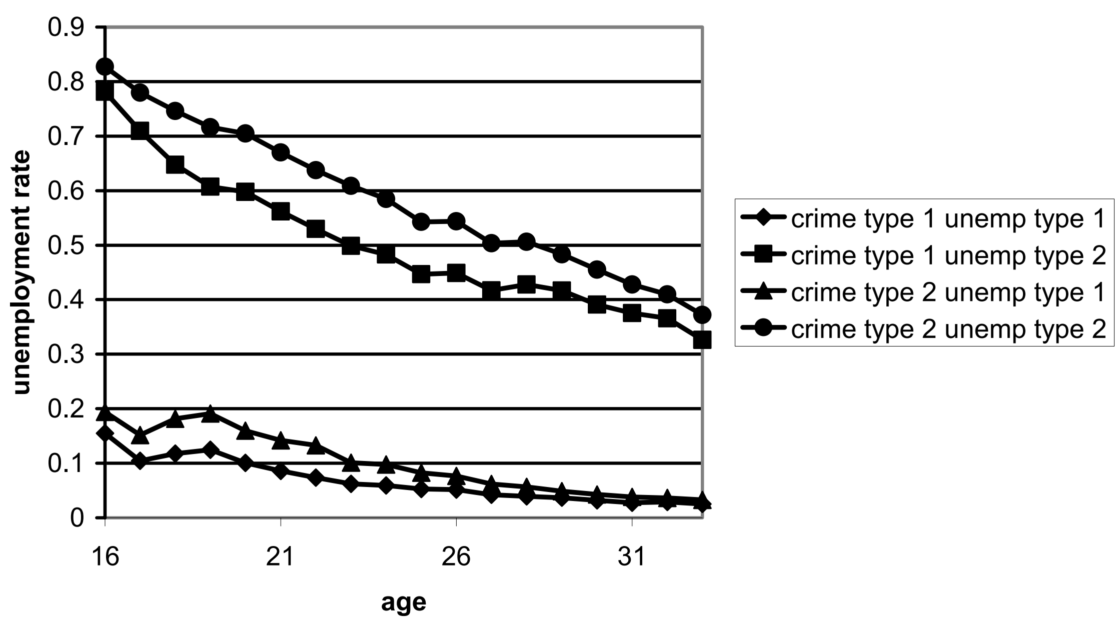

Figure 7: Simulated Age Wage Profiles of Various Types

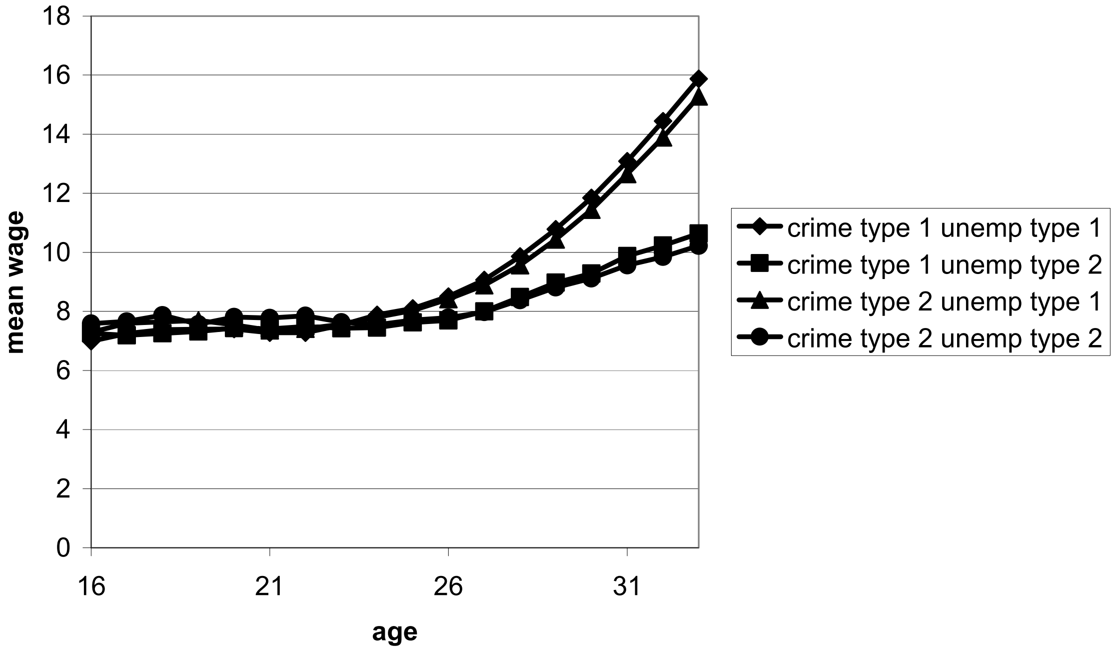

Figure 8: Simulated Age Arrest Profiles with Different Past Criminal Histories

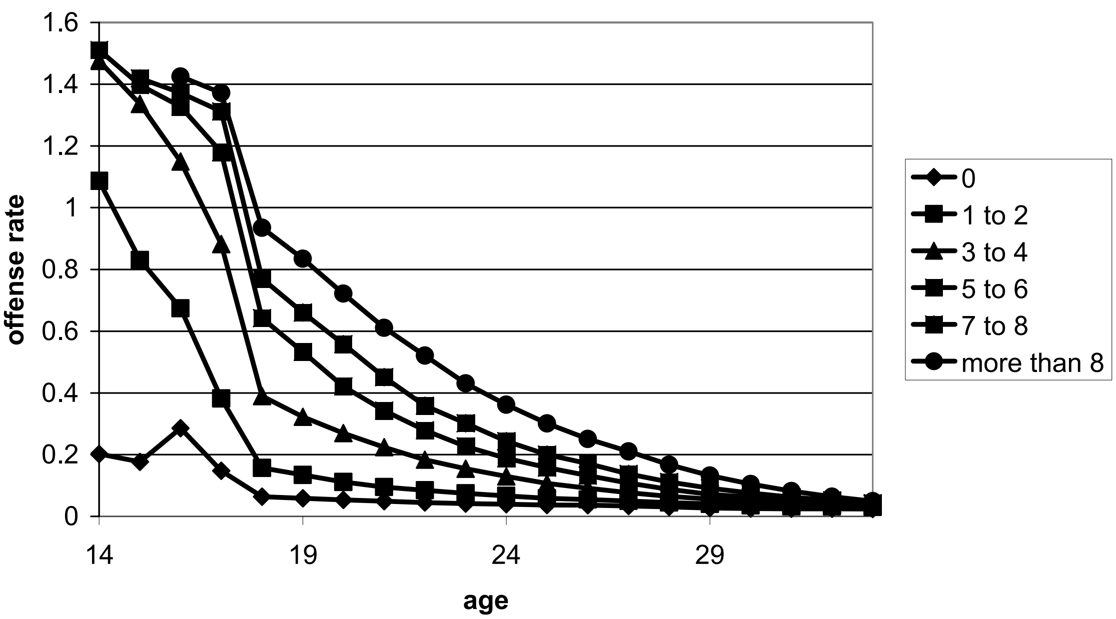

Figure 9: Ratio of Age Arrest Profiles: No Labor Market Effects of Arrests

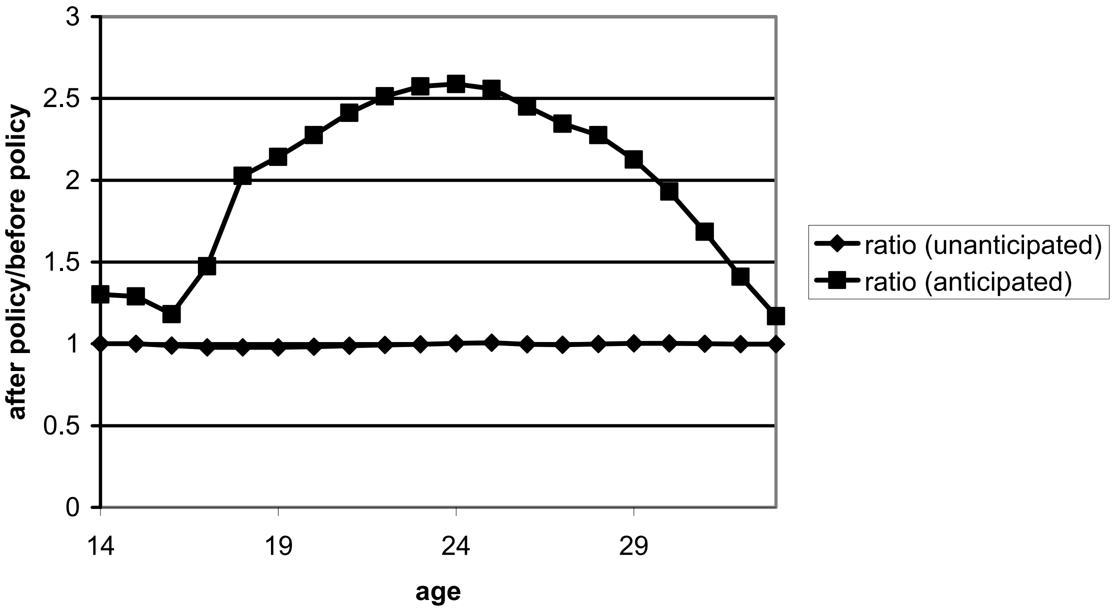

Figure 10: Ratio of Age Unemployment Profiles: No Labor Market Effects of Arrests

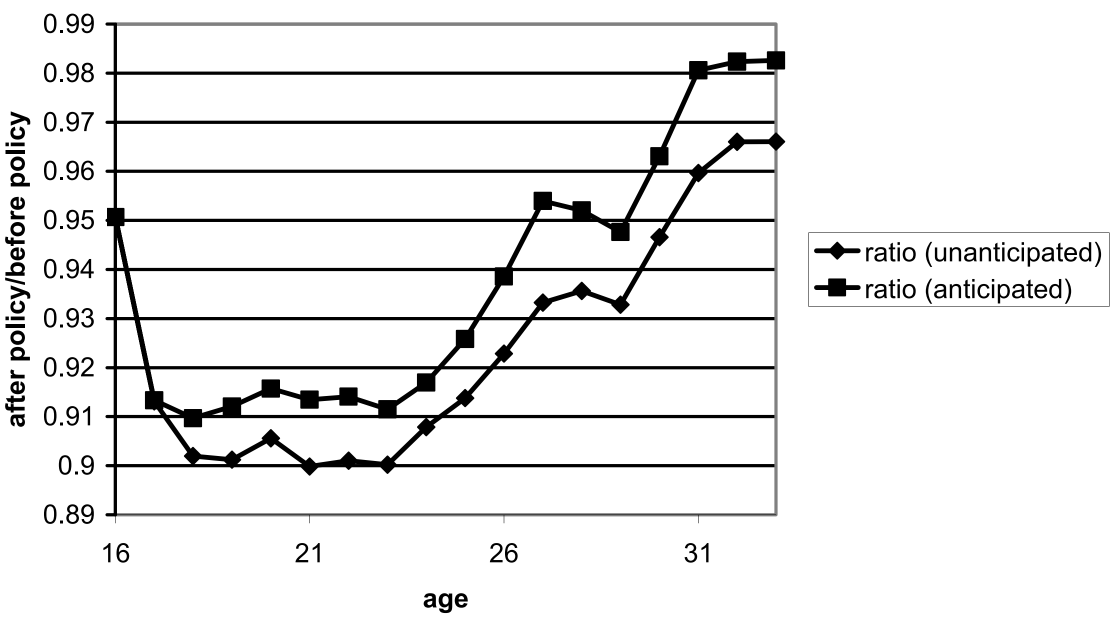

Figure 11: Ratio of Age Wage Profiles: No Labor Market Effects of Arrests

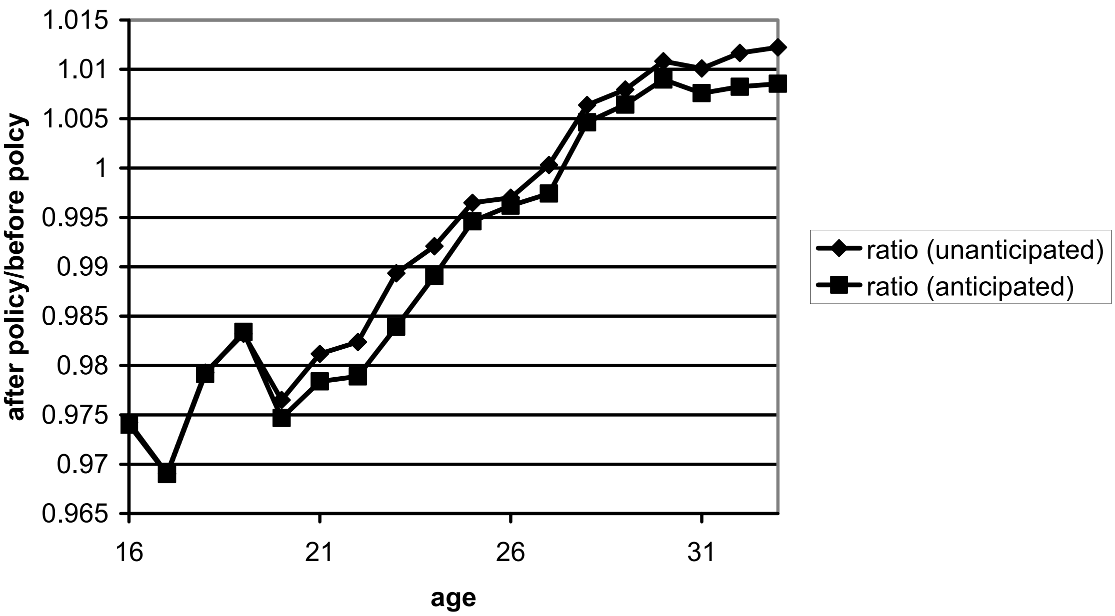

Figure 12: Ratio of Age Unemployment Profiles: 5% Decrease in Unemployment Transition Probability after Age 18

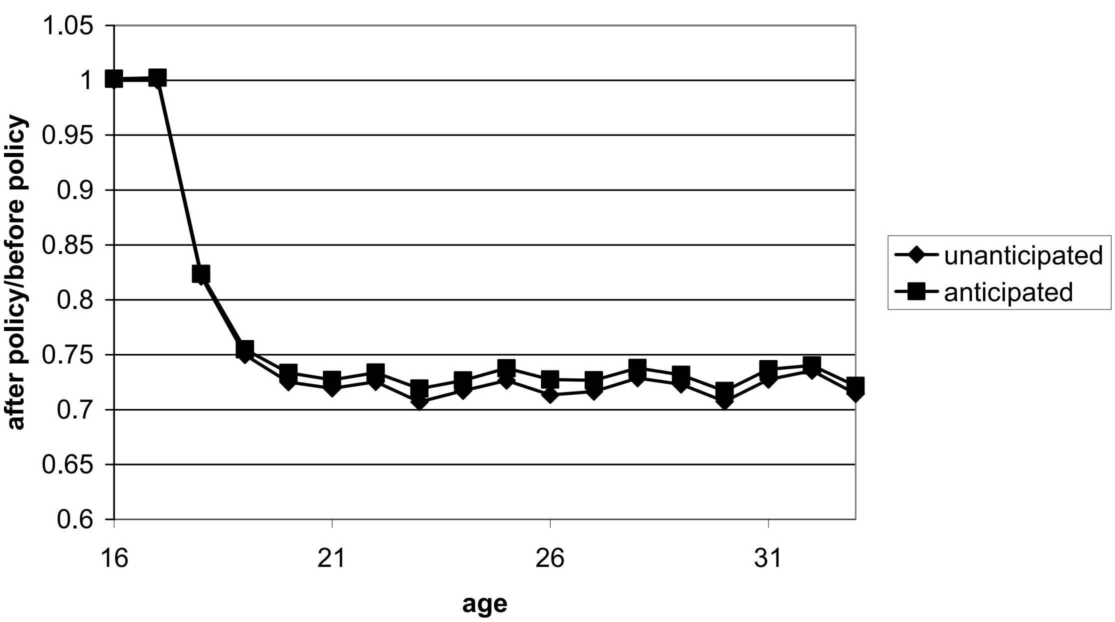

**Figure 13: Ratio of Age Arrest Profiles:
5% Decrease in Unemployment Transition Probability after 18**

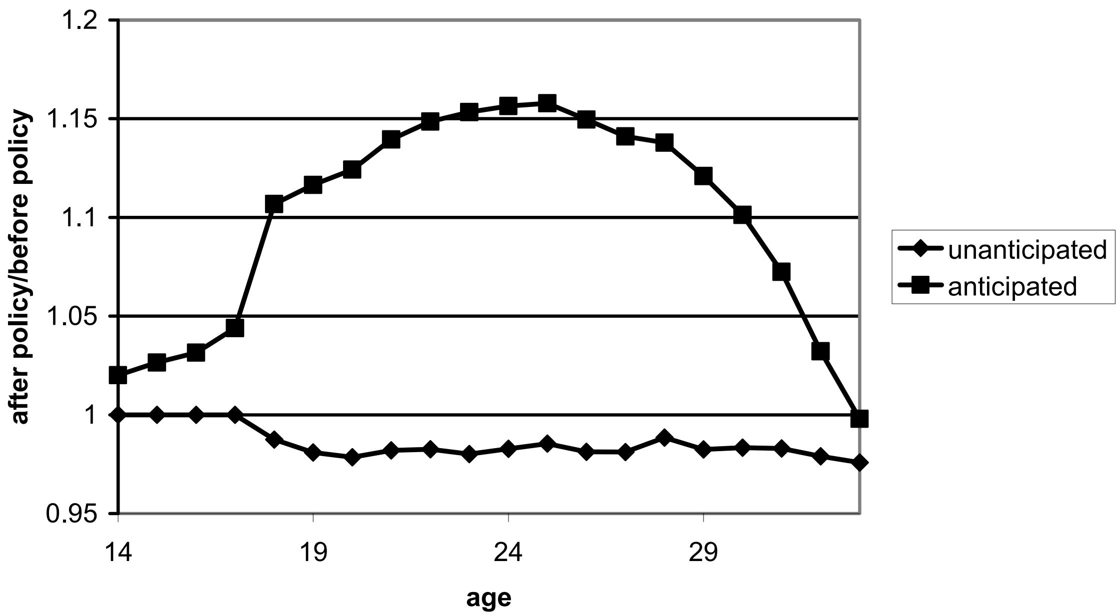

**Figure 14: Ratio of Age Arrest Profiles: Anticipated Temporary
Fluctuations in Unemployment Transition Probability**

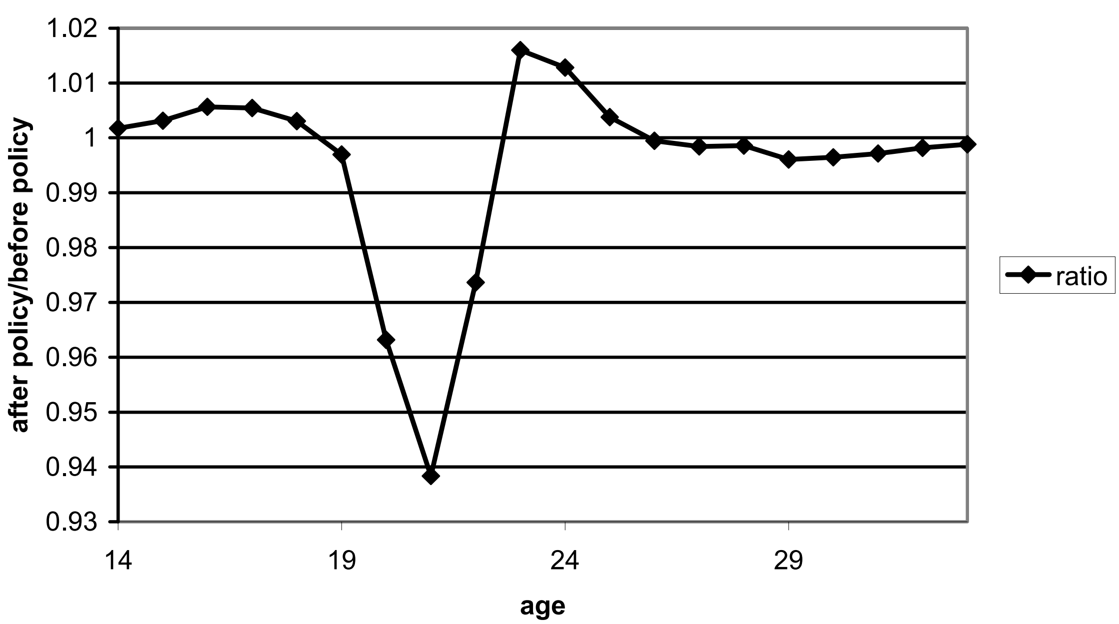

Figure 15: Ratio of Age Arrest Profiles: Anticipated 10% Increase in Probability of Getting Arrested

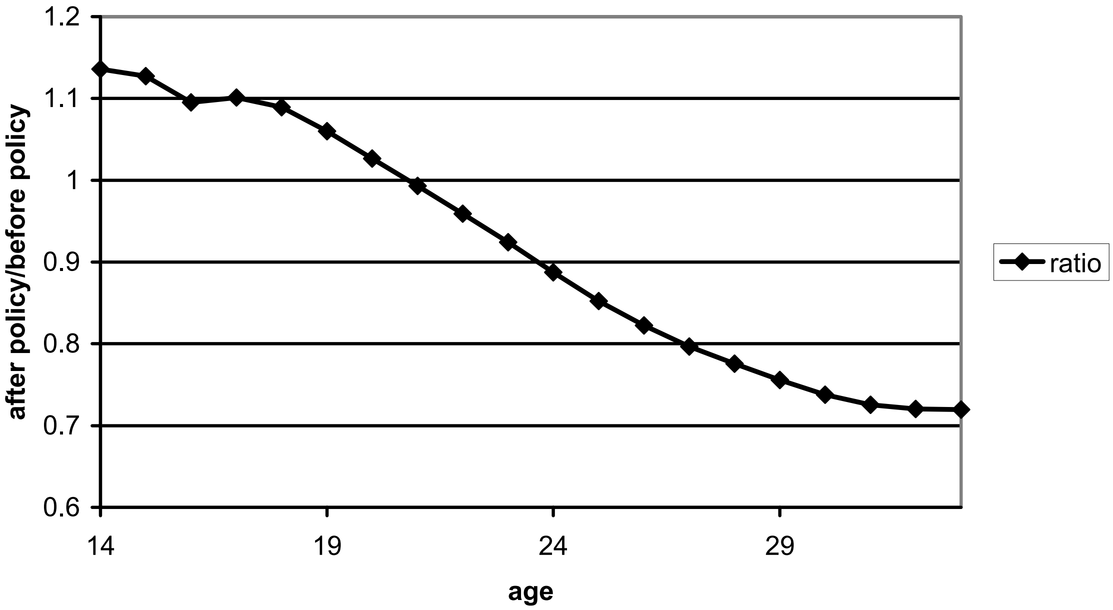

Figure 16: Ratio of Age Arrest Profiles: Anticipated Transformation of Juvenile Crime Records to Adult Records

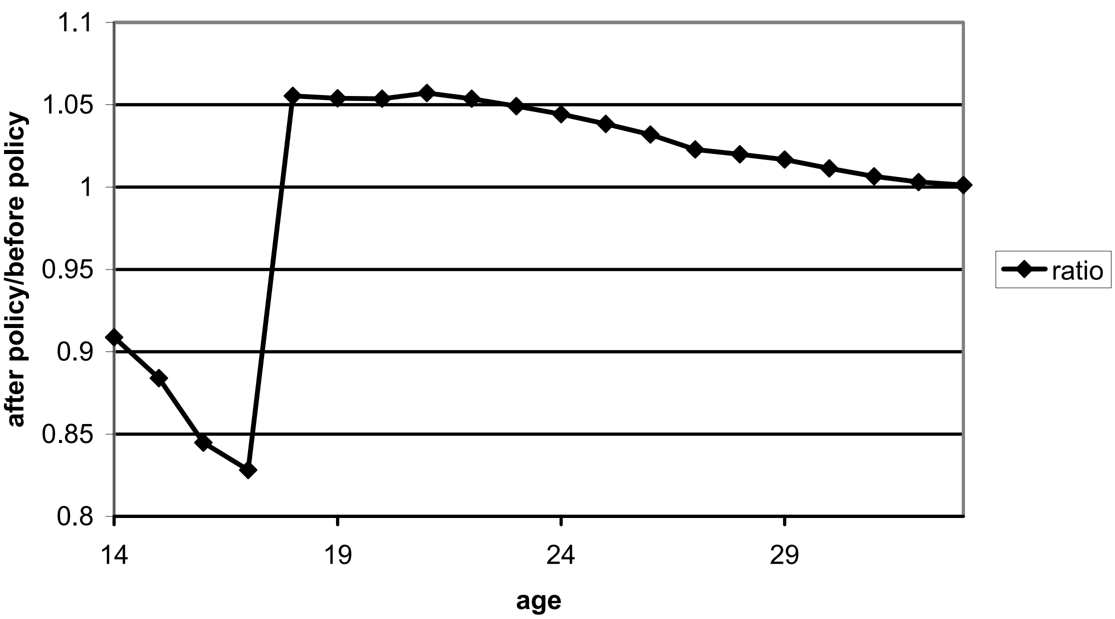

Figure 17: Ratio of Age Arrest Profiles: Anticipated 0.1% increase in all discount factors