NBER WORKING PAPER SERIES

CRIME AND YOUNG MEN: THE ROLE OF ARREST, CRIMINAL EXPERIENCE AND HETEROGENEITY

Susumu Imai
Hajime Katayama
Kala Krishna

Working Paper 12221
http://www.nber.org/papers/w12221

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2006

Crime and Young Men: The Role of Arrest, Criminal Experience, and Heterogeneity
Susumu Imai, Hajime Katayama and Kala Krishna
NBER Working Paper No. 12221
May 2006
JEL No. K42, J24

ABSTRACT

Using National Youth Survey (NYS) data, we examine the relationship of current criminal activity and past arrests using an ordered probit model with unobserved heterogeneity. Past arrests raise current criminal activity only for the non-criminal type, while past criminal experience raises current criminal activity for both types. Also, the age crime profile peaks at age 18 for non-criminal type individuals, but for criminal type individuals, it continues to rise with age. Past research indicates that age arrest profiles rise till age 18 and then fall for both types, suggesting lower apprehension rates for criminal type individuals.

Kala Krishna
Department of Economics
523 Kern Graduate Building
The Pennsylvania State University
University Park, PA 16802
and NBER
kmk4@psu.edu

1. Introduction

In this paper, we use data from the National Youth Survey (NYS) to investigate the determinants of the criminal behavior of young men.[1] Our work is, we believe, the first in the economics literature to incorporate three important factors affecting criminal behavior: arrests, criminal experience, and observed and unobserved individual characteristics such as family and social background and unobserved types. Past work has only looked at subsets of these and hence could have obtained misleading results. Distinguishing between these factors is of vital importance in designing effective anti-crime policies. For example, allowing for more than one type is important. Suppose that there are two types: the criminal type and the non-criminal type. The criminal type commits crimes repeatedly while the non-criminal type does not. If past arrests result in more crimes today by the non-criminal type, but not by the criminal type, then policies that are tough on first time offenders could well be self-defeating!

NYS contains self-reported data on arrests as well as on crimes committed. We use this information to identify the arrest effect and the criminal experience effect. We use the term "arrest effect" to mean the overall effect of past arrests on individuals' criminal behavior. The arrest effect has at least three components: the perceived probability effect, the stigma effect and the criminal training effect.

Upon arrest, an individual may update his perceived probability of being caught and punished upward, which makes him less likely to commit a crime in the future (i.e., the perceived probability effect, see Lochner, (2000)). Arrests and convictions cause stigma, which may decrease earnings in legitimate activities, making crime more attractive. Also, when incarcerated, he cannot commit crimes, though he may learn criminal skills from fellow prisoners and thus accumulate criminal training (see Bayer, Pintoff and Posen, (2004)). The former would reduce current crimes (and arrests) while the latter would raise current crimes (and reduce current arrests).

Similarly, the "experience effect" captures the overall effect of past criminal experience on current criminal behavior, other things being constant. The experience effect has at least two components: the learning by doing effect and the effect via perceived probability. An individual could become more skilled at committing crimes with experience, i.e., leaning by doing. In addition, an individual may update his perceived probability of arrest downwards: more past crimes, given arrests, would tend to reduce the perceived probability of arrest, and increase crimes today.

In order to capture observed heterogeneity, we include as explanatory variables information on schooling, employment and parental background. In order to allow for unobserved heterogeneities, we incorporate random effects on the constant term. In addition, we allow slope coefficient estimates to differ across the two unobserved types. The estimation results indicate that the two types could be interpreted as the non-criminal type and the criminal type.

It is well known that individuals differ in their criminal behavior: most people do not commit crimes regularly, whereas a small fraction of the population commits most of the crimes. One of our is to see how the criminal types differ from the non-criminal types, not only in terms of the overall likelihood of committing crimes, but also in terms of the effect of past arrests and past criminal experience on current criminal behavior. However, it is inappropriate to simply assign types to individuals based on their arrest or criminal history. We infer individual's types by estimating a mixture model that allows us to have different parameters for the two types. The behavioral parameters for the two types as well as the posterior probability (given characteristics) that an individual is of one type or of the other should be determined so as to best replicate the data.[2] This is basically the approach taken here in using the “mixture” model. In Section 5, we show that we do need to take this approach, as the results from taking the simplistic approach are very different from those when the model is appropriately specified.

We use an ordered probit setting. This allows for the fact that most people do not commit any crimes. In addition, it improves over the simple probit setting as it also utilizes information about the number of crimes committed.[3]

Our estimation results are summarized as follows: The criminal experience effect is significantly positive for the non-criminal type, both before age 18 and after. It is positive for the criminal type as well, though insignificant after age 18. Importantly, the arrest effect turns out to be quite different between the two types. In particular, after age 18, the arrest effect for the criminal type is significantly negative, while it is significantly positive for the non-criminal type.[4] We also find that controlling for other variables, the crime rate increases with age even after age 18 for the criminal type, whereas it decreases with age after age 18 for the non-criminal type.[5] This result highlights the importance of using data on criminal activity and not just on arrests. In Imai and Krishna (2004), for example, it is shown that arrests of the criminal type fall off even faster than those of the non-criminal type after the age of 18. Our results indicate that this should not be interpreted as evidence that their criminal activity follows a similar path, rather, that criminal types become better at evading arrest after 18. In terms of policy, our results suggest that tough on crime policy may be more effective for repeat offenders since arrests seem to serve as a deterrent for criminal types.

There has been some recent work by economists analyzing criminal behavior using individual level data. Viscusi (1986) uses cross sectional data to argue that criminal behavior responds to economic incentives. Tauchen, Witte and Griesinger (1994) use the Philadelphia Birth Cohort Study panel and run a probit (with random effects) of arrests on past arrests and other variables including police spending. They find that past arrests have a positive coefficient and police spending has a negative coefficient.

Self-reported criminal activity data is used by Grogger (1998), Lochner (2004) and Mocan and Rees (1999). Using a probit approach, Grogger (1998) shows that criminal activity rises when labor market conditions worsen. Lochner (2004) uses both arrest and criminal activity data to show that education reduces arrests as well as criminal activities. Mocan and Rees (1999) estimate criminal choice equations using an ordered probit approach. Their focus is on the possible determinants of crime including employment conditions that vary across regions. However, none of them look at the effect of past criminal activity or arrests on current crime.

The remainder of the paper is organized as follows. In Section 2, we present the ordered probit model of crime and the estimation strategy. In Section 3, we discuss our data. In Section 4, we present the estimation results and their implications. In Section 5, we conclude by outlining some work in progress that extends our model in various directions.

2. Model Specification

To capture the relationship between crimes, arrests and unobserved heterogeneity we estimate a standard ordered probit model of criminal choice: for individual i at time t, the following equation describes his criminal behavior:

$$y_{it}^{*} = b_i + \varphi_0 + \boldsymbol{x}_{it}'\varphi_1 + \varepsilon_{it} \tag{1}$$

where

$$\varepsilon_{it} \sim \Phi(\cdot).$$

We assume the error term is i.i.d. and its distribution function is standard normal. b_i is the mean-zero random effects on the constant term and φ_0 is the constant term. $\boldsymbol{x}_{it}$ is a vector of time-varying explanatory variables of individual *i* at time *t*. The elements of this vector include age, past arrest records,[6] number of past crimes committed[7] (for criminal experience), total past years of schooling, total past years of work, etc. y_{it}^{*} is the unobserved latent index variable representing the propensity to commit crimes. We only have limited information on the number of crimes individual *i* committed in period *t*: we know the category into which this number falls but not the exact count. Let y_{it} denote this category. We specify that

$y_{it} = 1$ (commit no crime) if $y_{it}^{*} \leq \mu_1$,
$y_{it} = 2$ (commit 1 to 2 crimes) if $\mu_1 < y_{it}^{*} \leq \mu_2$,
$y_{it} = 3$ (commit 3 to 11 crimes) if $\mu_2 < y_{it}^{*} \leq \mu_3$,
$y_{it} = 4$ (commit 12 to 30 crimes) if $\mu_3 < y_{it}^{*} \leq \mu_4$,
$y_{it} = 5$ (commit 31 crimes or more) if $\mu_4 < y_{it}^{*}$.

We define $\mu_0 = -\infty$ and $\mu_5 = +\infty$, and normalize μ_1 to be 0 for identification.

We allow for unobserved heterogeneities both in the constant term as well as the other coefficients.[8] First, we allow for two types with potentially different parameters so that their behavior could differ.[9] Second, we allow for the mean-zero random effects on the constant term, as done by Tauchen et al. (1994) among others. Even though we incorporate some observable characteristics of individuals that affect their chances of committing a crime, there may be many other individual characteristics that matter, but are unobservable to the econometrician. Including random effects in the constant term is a way of capturing such time-invariant effects, at least partially.[10]

As discussed above, in our estimation model, past criminal records are endogenous. Since the schooling and work choice are not explicitly modeled, both past years of schooling and past years of work are assumed to be strictly exogenous.[11] Because of the poor quality of schooling and work data in NYS, we decided to leave the estimation of a more formal model of schooling, work and crime choice for future research. We also assumed that given past criminal history, past arrests are exogenous. We believe that this assumption is reasonable since individuals do not directly choose the number of arrests.[12]

When lagged criminal experience enters into the right hand side, we face a problem known as the Initial Conditions Problem (see Heckman (1981)). Since initial criminal experience and the random intercept term are correlated, the coefficient estimate of the lagged experience is biased. We use a method similar to Wooldridge (2000) to deal with this problem. That is, we can rewrite the random effects term as follows:

$$b_i = \theta_0 + \theta_1 CE_{i0} + \theta_2 \overline{WE}_i + \theta_3 \overline{HE}_i + \boldsymbol{x}_i' \boldsymbol{\theta}_4 + \eta_i$$

where CE_{i0} is the initial period criminal experience, $\overline{WE}_i$ is the time-average of the number of years of employment after age 18, and $\overline{HE}_i$ is the time-average of the highest grade attended by the individual. $\boldsymbol{x}_i$ is a vector of time-constant covariates such as family background. η_i is assumed to be mean-zero normally distributed and uncorrelated with y_{i0}^*. Then, equation (1) can be rewritten as follows:

$$y_{it}^{*} = (\theta_0 + \varphi_0) + \theta_1 CE_{i0} + \theta_2 \overline{WE}_i + \theta_3 \overline{HE}_i + \boldsymbol{x}_i'\boldsymbol{\theta}_4 + \boldsymbol{x}_{it}'\varphi_1 + \eta_i + \varepsilon_{it} \qquad (2)$$

where $(\theta_0 + \varphi_0)$ is the constant term to be estimated, and η_i is the new random effects term.

Incorporating heterogeneities in the ordered probit estimation involves integration over both the latent error term ε, the random effect of the constant term η, and the type heterogeneity. We adopt a Bayesian approach suggested by Albert and Chib (1993). We choose to do so for computational reasons. With the Bayesian approach, integration over the latent error term and the random effect term can be done straightforwardly as part of the simulation process of the posterior distribution. This is known as the Markov Chain Monte Carlo (MCMC) method. Geweke, Keane and Runkle (1997), using the multinomial probit model as an example, argue that the Bayesian MCMC method works well compared to other standard estimation techniques, such as Simulated Maximum Likelihood or Simulated Method of Moments. The major difficulty in Bayesian estimation has been computing the posterior distribution if its dimension, which equals the number of parameters, 43 in our case, is very high. Using the MCMC procedure, more specifically, the Gibbs sampler algorithm, allows us to simulate the posterior distribution in manageable blocks.

The parameters that need to be estimated are those in equation (2) which determines y_{it}^{*}. This includes the random intercept term η_i. We denote the vector of other coefficients to be $\boldsymbol{\beta}$. They include constant terms, the coefficients for the past arrests and past criminal experience, and coefficients for the observed individual characteristics described in Table 2 to Table 4. We allow $\boldsymbol{\beta}$ to take on different values depending on the type. The other parameters we estimate are the threshold parameters γ_2 to γ_4 and the variance of the random effects denoted d. We also estimate the prior type probability λ which can roughly be interpreted as the proportion of criminal type individuals in the population.

The estimation procedures are described in detail in Appendix 1. We set up the priors for d and $\{\boldsymbol{\beta}^{(s)}, \lambda^{(s)}\}_{s=1}^{2}$ for types $s = 1,2$ as follows.

the variance of the random effect d	$d \sim INVWISH_1\left(v_0, (q_0)^{-1}\right)$[13]
Coefficients on explanatory variables $\boldsymbol{\beta}$	$\boldsymbol{\beta}^{(s)} \sim N_{19}\left(\gamma_0^{(s)}, (B_0^{(s)})^{-1}\right)$
Type probability (λ_1, λ_2)	$(\lambda_1, \lambda_2) \sim D_2(\alpha_1, \alpha_2)$

where $INVWISH(v, G)$ is the Inverted Wishart distribution with the degree of freedom v and the scale matrix G, and $D(\alpha_1, \alpha_2)$ is the Dirichlet distribution with parameters α_1 and α_2. The prior distribution of the threshold parameters γ is assumed to be completely diffuse.

3. Data

The National Youth Survey (NYS) we use contains a random sample of 1,725 individuals from ages 11 to 17 in 1976. Individuals were surveyed annually from 1976 to 1980; then again in 1983 and 1986.

From 76 to 80 as well as for 83 and 86, we have data on the number of crimes committed in that year. But in 1981, 1982, and 1984, 1985, the exact numbers of crimes committed are not available. Instead, we only know in which of the four categories (0, 1-2, 3-11, 12-) their criminal activity falls, which was part of the reason for using the ordered probit approach. In order to construct the past criminal activity variable, we impute the number of crimes committed in each category to be the lower end point.[14] Moreover, only the criminal activity of last three years is used. This is a simple way to model the depreciation of past criminal activities. The results seem to be robust to using other specifications.

The data on arrests are limited in that total arrests in the past two years are available only for years 1983 and 1986. Total arrests to date are recorded for 1980 and 1983. To fill in the missing arrest records, we evenly spread the incremental arrests over the intervening years.

We only use the male sample in the data as males commit most of the crimes. There are 918 male individuals in the data set. We exclude individuals who have incomplete responses for arrests in 1980, 1983, or 1986 since we cannot recover the arrest variable for these individuals from the data. The total number of male individuals in the data set is 1725, Furthermore, we exclude individuals who have incomplete responses for crimes committed on 1978, 1979 and 1980 since for these individuals, we cannot recover the criminal experience variable from the data. Having done this gives us a final sample of 612 individuals**.**

Table 1-4 show some sample statistics and Table 5 contains information on the types of crimes included. Figure 1 depicts the age mean crime profile which differs from the conventional age arrest profile. It has a peak at age 13, and is flat thereafter. This is surprising since most age arrest profiles in the literature have a pronounced peak around age 18. Figure 2 depicts the modified age mean crime profile, where for each age, the top 3% of the individuals, i.e., those who committed the most crimes, are removed. After such outlier removal, the age crime mean profile is a smooth function of age, and has a peak around age 18. It resembles the conventional age arrest profiles more than the original profile suggesting that the self-reported crime data has some outliers. We also plot the percentage of individuals at each age who committed at least one crime during that year. From the age crime participation profile in Figure 3, we can see that in one year, about 30% of the individuals commit crimes at least once. We also can see that the variation of the age-crime participation profile is smaller than that of the age mean crime profile. This is because the variation in numbers of crimes committed is ignored if we only focus on the discrete choice of whether the person committed crime at least once or not during one year. That is exactly why we use an ordered probit model for estimation.

4. Estimation Results

Table 6 (Model 1) contains the parameter estimates as well as the standard errors. The two unobserved types are called type 1 and type 2. We term these the non-criminal type and the criminal type respectively as type 1 commits fewer crimes at all ages. The probability of the criminal type is estimated to be about 0.12. This is similar to the estimate found in Imai and Krishna (2004).

Other things constant, the crime rate for the non-criminal type increases with age before age 18 and decreases with age after age 18. For the criminal type, however, the crime rate increases with age even after age 18. Imai and Krishna's results (2004) suggest that the arrest rate for the criminal type decreases more rapidly with age than for the non-criminal type. However, this need not imply that the age crime profile does the same if the criminal type becomes better at avoiding arrests after the age 18. Our results show that this is indeed the case, at least in the National Youth Survey data. In Figure 4, we plot the age mean crime profiles of the non-criminal type and the criminal type. This confirms the results that for the criminal type, crimes increase after the age 18, especially after the age 22.

The criminal experience effect (i.e., the coefficient on the crimes committed over the last 3 years) is positive both for the non-criminal type and the criminal type. Before age 18, the criminal experience effect is larger for the criminal type than for the non-criminal type.

The arrest effect (i.e., the coefficient on the past arrest record) is positive and significant for the non-criminal type both before and after age 18. It is negative for the criminal type, though only significant after age 18. One interpretation of this is that the non-criminal type would be more stigmatized upon arrest than the criminal type. Thus, the stigma and training effect discussed earlier could dominate the perceived probability effect for non-criminal type individuals but would not do so for criminal type individuals. If a criminal type individual already has a bad reputation, then it is not surprising that his stigma effect is smaller.[15]

The coefficient on schooling (the highest grade attended by the individual) is positive for the non-criminal type and is negative though not significant for the criminal type. This does not imply that schooling increases crime, because current schooling is not included in the variables. It could be that past schooling is negatively correlated with current schooling and current schooling decreases crime. The coefficients on employment (the number of years of employment after age 18) are insignificantly different from zero. To fully understand the effect of schooling and employment, we need to estimate a model of criminal behavior where schooling and employment choices are endogenous.[16]

Coefficients on the family backgrounds for both types have the expected sign, parents' presence in the home has the expected effect, even though not significant, in that it reduces criminal activity. Also note that for both types, individuals whose parents graduated from high school seem to commit fewer crimes, perhaps as the families are better off and so the children have more to lose. Other coefficients have expected signs or are not significantly different from zero.

It is worth noting that it is not straightforward to interpret coefficients on the time-constant covariates such as race and family background. That is because they also contain the Wooldridge initial conditions corrections term. In other words, even if the coefficient of a race dummy is positive, it is unclear whether the individual belonging to the race is more likely to commit a crime or to have high criminal experience prior to the sample period, leading to high crimes in the future. Therefore, one should not over emphasize the differences in the estimates across types though they have reasonable signs for the most part.

In Table 6 (Model 2), we report the estimation results in which, as is often done in the crime literature, we include past arrests, but do not include past crimes committed, as explanatory variables. Notice both before and after 18, the arrest coefficients for the non-criminal type are positive though not always significant. When criminal activity is included, both coefficients are positive and significant. The arrest coefficients for the criminal type are negative both before and after 18 but are not significant. When criminal activity is included, both coefficients are negative and the coefficient after age 18 is significant.

We conclude that adding past criminal activity to the regressors increases the significance of the arrest coefficients, and in some cases, dramatically changes their magnitude, for example, for the non-criminal type and for the criminal type after age 18. This is likely due to the omitted variable bias. The direction of the bias is less clear because past crimes are likely to be correlated not only to past arrests, but also to age, and random effects term. Furthermore, it is also worth mentioning that the precision of the random effects term is larger for the original specification $(d^{-1} = 1.086)$ than that for the specification which does not include past criminal experience in the regressor $(d^{-1} = 0.455)$. This suggests that part of criminal behavior attributed to past crimes should be attributed to unobserved heterogeneities at the individual level

Table 7 shows the marginal effects of the coefficient estimates for our model, i.e., the effect of a marginal change of a regressor on the probability of zero crimes, crimes less than or equal to 2, crimes less than or equal to 11, and crimes less than or equal to 30. The marginal effect measures the change in probability due to a small change in the variables.[17] It is evaluated by setting the parameters at the posterior mean and variables at the sample mean. Hence, being black raises the probability (by .13) of committing no crimes. All the signs and magnitudes of the marginal effects make sense and are consistent with the parameter estimates. For example, for the non-criminal type, a unit increase in age decreases the probability of not committing any crimes before age 18 by .0293 but increases it after age 18 by .0061. Recall that the coefficient for age for the non-criminal type was positive below age 18 and negative after 18. Before and after the age 18, an increase in log arrests and log past crimes reduces the probability of no crimes committed.

For the criminal type, an increase in age reduces the probability of committing crimes less than or equal to 30 times a year, i.e., an increase in age increases the probability of committing crimes more than 30 times a year. This is not so for the non-criminal type. Table 7 indicates that for the non-criminal type, most of the changes occur at the margin of criminal participation. This is because most of the non-criminal types do not commit any crimes. On the other hand, for the criminal type, most of the changes occur around the probability of committing 11 to 30 crimes, as can be seen from the fact that changes occur disproportionately at the cumulative probability of committing less than 30 crimes. This again comes from the fact that criminal type typically commits more than 10 crimes per year. It is interesting to see clear differences between the non-criminal and the criminal types with respect to which range of the numbers of crimes committed the marginal change has the most impact on.

In addition to our original model, we estimate a model without incorporating unobserved types and compare it to the original model. Table 8 provides the estimation results. The after age 18 coefficient on past arrests in this model is estimated in between the coefficient for the criminal type and that for the non-criminal type in our original model. Thus, it is essential for an empirical model of crime to allow unobserved types to have different coefficients. In this respect, past literature that does not allow for unobserved types can fail to suggest effective policies as discussed in the introduction.

In Table 9, we report the proportions of crimes that the non-criminal type and the criminal type commit. Here, we classified an individual as a non-criminal type if the posterior probability of being a non-criminal type is more than 50%. As can be seen, criminal type individuals tend to commit crimes like carrying hidden weapons and selling drugs. Even though drug related crimes also occur for the non-criminal type, they are much less likely, while minor crimes such as stealing less than $50 are more common.

Finally, we look at the results of sample splitting, the simplistic approach to dealing with heterogeneity. Specifically, we split the sample into the two groups. We ordered individuals by the number of crimes committed and estimated the model for the top 12 percent and the remaining 88 separately. We chose the 12 percent break based upon the type probability of criminals being 12 percent in the mixture model. Table 10 provides the estimation results. The results show that coefficients on arrests and crimes tend not to be significant: only the criminal human capital coefficient for the low end of the sample is significant. This is because the assignment of types is based on total criminal activities, which is endogenous. For example, individuals who, despite the low past criminal experience, belong to the criminal type must have high current and future crimes, resulting in negative bias on the past crime coefficients for the criminal type. This confirms the hazards of such a simplistic approach.

5. Conclusion

In this paper we have focused on the determinants of criminal activity as opposed to arrests. Our main results are threefold.

First, there is reason to believe that criminal type individuals and non-criminal type individuals behave differently. Not allowing for such differences leads to misleading results: as shown, the estimates when we do not allow for two types differ considerably from those when we do. Second, we show that criminal behavior depends not just on past arrests, but also on past criminal experience. Neglecting this tends to bias the effect of arrests on crime upwards as discussed earlier. Third, we show that it is inappropriate to look at the age arrest profile and assume it reflects the age crime profile. The age arrest profile differs from the age crime profile and differs across types as well. Just because the arrest rate for the criminal type decreases more rapidly with age than that for the non-criminal type does not mean that criminal behavior does.

Our work has a number of policy implications. First, since criminal types differ from non-criminal types in their behavior, policies towards them should differ. Recall that the coefficient on arrests for criminal type is negative and significant after the age 18 and positive and significant for the non-criminal type. This suggests that arrests increase criminal behavior for the non-criminal type, but not for the criminal type. Thus, tough on crime policies for the criminal type individuals look more attractive than for the non-criminal type individuals, suggesting leniency towards first time offenders and harshness towards repeat offenders, such as the "three strikes and you are out" one might be on the right track.

Second, our work suggests that habitual criminals seem to be able to evade arrests. The age crime profile of criminal type individuals is increasing in age while their age arrest profile is decreasing after 18. There seems to be a significant criminal human capital effect for both types. This suggests that targeting police activity towards known criminals might be useful.

To get a better understanding of the factors affecting criminal behavior, we need to consider at least two additional elements, namely past convictions and the perceived probability of punishment. If we had the latter, then the coefficient on criminal activity would more clearly capture criminal human capital. Moreover, the coefficient on arrests would better capture any stigma associated with arrests alone. Finally, the coefficient on conviction would capture both learned criminal behavior and stigma due to conviction.

It would also be worthwhile constructing and estimating the determinants of arrests, conviction as well as the perceived probability of punishment[18] simultaneously with that of criminal choice. Such a setup would help cast light on questions like: is there a race bias in arrests and convictions? Are habitual criminals indeed able to better evade arrest? If the coefficient on past criminal activity for criminal types in the arrest equation is negative, this might suggest habitual criminals are indeed better at evading arrests. If the coefficient on past arrests for them is positive, a "rap sheet effect" is suggested, i.e., known criminals are more likely to be caught.

6. References

Albert, James H., and Siddhartha Chib, "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88 (1993), 669-679.

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

Bayer, Patrick, Randi Pintoff, and David Posen, "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections," Mimeo, Yale University (1994).

Burdett, Kenneth, Ricardo Lagos, and Randall Wright, "An On-the-Job Search Model of Crime, Inequality, and Unemployment," *International Economic Review*, 45 (2004), 681-706.

Chay, Kenneth C., and Dean Hyslop, "Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence using Alternative Approaches," Mimeo, U.C. Berkeley (2004).

Geweke, John, Michael Keane, and David Runkle, "Statistical Inference in the Multinominal Multiperiod Probit Model," *Journal of Econometrics*, 80 (1997), pp. 125-165.

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

Heckman, James, "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time Discrete Data Stochastic Process," in C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Massachusetts (1981).

Heckman, James, and Burton Singer, "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52:2 (1984), 271-320.

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

İmrohoroğlu, Ayse, Antonio Merlo, and Peter Rupert, "What Accounts for the Decline in Crime," *International Economic Review*, 45 (2004), 707-730.

Levitt, Steven, and Lance Lochner, "The Determinants of Juvenile Crime," mimeo (2000).

Lochner, Lance, "Education, Work, and Crime: A Human Capital Approach," *International Economic Review*, 45 (2004), 811-844.

__________, “An Empirical Study of Individuals Perceptions of the Criminal Justice System,” Mimeo, University of Rochester (2004).

Mocan, Naci H., and Daniel I. Rees, “Economic Conditions, Deterrence, Determinant and Juvenile Crime: Evidence from Micro Data,” NBER Working Paper, No. 7405 (1999).

Tauchen, Helen, Ann D. Witte, and Harriet Griesinger, “Criminal Deterrence: Revisiting the Issues with a Birth Cohort,” *Review of Economics and Statistics*, 76:3 (1994), 399-412.

Viscusi, Kip W., “Market Incentives for Criminal Behavior,” in R. Freeman and H. Holzer (eds.), *The Black Youth Employment Crisis*, University of Chicago Press, Chicago, Illinois (1986).

Waldfogel, Joel, “The Effect of Criminal Conviction on Income and the Trust Reposed in the Workmen,” *Journal of Human Resources*, 29:1 (1994), 62-81.

Wooldridge, Jeffrey M., “A Framework for Estimating Dynamic, Unobserved Effects Panel Data Models with Possible Feedback to Future Explanatory Variables,” *Economics Letters*, 68 (2000), 245-250.

__________, *Econometric Analysis of Cross Section and Panel Data,* MIT Press (2002).

Appendix 1: Estimation Procedure

The Bayesian Markov Chain Monte Carlo (MCMC) algorithm is straight forward and avoids the difficulties inherent in dealing with posterior distributions which are complex high dimensional functions of the parameters. It simulates the posterior distribution by drawing sequentially from a series of conditional distributions of lower dimension which converge in distribution to the posterior distribution.

We also include a simulation step termed "Data Augmentation". This simulates the latent variable y_{it}^* which is a latent variable and not observable. Once the latent variable y_{it}^* is drawn, the Ordered Probit estimation becomes a straightforward Bayesian linear regression estimation.

Let us denote $Z \in \{1,2\}$ to be the type of the individual. The variables that need to be simulated are: $(Z, y^*, \eta, \beta, d^{-1}, \lambda, \gamma)$. y and X are given in the data. The MCMC algorithm in this case is a chain of the following steps:

Set the initial values $[Z^{(0)}, y^{*(0)}, \eta^{(0)}, \beta^{(0)}, d^{-1(0)}, \lambda^{(0)}, \gamma^{(0)}]$

STEP 1: Draw $Z^{(1)}$ given $[y^{*(0)}, \eta^{(0)}, \beta^{(0)}, d^{-1(0)}, \lambda^{(0)}, \gamma^{(0)}, y, X]$

STEP 2: Draw $y^{*(1)}$ given $[Z^{(1)}, \eta^{(0)}, \beta^{(0)}, d^{-1(0)}, \lambda^{(0)}, \gamma^{(0)}, y, X]$

STEP 3: Draw $\eta^{(1)}$ given $[Z^{(1)}, y^{*(1)}, \beta^{(0)}, d^{-1(0)}, \lambda^{(0)}, \gamma^{(0)}, y, X]$

STEP 4: Draw $\beta^{(1)}$ given $[Z^{(1)}, y^{*(1)}, \eta^{(1)}, d^{-1(0)}, \lambda^{(0)}, \gamma^{(0)}, y, X]$

STEP 5: Draw $d^{-1(1)}$ given $[Z^{(1)}, y^{*(1)}, \eta^{(1)}, \beta^{(1)}, \lambda^{(0)}, \gamma^{(0)}, y, X]$

STEP 6: Draw $\lambda^{(1)}$ given $[Z^{(1)}, y^{*(1)}, \eta^{(1)}, \beta^{(1)}, d^{-1(1)}, \gamma^{(0)}, y, X]$

STEP 7: Draw $\gamma^{(1)}$ given $[Z^{(1)}, y^{*(1)}, \eta^{(1)}, \beta^{(1)}, d^{-1(1)}, \lambda^{(1)}, y, X]$

Repeat Step 1 to Step 7 with the new initial values $\left[Z^{(1)}, y^{*(1)}, \eta^{(1)}, \beta^{(1)}, d^{-1(1)}, \lambda^{(1)}, \gamma^{(1)}\right]$.

Below, we discuss in more details the simulation algorithms of each conditional draws.

1. Draw $\left[z_i \mid y^*, \eta, \beta, d^{-1}, \lambda, \gamma, y, X\right]$

Let $P(i,s)$ be the probability that individual i is of type s for $s = 1,2$. Recall that λ_s denotes the type probability. Hence,

$$P(i,s) = \frac{\prod_{t=1}^{T} f\left(y_{it}^*, \eta_i \mid z_i = s, \beta, d^{-1}, \lambda, \gamma, y\right)\lambda_s}{\sum_{m=1}^{S} \prod_{t=1}^{T} f\left(y_{it}^*, \eta_i \mid z_i = m, \beta, d^{-1}, \lambda, \gamma, y\right)\lambda_m}$$

where

$$\begin{aligned} f\left(y_{it}^*, \eta_i \mid z_i = s, \beta, d^{-1}, \lambda, \gamma, y\right) = {} & f\left(y_{it}^* \mid z_i = s, \eta_i, \beta, d^{-1}, \lambda, \gamma, y\right) \\ & \times f\left(\eta_i \mid z_i = s, \beta, d^{-1}, \lambda, \gamma, y\right) \end{aligned}$$

$$f\left(y_{it}^* \mid z_i = s, \eta_i, \beta, d^{-1}, \lambda, \gamma, y\right) \propto \exp\left[-\frac{1}{2}\left(y_{it}^* - \eta_i - x_{it}'\beta^{(2)}\right)^2\right]$$

$$f\left(\eta_i \mid z_i = s, \beta, d^{-1}, \lambda, \gamma, y\right) \propto \frac{1}{\sqrt{d^{(s)}}} \exp\left[-\frac{1}{2}\frac{\eta_i^2}{d^{(s)}}\right]$$

The full conditional density of z_i is:

$$\sum_{s=1}^{S} I(z_i = s) P(i,s)$$

which is used to draw the value for z_i.

2. Draw $\left[y_{it} \mid Z, \eta, \beta, d^{-1}, \lambda, \gamma, y, X\right]$

The full conditional density of y_{it}^* is as follows. Given the draw for the type just obtained, and the remainder of the parameters and data, we obtain the conditional density function of y_{it}^* to be

$$I\left(\gamma_{j-1} < y_{it}^* \le \gamma_j\right)\phi\left(y_{it}^* \mid \eta_i + x_{it}'\beta^{(s)}, 1\right)$$

where $I(\cdot)$ is an indicator function and $\phi\left(y_{it}^{*} \mid \eta_i + x_{it}'\beta^{(s)}, 1\right)$ is a normal distribution with mean $\eta_i + x_{it}'\beta^{(s)}$ and a variance of unity. Thus, draw y_{it}^{*} from the normal distribution, $N\left(\eta_i + x_{it}'\beta^{(s)}, 1\right)$ truncated at the left by γ_{j-1} and to the right by γ_j.

3. Draw $\left[\eta_i \mid Z, y^{*}, \beta, d^{-1}, \lambda, \gamma, y, X\right]$:

The full conditional density of η_i is

$$\phi\left(\eta_i \mid 0, d^{(s)}\right)\prod_{t=1}^{T}\phi\left(y_{it}^{*} \mid \eta_i + x_{it}'\beta^{(s)}, 1\right)$$

Thus, draw η_i from $N\left(\hat{\eta}_i, \hat{d}_i\right)$ where

$$\hat{\eta}_i = \hat{d}_i\left(\sum_{t=1}^{T}\left(y_{it}^{*} - x_{it}'\beta^{(s)}\right)\right)$$

$$\hat{d}_i^{-1} = \left(\left(d^{(s)}\right)^{-1} + T\right).$$

4. Draw $\left[\beta^{(s)} \mid Z, y^{*}, \eta, d^{-1}, \lambda, \gamma, y, X\right]$:

The full conditional density of $\beta^{(s)}$ is the same as the posterior density for the regression parameter in the normal linear model

$$Y^{**(s)} = X^{(s)}\beta^{(s)} + \varepsilon^{(s)}, \varepsilon^{(s)} \sim N_{n(s)T}(0, I)$$

when the types are separated and where $n_{(s)}$ denotes the number of individuals of type s which defines the dimension of the above normal distribution. $X^{(s)}$ and $Y^{(s)}$ are the stacked matrix and vector for the separated types. That is

$$X^{(s)} = \left(x_{i_1 1}', \ldots, x_{i_1 T}', \ldots, x_{i_{n(s)} 1}', \ldots, x_{i_{n(s)} T}'\right)'$$

$$Y^{**(s)} = \left(y_{i_1 1}^{**}, \ldots, y_{i_1 T}^{**}, \ldots, y_{i_{n(s)}}^{**}, \ldots, y_{i_{n(s)} T}^{**}\right)'$$

where the new variable y_{it}^{**} is just y_{it}^{*} displaced by the random term, i.e.,

$y_{it}^{**} = y_{it}^{*} - \eta_i$.

If $\beta^{(s)}$ is assigned the proper conjugate $N_k\left(\beta_0^{(s)}, B_0^{(s)}\right)$ prior, then the we draw $\beta^{(s)}$ from

$N_k\left(\hat{\beta}^{(s)}, \hat{B}^{(s)}\right)$ where

$$\hat{\beta}^{(s)} = \left(B_0^{(s)} + X^{(s)\prime} X^{(s)}\right)^{-1}\left(B_0^{(s)}\beta_0^{(s)} + X^{(s)\prime} Y^{**(s)}\right)$$

$$\hat{B}^{(s)} = \left(B_0^{(s)} + X^{(s)\prime} X^{(s)}\right)^{-1}$$

5. Draw $\left[(d)^{-1} \mid Z, y^*, \eta, \beta, \lambda, \gamma, y, X\right]$:

The full conditional density of $(d)^{-1}$ is

$$f_w\left((d)^{-1} \mid v_0, q_0\right)\prod_{i:z_i=s}\phi(\eta_i \mid 0, d \quad)$$

Thus, we draw $(d \quad)^{-1}$ from

$$Wish\left(\sum_{i=1}^{N} I(z_i = s) + \nu_0, \left[(q_0)^{-1} + \sum_{i=1}^{N}\eta_i^2\right]^{-1}\right)$$

6. Draw $\left[\lambda \mid Z, y^*, \eta, d^{-1}, \beta, \gamma, y, X\right]$:

The full conditional density of λ is

$$f_D\left(\lambda \mid \{\alpha_s\}_{s=1}^{S}\right)\prod_{i=1}^{N}\sum_{s=1}^{S} I(z_i = s)\lambda^{(s)}$$

Thus, we sample λ from the Dirichlet distribution

$$D(\alpha_1 + n_1, \ldots, \alpha_s + n_s)$$

where n_i denotes the number of individuals of type *i*, i.e.,

$$\sum_{i=1}^{N} I(z_i = s) = n_s$$

7. Draw $\left[\gamma_j \mid Z, y^*, \eta, d^{-1}, \beta, \lambda, y, X\right]$:

For a formal derivation of this see Albert and Chib (1993). Since the draws of y^* contain all the information available about γ_j, the conditional density of γ_j is uniform on the following interval:

$$\left[\max\left\{\max\left\{y_{it}^* : y_{it} = j\right\}, \gamma_{j-1}\right\}, \min\left\{\min\left\{y_{it}^* : y_{it} = j+1\right\}, \gamma_{j+1}\right\}\right].$$

For estimation, we set up the prior distributions as follows: $d \sim INVWISH(3,0.1)$, $\beta \sim N(0, I)$ for type 1 and type 2, and $(\lambda_1, \lambda_2) \sim D(1,1)$. We run 20, 000 MCMC iterations. Convergence is achieved within the first 3000 iterations. The first 5,000 iterations are used as burn-in samples, and the next 20,000 iterations are used for computing the posterior distributions.

Table 1. Starting Age

Age	# individuals
11	98
12	96
13	87
14	86
15	85
16	90
17	70

Table 2. Race

Percent white	79.9
Percent black	15.7
Percent others	4.4

Table 3. Parental Background

Percent Both Parents High School Grads	55.23
Percent Poor	24.84
Percent Both Parents Living Together	71.57
Percent 4 years or More Neighborhood	67.16
Percent Rural	30.88
# Youths In the Household	2.84

Table 4. Cumulative Arrests

year	mean	stdev
1980	0.444	3.425
1983	0.717	2.994
1986	0.889	3.265

Table 5. Proportions of Crime Categories

Crime	Proportion
stolen motor vehicle	0.032
stolen > $50	0.037
bought stolen goods	0.049
carried hidden weapon	0.216
stolen < $5	0.077
attacked someone	0.039
been paid for sexual relations	0.058
been in gang fights	0.041
sold marijuana	0.160
hit parent	0.036
sold hard drugs	0.075
taken vehicle	0.036
sexual assault	0.064
stole things $5 < < $50	0.043
broken into a building	0.038

Table 6: Estimation Results

	Model 1				Model 2			
	Type 1 (Non-Criminal)		Type 2 (Criminal)		Type 1 (Non-Criminal)		Type 2 (Criminal)	
	mean	Stdev	mean	stdev	mean	stdev	mean	stdev
constant*D(age<18)	-0.632	0.444	**3.604**	1.979	-0.458	0.702	2.488	1.537
constant*D(age≥18)	-0.433	0.480	-0.199	1.810	-0.089	0.755	1.432	1.631
black	**-0.692**	0.184	0.806	0.741	**-1.056**	0.309	-0.001	0.858
others	0.059	0.273	-1.781	2.573	0.293	0.428	-0.956	1.503
age* D(age<18)	**0.121**	0.063	0.211	0.327	0.102	0.088	**0.363**	0.189
age* D(age≥18)	-0.025	0.067	**2.678**	0.740	-0.083	0.093	**0.679**	0.407
arrest* D(age<18)	**0.280**	0.165	-0.215	0.751	**0.381**	0.210	-0.236	0.540
arrest* D(age≥18)	**0.163**	0.091	**-1.772**	0.575	0.114	0.165	-0.544	0.602
past crime* D(age<18)	**0.250**	0.050	**0.446**	0.196				
past crime* D(age≥18)	**0.245**	0.037	0.052	0.105				
work experience	-0.025	0.071	**-1.674**	0.697	-0.087	0.097	0.074	0.375
highest grade	**0.175**	0.069	-0.115	0.343	0.143	0.089	0.254	0.208
crime(-1)	**0.002**	0.001	0.011	0.010	**0.004**	0.001	0.023	0.025
time mean work experience	0.052	0.051	**-1.287**	0.318	**0.182**	0.080	**-0.916**	0.226
time mean highest grade	**-0.212**	0.075	-0.003	0.352	**-0.178**	0.100	-0.405	0.250
parents high school grad	-0.022	0.125	**-2.930**	0.791	-0.290	0.199	-0.536	0.544
poor	0.161	0.154	**-1.362**	0.747	0.172	0.247	-0.142	0.740
# youth	0.019	0.039	-0.148	0.203	0.035	0.065	-0.036	0.185
living together	-0.067	0.138	**-2.636**	0.736	-0.087	0.239	**-1.370**	0.609
4 years	-0.064	0.129	**-1.660**	0.616	-0.041	0.214	-0.387	0.561
rural	**-0.497**	0.149	**2.005**	0.632	**-0.911**	0.214	**1.141**	0.561
probability (λ)	**0.880**	0.023	**0.120**	0.023	**0.791**	0.047	**0.209**	0.047

	Mean	stdev	mean	stdev
d^{-1}	**1.086**	0.17	**0.455**	0.056
γ_2	**0.661**	0.03	**0.722**	0.028
γ_3	**1.480**	0.06	**1.594**	0.058
γ_4	**2.605**	0.09	**2.796**	0.097

Table 7: Marginal Effects of Changes in the Regressors (Model 1)

	Type 1 (Non-Criminal Type)				Type 2 (Criminal Type)			
	$crime = 0$	$crime \leq 2$	$crime \leq 11$	$crime \leq 30$	Crime=0	$crime \leq 2$	$crime \leq 11$	$crime \leq 30$
black	0.1300	0.0480	0.0077	0.0001	-7.87E-08	-2.28E-06	-0.0001	-0.0039
others	-0.0147	-0.0062	-0.0012	-0.0001	1.00E-04	1.30E-03	0.0143	0.1420
age* D(age<18)	-0.0293	-0.0122	-0.0022	0.0000	-4.65E-08	-1.30E-06	0.0000	-0.0018
age* D(age≥18)	0.0061	0.0026	0.0005	0.0001	-5.90E-07	-1.65E-05	-0.0005	-0.0234
arrest* D(age<18)	-0.0679	-0.0282	-0.0052	-0.0002	4.73E-08	1.33E-06	0.0000	0.0018
arrest* D(age≥18)	-0.0395	-0.0164	-0.0030	-0.0001	3.90E-07	1.09E-05	0.0004	0.0156
past crime* D(age<18)	-0.0606	-0.0252	-0.0046	-0.0001	-9.82E-08	-2.75E-06	-0.0001	-0.0039
past crime* D(age≥18)	-0.0594	-0.0247	-0.0045	-0.0001	-1.14E-08	-3.19E-07	0.0000	-0.0004
work experience	0.0061	0.0026	0.0005	0.0001	3.69E-07	1.03E-05	0.0003	0.0146
highest grade	-0.0424	-0.0176	-0.0032	-0.0001	2.53E-08	7.07E-07	0.0000	0.0010
crime(-1)	-0.0005	-0.0002	0.0000	0.0000	-2.42E-09	-6.76E-08	0.0000	-0.0001
time mean work experience	-0.0126	-0.0052	-0.0009	0.0000	2.83E-07	7.91E-06	0.0003	0.0113
time mean highest grade	0.0514	0.0214	0.0040	0.0002	6.60E-10	1.85E-08	0.0000	0.0000
parents high school grad	0.0053	0.0022	0.0004	0.0000	2.49E-05	3.25E-04	0.0049	0.0732
poor	-0.0406	-0.0174	-0.0033	-0.0001	6.97E-06	1.07E-04	0.0021	0.0400
# youth	-0.0046	-0.0019	-0.0003	0.0000	3.26E-08	9.11E-07	0.0000	0.0013
living together	0.0165	0.0069	0.0013	0.0001	1.93E-06	3.78E-05	0.0008	0.0220
4 years	0.0157	0.0066	0.0012	0.0000	7.06E-07	1.57E-05	0.0004	0.0132
rural	0.1094	0.0436	0.0078	0.0004	-1.02E-06	-2.17E-05	-0.0005	-0.0160

Table 8: Estimation Results: Non-Mixture Model

	Mean	stdev
constant*D(age<18)	-0.336	0.374
constant*D(age≥18)	-0.269	0.394
black	**-0.444**	0.148
others	-0.033	0.230
age* D(age<18)	**0.103**	0.056
age* D(age≥18)	0.031	0.061
arrest* D(age<18)	0.165	0.142
arrest* D(age≥18)	0.056	0.077
past crime* D(age<18)	**0.288**	0.039
past crime* D(age≥18)	**0.271**	0.030
work experience	0.014	0.064
highest grade	**0.152**	0.060
crime(-1)	**0.001**	0.001
time mean work experience	-0.027	0.041
time mean highest grade	**-0.204**	0.067
parents high school grad	-0.115	0.099
poor	0.069	0.122
# youth	0.022	0.032
living together	**-0.237**	0.111
4 years	-0.084	0.099
rural	**-0.240**	0.108
d^{-1}	**1.322**	0.205
γ_2	**0.584**	0.029
γ_3	**1.280**	0.048
γ_4	**2.243**	0.064

Mean and stdev denote the mean and the standard deviation of the posterior distribution, respectively. Bold coefficient estimates indicate significance at the 10% level.

Table 9: Proportions of Crimes for the Non-Criminal Type and the Criminal Type

crime	non-criminal	Criminal
stolen motor vehicle	0.045	0.015
stolen > $50	0.051	0.017
bought stolen goods	0.054	0.023
carried hidden weapon	0.197	0.408
stolen < $5	0.091	0.026
attacked someone	0.047	0.023
been paid for sexual relations	0.046	0.015
been in gang fights	0.046	0.018
sold marijuana	0.122	0.234
hit parent	0.045	0.015
sold hard drugs	0.064	0.142
taken vehicle	0.047	0.015
sexual assault	0.045	0.016
stole things $5 < < $50	0.053	0.018
broken into a building	0.048	0.016

Table 10: Estimation Results of Splitting the Sample

	Bottom 88% Individuals		Top 12% Individuals	
	mean	Stdev	Mean	stdev
constant*D(age<18)	-0.414	0.374	1.888	1.437
constant*D(age≥18)	-0.291	0.400	1.941	1.436
black	**-0.341**	0.150	-0.952	0.669
others	-0.010	0.270	-0.158	0.742
age* D(age<18)	0.068	0.058	**0.345**	0.187
age* D(age≥18)	-0.004	0.068	0.054	0.148
arrest* D(age<18)	0.070	0.156	0.155	0.390
arrest* D(age≥18)	-0.053	0.100	-0.208	0.182
past crime* D(age<18)	**0.318**	0.051	-0.085	0.109
past crime* D(age≥18)	**0.314**	0.038	-0.007	0.066
work experience	0.039	0.072	-0.057	0.155
highest grade	**0.145**	0.066	0.271	0.171
crime(-1)	**0.031**	0.011	0.002	0.001
time mean work experience	-0.042	0.043	0.119	0.214
time mean highest grade	**-0.198**	0.073	**-0.400**	0.215
parents high school grad	-0.091	0.102	-0.279	0.531
poor	-0.044	0.130	0.715	0.565
# youth	0.019	0.034	0.108	0.166
living together	**-0.213**	0.122	-0.559	0.478
4 years	-0.038	0.109	-0.431	0.506
rural	**-0.182**	0.107	-0.289	0.595
d^{-1}	**1.594**	0.272	**0.324**	0.095
γ_2	**0.595**	0.032	**0.523**	0.066
γ_3	**1.344**	0.042	**1.246**	0.089
γ_4	**2.226**	0.080	**2.568**	0.131

Mean and stdev denote the mean and the standard deviation of the posterior distribution, respectively. Bold coefficient estimates indicate significance at the 10% level.

Figure 1. Age Mean Crime Profile

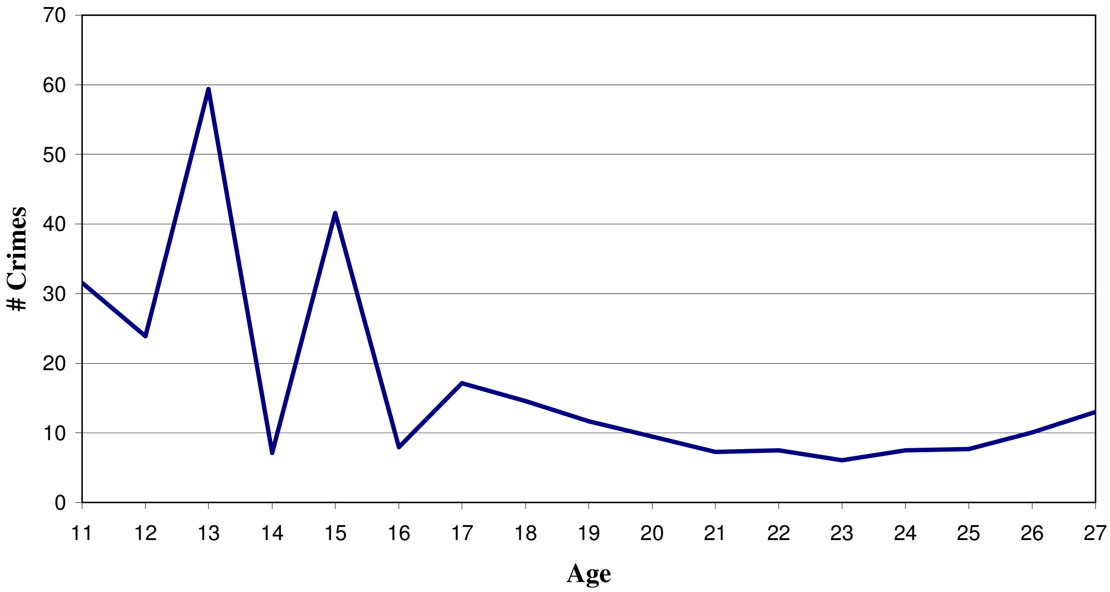


Figure 2. Age Mean Crime Profile (Top 3% Individuals Excluded)

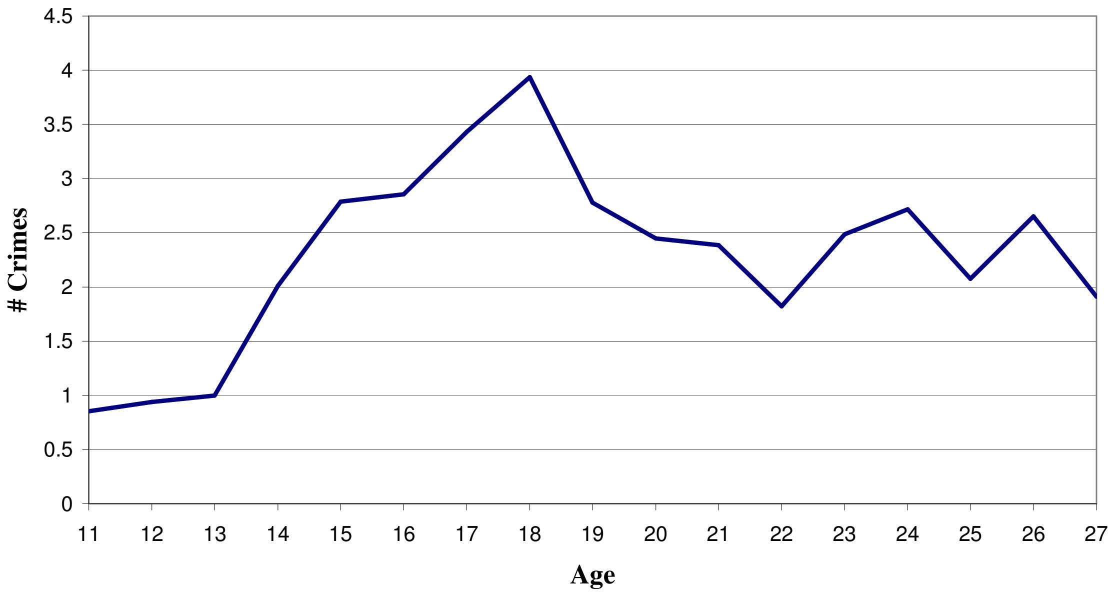


Figure 3. Age Crime Participation Profile

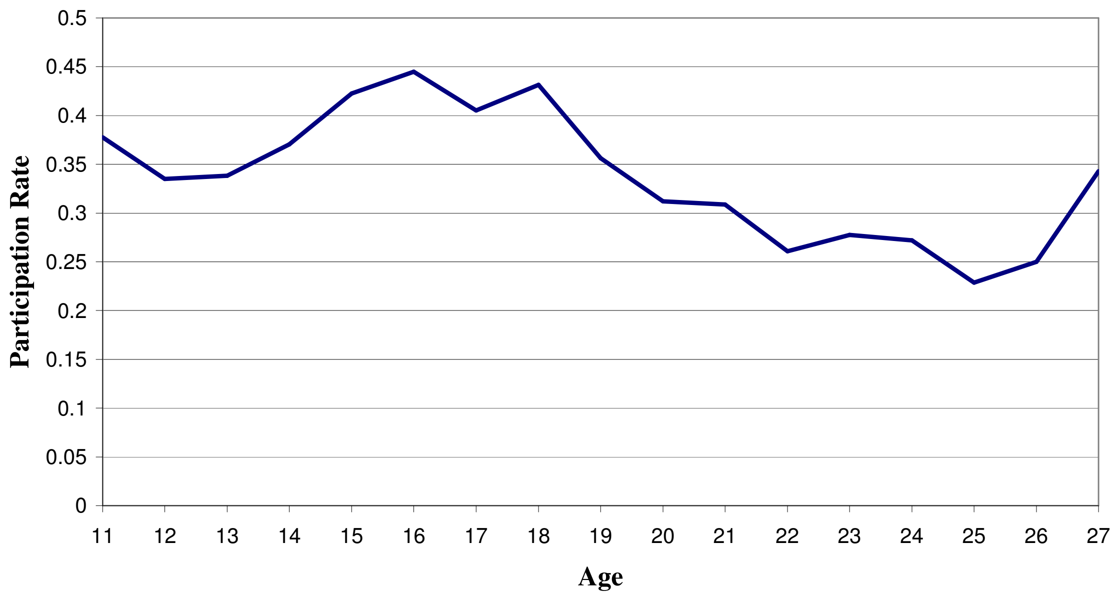
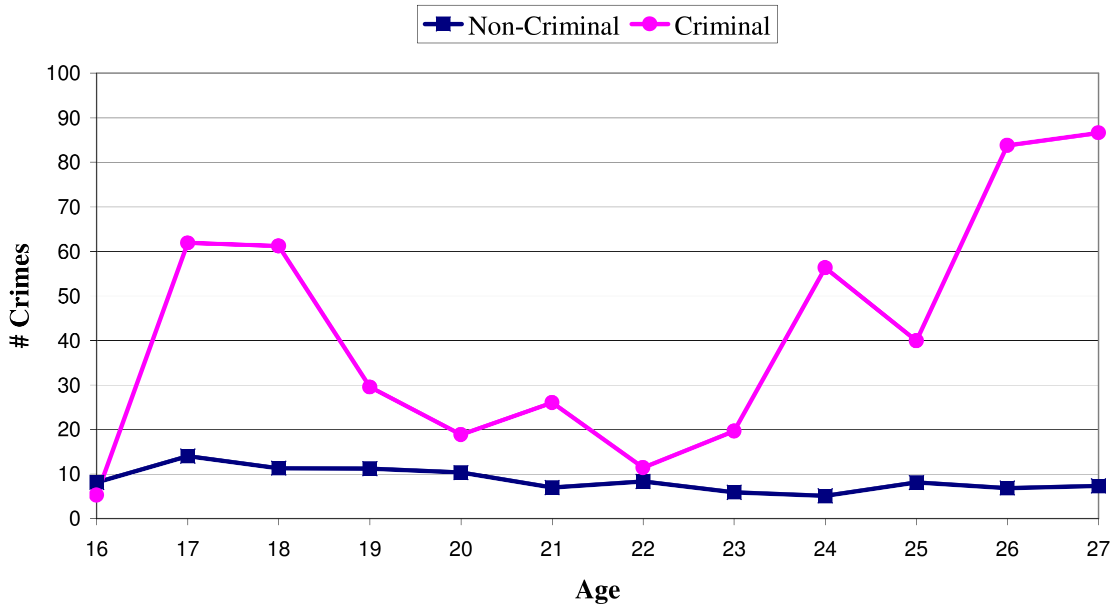


Figure 4. Age Mean Crime Profile



[1] Our work is empirical rather than theoretical. In addition to classic work, such as Becker (1968), there has been some interesting recent work on some of the issues we examine. For example, Burdett, Lagos and Wright (2004) develop a search model where unemployment, income inequality and crime are all endogenously determined and look at the effects of several policies to combat crime. İmrohoroğlu, Merlo and Rupert (2004) calibrate a general equilibrium model of crime to see what accounts for the fall in crime in the 90's.

[2] For a more detailed discussion, see Heckman and Singer (1984).

[3] A Tobit setting also uses this information. A count data setting does so as well and in addition respects the fact that only integral numbers of crimes occur. Our choice of the ordered probit setting is based on data considerations. In some years, information on criminal activity is not complete: only the range of criminal activity is given.

[4] If we estimate the ordered probit without allowing for unobserved heterogeneity in coefficients, we find a positive effect of arrests on criminal activity since non-criminal types dominate.

[5] Note the contrast to all existing work that uses age arrest data and finds the age arrest profile peaks at around 18.

[6] Past arrest record variable is: $\log(\text{number of past arrests} + 1)$.

[7] Past crimes committed variable is: $\log(\text{number of past crimes committed} + 1)$.

[8] The state dependence effect of past criminal experience on current criminal behavior is separately identified from the individual heterogeneity through the timing of past crimes committed. See Chay and Hyslop (2004) for more detailed discussions on identification.

[9] Since packaged programs cannot deal with this, we had to write the required program.

[10] It is tempting to use fixed effects in combination with a mixture model so that one need not assume strict exogeneity of education and work as done here. However, at present we are unaware of the existence of this technique.

[11] Lochner (2004) empirically investigates the joint relationship between schooling, work and crime.

[12] Notice that this assumption will give us some bias if the arrest probability depends on the individual unobserved characteristics.

[13] Subscripts on the density function denote the dimension of the random variables.

[14] We could have alternatively used the mean number in each category, but this rescaling would create problems at the upper end since some individuals commit many crimes inordinately affecting the mean. All the crime data in the tables below is based on this imputation.

[15] This is consistent with the work of Waldfogel (1994) who looks at the effect of past convictions on current wages. He finds that decreases in wage are more substantial for individuals in jobs that require credibility than in jobs that do not. If non-criminal type individuals are in jobs that require relatively more credibility compared to criminal type individuals, then our finding that the non-criminal type is more stigmatized is consistent with his results.

[16] Lochner (1999) estimated a simple dynamic lifecycle model where individuals jointly choose schooling, work and crime

[17] The formula for the marginal effect is similar to that for the probit model. That is, the marginal effect of the probability of the crime category being less than or equal to j is, $\frac{\partial prob(y \leq j \mid x = \bar{x})}{\partial x} = \phi\left(\bar{x}'\bar{\beta} - \bar{\mu}_j\right)\bar{\beta}$, where ϕ is the density function of standard normal distribution, $(\bar{\beta}, \bar{\mu}_j)$ is the posterior mean of the parameter estimates and $\bar{x}$ is the sample mean of variables.

[18] The perceived probability of punishment is not available and only that of arrest and the former has to be estimated as part of the model.