Delinquency and Gender*

Guyonne Kalb
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
and
Jenny Williams
Department of Economics
The University of Adelaide

Melbourne Institute Working Paper No. 13/01

ISSN 1328-4991

ISBN 0 7340 1517 8

November 2001

* We would like to thank Tim Fry for helpful discussion and suggestions and John Creedy for his comments. The Economics department at the University of Adelaide provided funding for travel and accommodation for us to work jointly on this project. Jenny Williams thanks the Economics department at the University of Illinois at Chicago for hosting her while part of this paper was written. Responsibility for any errors remains with the authors.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 8344 3701**
***Fax* (03) 8344 5630**
***Email* melb.inst@iaesr.unimelb.edu.au**
***WWW Address* http://www.melbourneinstitute.com**

Abstract

This paper investigates the determinants of juvenile delinquency for males and females using the 1958 Philadelphia Birth Cohort Study. Ordered probit models for juvenile arrest are estimated separately for males and females. An adaptation of the EM-algorithm is used to estimate the model for females in order address a problem of missing values for the variable linking demographic and arrest data. The results indicate that juvenile arrests for both males and females are more likely for non-whites and for those who leave education early. Males and females behave differently, in that males are more likely to be repeat offenders.

1. Introduction

This paper investigates the factors that are commonly believed to increase the risk of juvenile delinquency. From a policy perspective, this is of importance given that 32 per cent of people arrested for a crime against property are juveniles.[1] Furthermore, juvenile delinquency is seen as the most common pathway to adult criminality.[2]

Quantitative studies, which focus on the determination of juvenile delinquency, are rare in the economics literature. We are only aware of two papers, Phillips and Votey (1987) and Levitt and Lochner (2001). The criminology literature provides some salient studies on the determinants of juvenile delinquency such as, Good, Pirog-Good and Sickles (1986) and Laub and Sampson (1988). The former paper focuses on the crime-work nexus, while the latter investigates the role of family background and family process in causing delinquency. Furthermore, research using individual level data is almost exclusively based on males (see for example, Witte and Tauchen, 1994; Grogger, 1998; Freeman, 1991, 1996; Williams and Sickles, 2000).[3] Widom (2000) and Ritchie (2000) suggest that the major factors contributing to delinquency and subsequent criminality in females and males differ. In particular, physical and sexual abuse is identified as the primary cause of delinquency in girls. This has lead to calls for the criminal justice system to define justice for females and males differently, due to the victim status of female offenders. We attempt to shed some light on this issue, by carrying out a comparative analysis of the determinants of delinquency in males and females.

---

[1] See Snyder (2000).
[2] See for example Williams and Sickles (2000); Grogger (1998); Witte and Tauchen (1994); Good, Pirog-Good and Sickles (1986).
[3] Jurik (1983) is an exception.

The empirical analysis draws on data from the Delinquency in a Birth Cohort II study. These data were collected for the purpose of examining delinquent activities of a birth cohort, and contain juvenile arrest records for the population of individuals born in the city of Philadelphia in 1958. The analysis is based on a sample of males and females from this population, who were administered a retrospective survey in 1988. The retrospective survey includes detailed information on family composition during childhood, and the respondent's history of physical and sexual abuse, in addition to more standard variables such as the respondent's education. This rich source of individual level background information for the sample, along with the complete official records of juvenile arrests make this data well suited to investigate the determinants of juvenile delinquency.

Estimation of the model of juvenile delinquency for females is complicated by missing observations on the variable which links criminal justice arrest records and follow-up survey information. As a consequence 30 per cent of women in the sample cannot be uniquely matched to criminal justice records. The issues arising in estimating a model for juvenile arrests from these "scrambled data" are similar in nature to those generated by grouped or censored data. We therefore use an adaptation of the EM algorithm[4] to estimate the model for female juvenile delinquency rather than maximizing the likelihood directly. A second contribution of this paper is that it shows how the observations of female respondents in this data set can be used in analysis. So far, only the male observations have been used.

[4] EM stands for the two steps involved in this method an Expectation and a Maximization step. Dempster, Laird and Rubin (1977) were the first to recognize the usefulness of the method for incomplete data.

The next section describes the data used in the analysis. Section 3 discusses the methodology and Section 4 presents the results from estimation. The paper finishes with some concluding remarks in Section 5.

2. Data

2.1 Description of the 1958 Philadelphia Birth Cohort Study

Inclusion in the 1958 Philadelphia Cohort Study (Figlio, Tracy, and Wolfgang, 1991) is based on the criteria of being born in 1958 and living in Philadelphia between one's tenth and eighteenth birthdays. The 27,160 members of this universe were identified using the Philadelphia school census, the U.S. Bureau of Census, and school records. Once the members of the cohort were identified, data collection by Figlio and his team occurred in two phases.

The first phase involved obtaining the complete official criminal history of the cohort. These data were collected between 1979 and 1984 and cover the criminal careers, as recorded by the police, and juvenile and adult courts, for all 27,160 members of the cohort.[5] The second stage of the Study entailed a retrospective follow-up survey for a sample from the cohort. Figlio, Tracy and Wolfgang (1991) employed a stratified sampling scheme to ensure that they captured the most relevant background and juvenile arrest characteristics of the cohort, and yield a sample size sufficient for analysis. The male population was stratified by race, socio-economic status, arrest history (0, 1, 2-4, 5 or more arrests), and juvenile "status" arrests, which are arrest categories only applicable to individuals less than

[5] The information for juveniles was obtained from the Philadelphia police, Juvenile Aid Division (JAD). Once individuals reach the age of 18, police encounters are recorded on regular police forms (rap sheets) and reported to the FBI. Information about adult arrests was obtained from the Philadelphia Police Department, the Common and Municipal Courts, and the FBI, ensuring arrests both within and outside the boundaries of Philadelphia are included in the data set.

18 years of age. The female population was stratified by race, socio-economic status, and arrest history (0, 1, 2 or more arrests). From the resulting strata, a sample of males and females were randomly selected, with equal draws from all strata. This framework oversamples from the more sparsely populated strata covering repeat and chronic juvenile offenders. Of the 1,992 individuals randomly selected for the survey, 575 men and 201 women were interviewed in 1988. Areas of inquiry covered by the survey, which are of interest to this research, include composition of current and childhood households, history of sexual and physical abuse, personal and social demographic characteristics, parent's education, respondent's education, and respondent's marital status.

By combining the information from the retrospective survey and official arrest records, we have information on the respondent's family structure while growing up, such as whether the mother worked in paid employment outside the home, presence of both parents and number of siblings; characteristics of the respondent's parents such as their education; self-reported information on physical and sexual abuse; other important variables, such as the respondent's educational attainment, marital history and number of children; and the official arrest records for 575 men and 201 women who were surveyed. The dependent variable for the following analysis is an ordered categorical variable for juvenile arrest (0 if not arrested, 1 if arrested once, 2 if arrested more than once)[6]. Sample statistics for these data are provided in appendix 1. The unweighted statistics are based on sample averages. The weighted statistics take the stratified nature of the sample into account and reflect statistics for the underlying population.

[6] In order to facilitate the comparison between the determinants of juvenile arrest for males and females, we collapse the strata categories for males into those for females.

2.2 Missing data

The observations on the dependent variable (juvenile arrests) and the independent variables were collected at different points in time (phase 1 and phase 2 of the data collection) and are physically contained in different files. The file containing the dependent variables has observations on all 27,160 persons in the 1958 Philadelphia birth cohort from which people are sampled for the retrospective survey. Most observations from the retrospective survey can be linked to the respondent's arrest records through a common identification number called the link code. However, a group of 59 women who were surveyed about family background and other characteristics has not been provided with this link code. As a result, the independent and dependent variables cannot be matched for these women. The observations from the arrest file that are potential matches for the unlinked women can be reduced considerably by the information we have. This information includes variables that occur in both the criminal and retrospective survey files, such as birth month and race, and the total number of individuals surveyed from each of the strata. A more detailed discussion of the available information can be found in appendix 2. Rather than dropping the unmatched women from the sample, we address the problem by adopting an estimation technique, which accounts for the scrambled nature of the data. This is discussed in section 3.3.

2.3 Selection of variables

In the following analysis, we draw on previous research for identifying key variables of interest in determining the probability of juvenile arrest. Although juvenile delinquency is often not modelled explicitly, most studies of crime control for a range of background characteristics that may contribute to criminality in both childhood and adulthood. These background characteristics include family structure during the individual's childhood, such as: whether the father, mother, or both parents were present in the childhood home, whether the mother worked; and the number of siblings the individual has. Another class of variables measures the socio-economic status of the respondent's family during childhood, such as: ethnicity, race, occupational status of the household head during high school and household income. The effect of role models and peer effects are captured using variables such as: gang membership, criminal history of family members, drug and alcohol problems of family members, parent's education and religiosity. Whether the respondent is in school and educational attainment are also commonly included to capture the opportunity cost of time. In light of the recent research linking female criminality to physical and sexual abuse in childhood, (see for example Ritchie, 2000; Widom, 2000) we also include measures of abuse in the model for juvenile delinquency.

3. Methodology

3.1 Ordered Probit

Suppose that latent juvenile criminal activity, denoted y_1^*, depends on a vector of observable characteristics x_1, such as race, family structure, history of physical or sexual abuse, parent's education, whether the respondent's mother participated in paid employment, and whether the respondent dropped out of school, and unobservables ε. Assuming ε is a standard normal random variable, the model for juvenile arrests can be written as:

$$y_1^* = \beta_1 x_1 + \varepsilon \qquad (1)$$

While delinquent activity is not directly observed in the data, we do observe whether an individual has, no arrests, one arrest, or two or more arrests. We denote this categorical variable y_1, where

$$
\begin{aligned}
y_1 &= 0 \quad \text{if not arrested as a juvenile} \\
&= 1 \quad \text{if arrested once as a juvenile} \\
&= 2 \quad \text{if arrested more than once as a juvenile}
\end{aligned}
$$

y_1 is related to latent juvenile criminal activity, according to the following rules:

$$
\begin{aligned}
y_1 &= 0 \ \text{ if } y_1^* \le 0 \\
y_1 &= 1 \ \text{ if } 0 < y_1^* \le c \\
y_1 &= 2 \ \text{ if } y_1^* > c
\end{aligned}
$$

where, c is a threshold parameter to be estimated together with the other parameters.

Then the likelihood for this ordered probit model is given by:

$$f(y_1) = P(y_1 = 0)^{(1-y_{11})(1-y_{12})} P(y_1 = 1)^{y_{11}} P(y_1 = 2)^{y_{12}} \qquad (2)$$

where y_{11} is an indicator for being arrested once as a juvenile (that is y_{11}=1 if y_1=1, and zero otherwise), and y_{12} is an indicator for being arrested more than once as a juvenile (that is y_{12}=1 if y_1=2, and zero otherwise).

3.2 Dealing with stratification

This model can be estimated by maximum likelihood using data on a random sample of individuals. However, the data used in this study were generated by a stratified random sample, where stratification is based on the number of juvenile arrests. This variable appears in the model as a dependent variable. Manski and Lerman (1977) and Manski and McFadden (1981) show that a simple weighting of the observations and a correction of the covariance matrix are sufficient to deal with this type of endogenously stratified data. The weights are calculated by dividing the population proportions by the sample proportions. The covariance matrix is calculated as HGH, where H is the negative inverse of the hessian of the (weighted) log-likelihood and G are the summed outer products of the first derivatives of the (weighted) log-likelihood.

3.3 Using the EM Algorithm to Overcome the Problem of Scrambled Data

A second issue for estimation, only arising when estimating the model for females, is that the variable linking the follow-up survey data with the juvenile arrest information is missing for 59 out of the 201 observations for women. As a consequence, the dependent variable is not uniquely observed for this subsample of women.

The EM algorithm is often used in situations where data are missing or where variables are censored. The problem in this paper is of a similar nature. The difference between the standard missing or censored variable problem and the problem in this research is that the number of possible values for each variable is finite. The EM algorithm is normally used in a continuous context, but because of this finite number of possible values, a discrete approach is more appropriate. Kalb (1998) gives a full description of the derivation of this discrete version of the EM algorithm and investigates its performance[7].

The variable vector $Y_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix}$ (where Y_1 is the number of juvenile arrests and Y_2 is the number of adult arrests) is in a separate file from the independent variables, X, for the 59 unmatched women. Although this paper only analyses juvenile arrests, the information on adult arrests can be used to limit the possible combinations that can be made between the records in the two data files. There is a choice from at most six different possible values for

[7] The derivation of the discrete version of the EM algorithm hinges on the validity of a similar lemma as is underlying the derivation of the EM algorithm in the continuous case (lemma 1e.6 (i) and (ii) in Rao (1973)).

Y_i. That is, $Y_i \in \left\{ \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}2\\0\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix} \right\}$. The number of possible matches from the arrest file is restricted for some observations because of the limited occurrence of women with adult arrests (Y_{2i}=1) and of women with more than one juvenile arrest (Y_{1i}=2) (see Appendix 1), and because the respondent's birth month is contained in each file and has to be matched. Each record in the cohort file can at most be linked once. Thus, selecting a value for one woman may restrict the choice of values for the other 58 women. This means that for the women without link codes, a combined likelihood expression has to be evaluated instead of an expression for each individual, because the matching to values for the dependent variables for one woman cannot be done independently from the matching for the other women. The unit of observation is this group of 59 women rather than each woman individually. The information available is observed for the whole group at once instead of for each woman separately. Appendix 2 reports on the available information. The EM algorithm adapted for use in a discrete context is discussed in Appendix 3.

4. Results

The results for the ordered probit models of juvenile arrest for boys and girls are contained in Table 1. The dependent variable for the model is an ordered categorical variable taking on a value of zero if the individual was not arrested, one if the individual was arrested once, and two if the individual was arrested more than once. The upper panel of Table 1 contains results for males, and the lower panel presents results for females[8].

[8] For women, a joint model with adult arrests, including a constant only (given the small number of women with adult arrests), is estimated to enable use of all available information in the EM algorithm. It is assumed that the two equations are independent. The latter means the information on adult arrests is only used in determining the probability of a combination occurring and it will not affect the results for juvenile arrests. This approach will prevent the selection of too many observations with adult arrests. The adult arrest equation is estimated through a simple probit model (0 is no adult arrests, 1 is one or more adult arrests).

Table 1. — Ordered Probit of Juvenile Arrests for Men and Women

	Men				
	Estimated model		Marginal effects[a]		
	Parameter	t-value	JA=0	JA=1	JA=2
Constant	-1.1386	-6.39	0.7087	0.1302	0.1612
Race is non-white	0.3496	2.32	-0.1091	0.0333	0.0759
No father in childhood home	0.1697	0.85	-0.0542	0.0151	0.0391
Stay at home mother	0.2242	1.43	-0.0708	0.0199	0.0509
Physically or sexually abused	-0.1617	-0.63	0.0482	-0.0148	-0.0333
No mother in childhood home	0.2806	0.61	-0.0920	0.0233	0.0687
Father's education is $\geq$ hs grad	-0.3041	-1.93	0.0946	-0.0291	-0.0656
Number of siblings	0.0677	1.91	-0.0190	0.0065	0.0125
Left school < 17 years old	0.7527	3.38	-0.2582	0.0617	0.1965
Left school at 17 years old	0.3142	1.96	-0.0990	0.0311	0.0678
Threshold parameter	0.4903	10.03			

	Women				
	Estimated model		Marginal effects[a]		
	Parameter	t-value	JA=0	JA=1	JA=2
Constant	-1.5790	-5.19	0.8640	0.1181	0.0220
Race is non-white	0.4889	2.73	-0.0941	0.0736	0.0205
No father in childhood home	0.3284	0.87	-0.0720	0.0546	0.0175
Stay at home mother	0.4767	2.17	-0.1040	0.0781	0.0259
Physically or sexually abused	-0.3160	-0.90	0.0538	-0.0427	-0.0111
No mother in childhood home	-0.5347	-1.22	0.0801	-0.0646	-0.0155
Father's education is $\geq$ hs grad	-0.0951	-0.41	0.0186	-0.0144	-0.0042
Number of siblings	-0.0383	-0.73	0.0082	-0.0062	-0.0021
Left school < 17 years old	1.0905	2.15	-0.2968	0.2062	0.0906
Left school at 17 years old	0.2492	1.17	-0.0471	0.0384	0.0088
Threshold parameter	1.0799	6.88			

Note a: In the row for the constant, the predicted proportions of the sample with 0, 1 and 2 or more juvenile arrests are presented.

Parameter estimates and t-statistics can be found in the first two columns of the table. Columns 3, 4 and 5 contain the marginal effect of a change in an explanatory variable on the probability that the individual is observed to have zero, one, or greater than one juvenile arrest, respectively. The marginal effect of a discrete variable is defined as the change in the probability of zero, one, or greater than one juvenile arrest associated with the variable of interest changing from zero to one, while keeping all other characteristics at their observed values. The change in probability is averaged over all observations. The predicted average probability of being in each juvenile arrest category is contained in the row corresponding to the constant term.

The following subsections discuss the results and compare them to results found in the previous literature. Some of these previous results relate to adult rather than juvenile criminality.

4.1 Pattern of Delinquent Behaviour

The first row of columns 3, 4 and 5 in Table 1 reports the predicted proportion of the sample with zero, one, or greater than one juvenile arrest. This information reveals that males and females differ in the pattern of delinquent behaviour. Overall, 29 per cent of males are predicted to have at least one juvenile arrest, compared to 14 per cent of females. While this is consistent with the accepted wisdom that delinquency is a larger problem for males than females, it nonetheless indicates that delinquency is a substantial problem for females as well. The higher rate of juvenile arrest experienced by males can be attributed to a greater prevalence of repeat offenders. Males are much more likely to have more than one arrest compared to females (0.16 and 0.02 respectively). The proportion of males and females who have only one juvenile arrest is similar (0.13 for males, and 0.12 for females).

This greater tendency for males to experience multiple arrests is also evident in the marginal effects. The marginal effect of changing a characteristic mostly shifts males between no arrests and more than one arrest, whereas females mostly move between no arrests and one arrest. This is reflected in the size of the threshold parameter, which is much higher for females than for males, indicating that women are unlikely to be in the group with the highest arrest rate.

4.2 Race

The results in Table 1 are informative in terms of the determinants of juvenile delinquency, and the extent to which these determinants differ across gender. We find that there is substantial concordance between the two estimated models. In terms of what matters, we find that both boys and girls who are not white are significantly more likely to be arrested than white children. The literature reports mixed findings with respect to race.[9] For example, similar to the results reported above, Phillips and Votey (1987) and Witte and Tauchen (1994) find that blacks are significantly more likely to be involved in crime. In contrast, Comanor and Phillips (1999) find that being black is associated with significantly fewer encounters with the law, a result confirmed by Levitt and Lochner (2001) for men, while both Williams and Sickles (2000) and Grogger (1998) find a positive but insignificant relationship, which is also found by Levitt and Lochner (2001) for women.

Being non-white decreases the probability of no juvenile arrests by 10.9 percentage points on average for males, and 9.4 percentage points for females. The probability of having exactly one juvenile arrest is 3.3 percentage points greater for non-white males and 7.4 percentage points greater for non-white females, compared to their white counterparts.

[9] With respect to race, differences in the sample analyzed and definition of participation in crime also lead to conflicting results between studies based on the same data sets.

The probability of more than one juvenile arrest is estimated to be 7.6 percentage points higher for non-white males, and 2.1 percentage points higher for non-white females compared to whites. These are large effects, given that the average predicted probability of having more than one arrest is only 16.1 per cent for males, and 2.2 per cent for females.

4.3 School-leaving age

Leaving school before the age of 17 is also associated with a greater probability of juvenile arrest for both boys and girls. This is similar to the results of Witte and Tauchen (1994) and Phillips and Votey (1987) who find that being in school has a significantly negative effect on the probability of arrest. For males, leaving school before the age of 17 decreases the probability of no juvenile arrests by 25.8 percentage points. The decrease is 29.7 percentage points for females. As with the effect of race, we find that the decrease in the probability of no juvenile arrest associated with leaving school early is accompanied by a large increase in the probability of having more than one arrest for males (19.7 percentage points), and a large increase in the probability of having a single arrest for females (20.6 percentage points)[10]. The interaction between race and leaving school early is insignificant.

4.4 Presence of Parents

The results in Table 1 are also informative about some commonly held beliefs about the causes of juvenile delinquency. We find that growing up without a father in the childhood home has a positive but statistically insignificant effect on the probability of juvenile arrest for both boys and girls. Although, the presence of a father in the childhood

[10] This is clearly an important result. This finding is consistent with several hypotheses regarding the influence of education on juvenile delinquency. These hypotheses include schooling, which works through the time constraint as in "the Devil makes work for idle hands" hypothesis; through shaping preferences away from criminal activities which may be the peer effect; through household budget constraints, if leaving school before age 17 is because parents cannot afford to keep their children in school; and through having poor earning opportunities because of low levels of human capital.

home is generally found to have a negative impact on measures of criminal involvement (Comanor and Phillips, 1999), Williams and Sickles (2000) also find that the father's presence has an insignificant effect on the probability of adult arrest. Similarly, Case and Katz (1991) find that having both parents present in the family home has a negative but insignificant effect on the probability of participating in crime, whereas Levitt and Lochner (2001) find a significant negative effect.

Due to a focus on the absence of a father in the literature, we experimented with disaggregating the measure no father present to no father present due to divorce or separation, and no father present due to death. These variables were neither individually nor jointly significant. We also investigated whether the presence of a stepfather affected the respondent's juvenile delinquency. This variable was also not significant.

The results also fail to find evidence of the popularly held belief that working mothers contribute to higher levels of juvenile delinquency. On the contrary, the model for females finds statistically significant evidence that having a mother who stays at home increases the probability of juvenile arrest. Moreover, the point estimate for males also suggests that having a mother who stays at home is associated with an increased probability of juvenile arrest, and this effect is statistically significant at the 10%-level on the basis of a one sided test. Levitt and Lochner (2001) find a similarly signed small insignificant effect. Although not reported, we investigated whether this result was sensitive to whether the respondent's mother was a single mother by including an interaction term between mother not working and no father in the respondent's childhood home. This interaction term was not statistically significant.

4.5 Childhood Abuse

Contrary to the findings of Ritchie (2000) and Widom (2000), we fail to find any statistical evidence that physical or sexual abuse during childhood is associated with increased probability of juvenile delinquency for girls or boys. However, with just 9 per cent of the sample of 575 males and 7 per cent of the sample of 201 females reporting sexual or physical abuse, we simply might not have the power to identify this effect.

4.6 Parental Education

While the estimated models of male and female delinquency share many common findings, there are nonetheless distinct differences. In particular, having a father who graduated from high school reduces the probability that a male is arrested as a juvenile, but is not a significant determinant of juvenile arrests for females. For males, the effect of having a father who has graduated from high school is of a similar magnitude as the effect of race, although opposite in sign. Because education is generally correlated with employment and income, we would expect that fathers with more education are more likely to be employed and earning a higher wage than those with lower levels of education. This suggests that the father's education may be a proxy for socio-economic status. An alternative explanation is that fathers with higher levels of education are better able to act as a positive role model for their sons and provide information about legitimate opportunities to their sons. Although not included in the results in Table 1, the mother's education was also included in the model for males and females. We found that the mother's education had no significant effect on the probability of juvenile arrest, and it was therefore excluded from the model.

In terms of peer effects and role models, Case and Katz (1991) find that the parent's educational attainment has a negative but insignificant impact on criminality. Levitt and Lochner (2001) find no effect for men and a small positive effect of the father's education on female criminality. Interpreting parental education as a proxy for socio-economic status, the result found here is similar to Phillips and Votey's (1987) result, who find that the socio-economic status of the respondent's family is negatively related to the probability that an individual engages in crime.

4.7 Number of Siblings

Finally, we find that the probability that a male is arrested as a juvenile is increasing with the number of siblings he has. This result is common in the literature, which is limited to males, and is generally interpreted as reflecting the fact that children from larger families receive less individual attention. Williams and Sickles (2000) and Grogger (1998) found positive although generally insignificant effects of family size on the probability of engaging in crime. It is therefore somewhat surprising that the number of siblings does not have the expected sign (although it is insignificant) in the model for female juvenile arrests.

5. Conclusion

This research used data from the Delinquency in a Birth Cohort II study to investigate a number of popular hypotheses regarding the determinants of juvenile delinquency, as well as examining whether the determinants of criminality differ by gender.

The results for the comparative analysis of juvenile delinquency indicate that some determinants of delinquency are common across gender, while others are gender specific. In terms of common determinants, we find that the probability of juvenile arrests for both males and females is greater for non-whites, who leave school before the age of seventeen.

Contrary to recent research, we fail to find evidence that sexual or physical abuse is a significant determinant of either female or male delinquency, or that coming from a single parent household is associated with a greater likelihood of delinquency. We also find important differences both in the determinants and patterns of male and female delinquency. In particular, father's education and number of siblings are significant determinants of male –but not female– delinquency, with juvenile arrest less likely for boys whose father has graduated from high school and who have fewer siblings. A major difference between males and females is that a change in characteristics mostly shifts males from no arrests to more than one arrest, whereas females move between no arrests and one arrest. This results in a pattern of delinquency where a similar proportion of males and females experienced a single juvenile arrest, but where males were about twice as likely as females to have more than one arrest.

An issue not addressed in this paper is whether education can be treated as exogenous to the crime decision. If characteristics that lead people to stay in school also result in them being less likely to participate in crime, then the results overstate the impact of education on criminal outcomes. This remains an issue for future research.

References

Case, A., and L. Katz, "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths," Working Paper, National Bureau of Economic Research, Cambridge, MA (1991).

Comanor, W.S., and L. Phillips, "The Impact of Income and Family Structure on Delinquency," Unpublished manuscript, University of California, Santa Barbara (1999).

Dempster, A.P., N. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society B* 39 (1977), 1-38.

Figlio, M., P.E. Tracy, and M.E. Wolfgang, "Delinquency in A Birth Cohort II: Philadelphia 1958-1986: Computer File," Philadelphia, PA: Sellin Center for Studies in Criminology and Criminal Law and National Analysts, Division of Booz-Allen and Hamilton, Inc.: Producers, 1990. Ann Arbor, MI: ICPSR (Second Release 1991).

Freeman, R.B., "Crime and the Economic Status of Disadvantaged Young Men," Mimeo, Conference on Urban Labor Markets and Labor Mobility, Virginia: Arlie House (1991).

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

Good, D., M. Pirog-Good, and R.C. Sickles, "An Analysis of Youth Crime and Employment Patterns," *Journal of Quantitative Criminality* 2 (1986), 219-235.

Grogger, J., "Market Wages and Youth Crime," *Journal of Labor Economics* 16 (1998), 765-791.

Jurik, N.C., “The Economics of Female Recidivism,” *Criminology* 21 (1983), 603-622.

Kalb, G., An Australian Model for Labour Supply and Welfare Participation in Two-adult Households. Unpublished PhD thesis, Monash University, Department of Econometrics (1998).

Kloek, T., and H.K. van Dijk, “Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo,” *Econometrica* 46 (1978), 1-19.

Laub, J.H., and R.J. Sampson, “Unraveling Families and Delinquency: A Reanalysis of the Glueck’s Data,” *Criminology* 26 (1988), 355-379.

Levitt, S.D., and L. Lochner, "The Determinants of Juvenile Crime", in J. Gruber (ed.), *Risky Behavior Among Youths: An Economic Analysis*, (Chicago: University of Chicago Press, 2001).

Manski, C.F., and S.R. Lerman, “The Estimation of Choice Probabilities from Choice Based Samples,” *Econometrica* 45 (1977), 1977-1988.

Manski, C.F., and D. McFadden, “Alternative Estimators and Sample Designs for Discrete Choice Analysis,” in C.F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*. (Cambridge, Massachusetts: MIT press, 1981).

Phillips, L., and H.L. Votey, “Rational Choice Models of Crime by Youth,” *Review of Black Political Economy* 16 (1987), 129-187.

Rao, C.R., *Linear Statistical Inference and Its Applications.* (New York: John Wiley & Sons, 1973).

Ritchie, B.E., “Exploring the Link Between Violence Against Women and women’s involvement in Illegal Activity,” in *Research on Women and Girls in the Criminal*

Justice System: Plenary Papers of the 1999 Conference on Criminal Justice Research and Evaluation – Enhancing Policy and Practice Through Research, Vol 3, NCJ 180973 (2000).

Snyder, H. N., "Juvenile Arrests 1999," U.S. Department of Justice, (2000)

Tanner, M.A., *Tools for Statistical Inference; Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer series in statistics, (New York: Springer-Verlag, 1993).

Van Dijk, H.K., and T. Kloek, "Further Experience in Bayesian Analysis using Monte Carlo Integration," *Journal of Econometrics* 14 (1980), 307-328.

Widom, C.S., "Childhood Victimization and the Derailment of Girls and Women to the Criminal Justice System," In *Research on Women and Girls in the Criminal Justice System: Plenary Papers of the 1999 Conference on Criminal Justice Research and Evaluation – Enhancing Policy and Practice Through Research,* Vol 3, NCJ 180973 (2000).

Williams, J., and R.C. Sickles, "An Analysis of the Crime as Work Model: Evidence from the 1958 Philadelphia Birth Cohort Study," Working Paper, University of Adelaide (2000).

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

APPENDIX 1— SUMMARY STATISTICS

TABLE A.1.— SUMMARY STATISTICS OF VARIABLES IN THE SAMPLE USED

	Men (N=575)		Women (N=201)	
	weighted	unweighted	weighted	unweighted
Indicator for juvenile arrests (=1)†	0.1290	0.2383	0.1075	0.3028
Indicator for juvenile arrests (>1)†	0.1606	0.3826	0.0205	0.3099
Indicator for adult arrests†	0.1186	0.2104	0.0028	0.0423
E(juvenile arrests (=1))			0.1193	0.3117
E(juvenile arrests (>1))			0.0215	0.3650
E(adult arrests)			0.0031	0.0409
Race is non-white	0.5274	0.4609	0.5263	0.4478
No father in childhood home	0.1659	0.1896	0.1309	0.2090
Stepfather	0.0844	0.0904	0.0789	0.0945
No father (divorced)	0.1047	0.1113	0.0758	0.1244
No father (deceased)	0.0237	0.0383	0.0454	0.0448
No father (other)	0.0374	0.0400	0.0100	0.0398
Stay at home mother	0.3359	0.3704	0.2394	0.3383
No father*stay at home mother	0.0593	0.0661	0.0477	0.0597
Physically abused in childhood	0.0769	0.0713	0.0144	0.0348
Sexually harassed in childhood	0.0143	0.0139	0.0604	0.0597
No mother in childhood home	0.0301	0.0261	0.0294	0.0299
No mother (divorced)	0.0244	0.0157	0.0268	0.0199
No mother (deceased)	0.0026	0.0070	0.0001	0.0050
No mother (other)	0.0030	0.0035	0.0017	0.0050
Don't know mother's education	0.0570	0.0765	0.0310	0.0448
Mother's education <hs grad	0.2627	0.3113	0.3061	0.3433
Mother's education is hs grad	0.6194	0.5600	0.5713	0.5522
Mother's education > hs grad	0.0308	0.0261	0.0623	0.0299
Don't know father's education	0.0849	0.1148	0.0917	0.0945
Father's education <hs grad	0.3034	0.3183	0.2916	0.3035
Father's education is hs grad	0.3866	0.3391	0.4081	0.3433
Father's education > hs grad	0.0592	0.0383	0.0777	0.0498
Number of siblings	3.0500	3.2243	3.2571	3.2736
left school <16 years old	0.0392	0.0574	0.0326	0.0896
left school at 16 years old	0.0891	0.1635	0.0470	0.1642
left school at 17 years old	0.2794	0.2730	0.3199	0.3085

† For women, the statistics for these variables are only calculated over the 142 complete observations.

APPENDIX 2— INFORMATION USED FOR MATCHING

The following information is used for matching:

- the unmatched records have a cohort identification number over 24780
- the unmatched records are white women
- birth month (occurs in both files)
- we know the number of white females in the birth cohort who have no arrests, one arrest and more than one arrest, and the number of women in these categories who were surveyed and hence that amongst the unlinked women there are 10 without arrests, 20 with one arrest and 30 with more than one arrest.
- assuming that people do not report more arrests than they have actually had (but rather underreport their arrests), the number of self-reported arrests in the survey identifies in some cases the strata the women are in. We found that six women are in strata 5 (one arrest) and 14 in strata 6 (more than one arrest)
- there are only a restricted number of people with adult arrests in each subgroup, which further limits the number of possible values to choose from.

Of the 60 women without a link code one woman was dropped. Since we do not know to which strata the woman belongs, the above information remains unchanged.

The cohort file, from which we can choose a match for the 59 surveyed women with a missing link code, contains 2046 records for white women with no juvenile arrests, 33 of whom have had an adult arrests; 282 records for white women with a single juvenile arrest, 27 of whom had adult arrests; and 52 records for white woman with greater than one juvenile arrest, eight of whom had adult arrests.

APPENDIX 3— THE EM ALGORITHM IN A DISCRETE CONTEXT

Instead of integration over a range of possible values for a missing or censored variable, in this discrete version of the EM algorithm, a summation over all possible combinations of values for the variable Y is performed. Explicitly writing down this set of possible combinations is cumbersome, since the constraints on the possible combinations need to be fulfilled for all 59 observations simultaneously. The choice for one observation can affect the possible choices for another observation. Therefore for ease of notation, a set A is defined which contains all possible combinations that fulfil all the requirements. We defer the discussion of computational issues till the end of this appendix.

The contribution to the likelihood function of the 59 unmatched women can be constructed by taking the joint probability density function for the 59 observations without link code and summing over all combinations contained in the set A:

$$L(\theta|A)=\sum_{i,k,\ldots,p\in A} \mathrm{pdf}\left(Y_1=y_i, Y_2=y_k,\ldots,Y_{59}=y_p|\theta, X_1,\ldots,X_{59}\right)$$

$$=\sum_{i,k,\ldots,p\in A} \mathrm{pdf}\left(Y_1=y_i|\theta,X_1\right)\ldots\mathrm{pdf}\left(Y_{59}=y_p|\theta,X_{59}\right) \qquad \text{(A.1)}$$

Taking logarithms to obtain the log likelihood:

$$l(\theta|A)=\ln\left\{\sum_{i,k,\ldots,p\in A} \mathrm{pdf}\left(Y_1=y_i|\theta,X_1\right)\ldots\mathrm{pdf}\left(Y_{59}=y_p|\theta,X_{59}\right)\right\} \qquad \text{(A.2)}$$

where θ is the parameter vector, consisting of β_1, c and β_2 (the parameter vector in the probit model for adult arrests), that has to be estimated; A is the set of combinations for the

59 non-matched observations that fulfil all constraints[11]; and pdf stands for the probability density function.

Using the EM algorithm results in a rewritten log likelihood expression (see Dempster, Laird and Rubin, 1977):

$$l(\theta|\,\mathrm{A}) = \mathrm{Q}(\theta, \varphi\,; \mathrm{A}) - \mathrm{H}(\theta, \varphi\,; \mathrm{A}) \qquad \text{(A.3)}$$

where φ is a vector defined over the same domain as θ and the above is valid for any value of φ. The function H is not relevant when searching for the maximum of $l(\theta|\,\mathrm{A})$. In the function Q, the contribution $\log(f(\theta,y))$ of an unobserved latent variable y to the log likelihood function is replaced by its expectation over the set of values in which its true value is known to lie:

$$\mathrm{Q}(\theta,\varphi;\mathrm{A}) = \sum_{i,k,\ldots,p\in A} \{\ln[f(y_i|\theta,X_1)] + \ln[f(y_k|\theta,X_2)] + \ldots + \ln[f(y_p|\theta,X_{59})]\} \bullet \Pr(Y_1 = y_i,\ldots,Y_{59} = y_p|\varphi,A) \qquad \text{(A.4)}$$

Applying the EM algorithm to the problem of the scrambled data sets and assuming a distributional form: $\mathrm{pdf}(y|\,\varphi,X) = f(y|\,\varphi,X)$, the expression Q can be calculated. Assuming that $f(y|\,\varphi,X)$ is known, $\Pr(Y_1, Y_2,\ldots,Y_{59}\,|\,\varphi,A)$ can be constructed:

$$\Pr(Y_1 = y_i,\ldots,Y_{59} = y_p|A,\varphi) = \frac{\mathrm{pdf}(Y_1 = y_i|X_1,\varphi)..\mathrm{pdf}(Y_{59} = y_p|X_{59},\varphi)}{\sum_{i,k,\ldots,p\in A} \mathrm{pdf}(Y_1 = y_i|X_1,\varphi)..\mathrm{pdf}(Y_{59} = y_p|X_{59},\varphi)}$$

$$= \frac{f(y_i|X_1,\varphi)..f(y_p|X_{59},\varphi)}{\sum_{i,k,\ldots,p\in A} f(y_i|X_1,\varphi)..f(y_p|X_{59},\varphi)} \quad \text{for } \{y_i,\ldots,y_p\}\in A \qquad \text{(A.5)}$$

$$= 0 \qquad \text{elsewhere}$$

[11] See appendix 2.

The Q-function in equation (A.4) has to be maximized with respect to θ, where φ is given. Dempster, Laird and Rubin (1977) have shown that iteratively maximizing this function using the previous optimal values θ^{q-1} for φ leads to convergence. Only an arbitrary value $\varphi=\theta^0$ is needed to start the process and the iterations are finished when $\theta^q = \theta^{q-1}$.

As previously noted, writing down the set A of possible combinations would be quite complicated. The requirements need to be checked simultaneously for all 59 observations for each possible match that can be made. The exact calculation in the E-step can be replaced by a Monte Carlo implementation of the E-step (Tanner, 1993). Using a simulation technique to approximate the expected value will be convenient, since we can then just check the requirements for each simulated combination and only allow those that fulfill all conditions.

However, before observations can be sampled from the function, all elements of the above probability function would have to be calculated. Therefore, to circumvent having to write down all possible combinations a further step is needed, since the probability of each possible combination occurring still contains the sum over all possible combinations in the denominator. Instead of using the exact probabilities in the simulation, we use an approximation and correct for the approximation by weighting the draws with the ratio of the true probability and the approximated probability. In Bayesian literature much use is made of this method named importance sampling[12], which under certain conditions allows one to draw from a simpler distribution (the importance function) than the actual one. In order to get good results the approximating distribution has to be sufficiently similar to the original distribution. This means that the approximating distribution should have large probability where the actual distribution has large probability, so that all important

[12] See Kloek and Van Dijk (1978) and Van Dijk and Kloek (1980).

combinations will be drawn from the importance function.[13]

The importance function considered here is constructed from a sequence of probabilities. The procedure for generating data for Y goes as follows:

Choose one unmatched record containing a vector X with characteristics from the main data set to start from and calculate probabilities of observing each of the attainable values for Y in the cohort file (CF) given X and a value for $\theta = \theta^q$:

$$\Pr\left(Y_1 = y_i \middle| X_1, \theta^q\right) = \frac{f\left(y_i \middle| X_1, \theta^q\right)}{\sum_{k \in A} f\left(y_k \middle| X_1, \theta^q\right)}, \text{ for } y_i \in CF \tag{A.6}$$

Draw a value for Y_1 from this discrete distribution. Then go to the next record on X and repeat this procedure after removing the value $y_{1,t}{*} = y_{t_1}$ that is drawn in the previous step from the cohort file. t_1 indicates the t^{th} simulated y value belonging to X_1. There are at most six different values to choose from.

The probability distribution function then looks like:

$$\Pr\left(Y_2 = y_i \middle| X_2, \theta^q\right) = \frac{f\left(y_i \middle| X_2, \theta^q\right)}{\sum_{\substack{k \in A \\ k \neq t_1}} f\left(y_k \middle| X_2, \theta^q\right)}, \text{ for } y_i \in CF \setminus \{y_{t_1}\} \tag{A.7}$$

where A \ B means the set A after deletion of the elements in set B.

From this distribution function y_{t_2} is drawn. In the next step y_{t_1} and y_{t_2} are excluded from the set of possible values for Y. After each draw the set of attainable values

[13] The number of drawings T from the importance function $I(Y_1, Y_2, \ldots, Y_{59} | A, \theta_q)$ can be less when this function is more similar to the actual distribution.

for Y contains fewer value. Following this procedure, a value for Y is drawn for all records on X in the matching group in a sequential way. For each series of draws of y-values the procedure starts from a different record in the main data set. This is done since the probability of certain combinations X and Y occurring is dependent on the order of observations X_i (i=1,...,59) for which y-values are drawn. The last observation on X potentially has fewer values to choose from since several possible values have already been combined with other observations on X. Always starting from the same observation on X could disadvantage certain combinations of values y and x. By alternating the starting point it is hoped that all combinations that would occur with high probability according to the actual probability distribution, will also be drawn from this simpler importance function.

Although an attempt is made to get as close as possible to drawing from the actual probability distribution function, we know that this will never exactly be the case. To correct for this drawing from an incorrect distribution, the simulated contributions to the log likelihood have to be weighted. The appropriate weights can be found by dividing the value of the actual probability density function by the approximated value at the simulated data point[14]:

$$w_t = \frac{f\left(Y_1 = y_{t_1} \middle| X_1, \theta^q\right) .. f\left(Y_{59} = y_{t_{59}} \middle| X_{59}, \theta^q\right)}{\dfrac{f\left(Y_1 = y_{t_1} \middle| X_1, \theta^q\right)}{\sum\limits_{k} f\left(Y_1 = y_k \middle| X_1, \theta^q\right)} \cdot \dfrac{f\left(Y_2 = y_{t_2} \middle| X_2, \theta^q\right)}{\sum\limits_{\substack{k \\ k \neq t_1}} f\left(Y_2 = y_k \middle| X_2, \theta^q\right)} \cdots \dfrac{f\left(Y_{59} = y_{t_{59}} \middle| X_{59}, \theta^q\right)}{\sum\limits_{\substack{k \\ k \neq t_1, t_2, \ldots, t_{58}}} f\left(Y_{59} = y_k \middle| X_{59}, \theta^q\right)}} \tag{A.8}$$

[14] Note that only the numerator of both density functions appears in (A.8). The denominators are constants and can be left out of the formula.

We only accept those series of draws that fulfil all requirements as described in Appendix 2, that is $\{y_{t_1},..., y_{t_{59}}\} \in A$.

Using the weights w_t, $Q(\theta,\theta_q;A)$ is now approximated by:

$$\frac{1}{\sum_{t=1}^{T} w_t} \sum_{t=1}^{T} w_t \left\{\log\left[f\left(y_{t_1} | \theta, X_1\right)\right] + ... + \log\left[f\left(y_{t_{59}} | \theta, X_{59}\right)\right]\right\} \quad \text{(A.9)}$$

and the information matrix can be approximated by using (see Tanner, 1993):

$$\left.\frac{\partial^2 l\left(\theta | A, X_1, ..., X_{59}\right)}{\partial\theta^2}\right|_{\theta^q} = \frac{1}{\sum_{t=1}^{T} w_{tj}} \sum_{t=1}^{T} w_{tj} \left.\frac{\partial^2 \left\{\log\left[f\left(y_{t_1} | \theta, X_1\right)\right] + ... + \log\left[f\left(y_{t_{59}} | \theta, X_{59}\right)\right]\right\}}{\partial\theta^2}\right|_{\theta^q} +$$

$$\frac{1}{\sum_{t=1}^{T} w_t} \sum_{t=1}^{T} w_t \left(\left.\frac{\partial\left\{\log\left[f\left(y_{t_1} | \theta, X_1\right)\right] + ... + \log\left[f\left(y_{t_{59}} | \theta, X_{59}\right)\right]\right\}}{\partial\theta}\right|_{\theta^q}\right)^2 - \quad \text{(A.10)}$$

$$\left(\frac{1}{\sum_{t=1}^{T} w_t} \sum_{t=1}^{T} w_t \left.\frac{\partial\left\{\log[\left[f\left(y_{t_1} | \theta, X_1\right)\right] + ... + \log\left[f\left(y_{t_{59}} | \theta, X_{59}\right)\right]\right\}}{\partial\theta}\right|_{\theta^q}\right)^2$$