Keskustelualoitteita #48
Joensuun yliopisto, Taloustieteet

Criminal career and age-crime profile

Mikael Linden

ISBN 978-952-458-979-6
ISSN 1795-7885
no 48

CRIMINAL CAREER AND AGE-CRIME PROFILE

Mikael Linden

Economics and Business Administration
University of Joensuu, Finland
(mika.linden@joensuu.fi)

JUNE 2007

Abstract

Two classical issues in criminology, criminal career paradigm and age-crime curve rule, are analyzed in terms of economics. Age-crime curve sums up the crime intensity and participation aspects of individual crime behaviour during his/her lifetime. It is argued that crime intensity must be the starting point of economic analysis if we want to understand crime career and age-crime curve regularities. A simple model with crime intensity as age dependent decision variable was proposed to generate non-homogenous life time crime intensity rate. The bell shaped age-crime alternative was tested empirically with arrest count data of felony defendants in large urban counties in U.S year 1998. Poisson and NegBin count data regression models gave unsatisfactory results concerning the age dependency of number of arrest counts. NLS regression on individual age standardized arrest counts and semi-parametric Poisson regression on individual arrest counts provided acceptable results. The semi-parametric model estimation allowed for non-homogenous crime intensity presentation in the form of age class average proportional intensities. Regardless of using age standardized counts or controlling for avarage age-dependent intensities the bell shaped age-crime profile was not rejected for the sample.

Key words. Criminal career, age-crime profile, non-homogenous Poisson process, count data regression models, semi-parametric estimation.

I. Background

There are two topics that have excited much debate in recent criminology literature. The older one, criminal career paradigm, refers to basic empirical requirement for characterizing criminality as a career, the autocorrelation of criminal behavior. The probability to commit a crime is higher for individuals that have already a criminal record compared to individuals without one. The second topic is the most important empirical regularity in criminology, the age-crime curve: however or wherever measured the rates of criminal involvement in the population is highest for the young (males) between ages 17 to 25 (Land & Nagin 1993). Although these facts are not in accordance with each other in all cases (e.g. life-course criminals) they are typically closely inter-connected. Age-crime curve predicts that a gradual decline will take place in average crime rates among the older age cohorts. This fact does not rule out the state dependency predicted by criminal career paradigm. State dependency can still occur among the careerists although increasing number of criminals terminates their criminal career and crime intensity is small in the old age cohorts.

Basically criminal age or time dependency (age-crime curve) and state dependency (criminal career and recidivism) are conceptually distinct - the former refers to particular age (or time) structure among the criminals, and the latter refers to non-time dependent clustering (stigmatization, lock-in or hysteresis phenomena) of criminal actions among some criminals. Nevertheless they are empirically closely related since criminal activity is determined by the criminals' discounted lifetime, and once started criminality at youth breeds too often new crimes. Thus overall volume of crime is therefore a function of frequency and prevalence (participation). The crime frequency does not decrease as much with age as participation (Kyvsgaard 2003, Ch. 7-8). This particular two dimensionality of criminality is also harmed by many measurement biases (e.g. what is the proper and reliable unit of measurement of criminality, sample selection in analyzed data, individual unobserved heterogeneity, etc.) and conceptual problems (e.g. how to define the criminal career, changing legal norms and sanctions, etc.). In spite of these problems criminal career paradigm and age-crime curve are still major research agendas in criminology (e.g. Fagan & Western 2005, D'Unger, Land & McCall 2002, Levitt 1999).

However a more deeply problem lies in the disagreement concerning the theoretical understanding of age-crime curve and criminal career. Criminal (choice) career theory (Blumstein & Cohen 1979) underrates the age-crime curve much as statistical artifact but the criminal propensity theory (Gottfredson & Hirschi, 1990, 1983) stresses the lack of self-control and unsocial behavior of teenagers. Recently evolutionary psychology has give support the propensity theory with argument that age-crime curve is produced by the difference between the reproductive benefits and costs of competition among young men (see Campbell, 1995, Kanazawa 2002). The well developed modern statistical theory provides models that can test the implications of both theories (e.g. Land 1992, Nagin & Land 1993, Blumstein 2005). The results support both theories. The outcome is expected remembering the dimensions of criminality. Aggregate age-crime curve is a product of both the high rate of youth full participation in crime, and the high activity level (crime frequency or intensity) of these young offenders. Age specific differences in participation, however, rather that frequency, are primarily responsible for the basic form of the age-crime curve (Kyvsgaard, 2003, p. 106).

Economic theory does not have much to say about age-crime curve phenomena but criminal career issue can be understood by it in some extension. Heineke (1978) categorizes economic models of crime as either (1) portfolio problems, in which the agent must decide how much wealth to put risk through involvement in crime, or (2) labour supply problems, in which the agent must choose the amount of time to be allocated to illegal activity. In both cases criminal activity onsets if the expected net gain or utility from it exceeds the benefits from legal activity. In principle a repeated activity (recidivism) can be modeled as a sequence of crimes where prolonged activity can be seen as means to accumulate (illegal) capital or consumption resources. As the activity can be halted by imprisonment the incentive to return to illegal activity after sentence is high as typically the capital gained from crimes is confiscated by state. This kind of career optimization can been modeled with dynamic programming methods including three conditional states (terminated activity, halted activity, repeated activity) where the life time expected net income is maximized. Leung (1991, 1995) provides a nice starting point to these models. His approach resembles closely to technically demanding job change and career literature found in modern labour economics (e.g. Berkovic & Stern 1991, Hopenhayen & Rogerson 1993, Rust 1996, Adda & Cooper 2003). However many questions are still open and much work is needed to get transparent and robust economic models of recidivism.

The main problem with the approach above is that it can reveal only some issues on property crimes – not crimes in general. Secondly, it is silent of age-crime curve or profile regularity. By adding some elements of human capital and portfolio theory to the career optimization model would support age-crime curve, since relative expected gains from crimes of young person are high as their wealth is very low and first round sentences are not severe. Leung's paper (1994) on economics of age-crime profile is a promising starting point in this context. He observes first that two dimensionality of unimodal and positively skewed age-crime profile relates to the above underlying concepts of 'age-participation' and 'age-intensity'. The former can easily explain the age-crime profile but surprisingly Leung argues that changes in the latter can alone explain the age-crime profile.

The key of this result lies in the selection effect, i.e. intensity rate of offending increases with age and the hazard rate of arrest rises over time. An increasing hazard rate of arrest implies that fewer offenders can successfully avoid arrest as they got old. As a result, the proportion of older offenders is smaller because many of them had already been apprehended when they were young. The observed age-crime profile indicates a more intense *selection* among the young offenders because it is more difficult to sustain criminal career involving more serious offenses for an extended period of time (Leung 1994, p. 483-84). Leung's basic economic argument for increasing crime intensity rate stems from the consumption-investment decision theory. The probability that criminal will *not* be arrested by time *t* is like a capital asset that has to be consumed in the finite lifetime. A dynamic trade-off exists between consumption and capital de-accumulation.

An optimal strategy is to increase the intensity (i.e. consumption) over life time as the expected returns form criminal activity fall over the finite life time. If criminal would exhaust the stock of capital too early, he/she would not be able enjoy any later gains at all, resulting in a lower lifetime expected utility. The increase in the intensity rate over time is driven by the end-of-horizon effects.

Although Leung's approach is challenging and novel it has some drawbacks. First, the analogy to capital theory is somewhat awkward. If the criminal's subjective interest rate (time preference) is high and he/she is risk-aversive the increasing intensity rate is not necessarily sustained. Second, to derive the observed age-crime profile it is enough that the age-intensity profile has the same form, i.e. the intensity rate of offending increases rapidly in the teens and then start to decline after 25 years. However we do not have yet an economic theory to support this behaviour. Leung's trick to derive observed age-crime profile with increasing intensity rate is the selection effect. It is difficult to maintain a criminal career with an increasing intensity rate of offending over an extended period of time. Thus there exist less offenders in old age cohorts compared to young ones. Third, Leung does not deal with the issues of criminal career and recidivism. He only derives optimal time for first crime action with increasing intensity rate. Thus Leung's approach can give one answer to observed age-crime statistics but many other underlying aspects remain unexplained. However other economic papers that pay theoretical attention to dynamic complexity of criminal activity are few (Davis 1988, Polinsky & Rubinfield 1991, Lee & McCrary 2005, Jacob, Lefgren & Moretti 2004) [1)].

In this context a simple model that tackles the salient features of crime career and age-crime curve is proposed. The starting point of the model is the fact that crime intensity per time and age varies across the individuals and they control their intensity. Higher activity increases the probability of imprisonment. The criminal sets his crime rate at level that minimizes his expected time devoted to criminal activity incurring some costs.

---

1) Two last mentioned references give also partial review of current related literature. The special number of Int. Economic Review (Vol. 45, Number 3, 2004) on Economic Models of Crime contains also some important papers.

By augmenting the basic model with fixed expected life time during the total number of crimes can be committed we are able to show that increased number of crimes increases the expected life time with rate that increases also the life time crime intensity. This result gives an explanation to question why *old criminals* necessarily do not commit less age standardized crime than young ones (the intensity effect). The assumption of age dependent subjective time preference stresses the model results that crimes are committed at the beginning of life time. The model emphasizes the noticed importance of finding a connection between behavioral models and stochastic models of crime.

In the empirical part of paper some statistical models related to Poisson regression are suggested and estimated. However used crime career data (BJS, 1998) and implications of economic model refer to non-homogenous Poisson process for crime intensity. Thus some modifications of the basic Poisson regression – both parametric and semi-parametric - are proposed. Results show that age-crime curve is still valid although we control for observed personal characteristics and for average age cohorts effects of individual recidivism.

II. Model of crime intensity, criminal career, and age-crime profile

II.1. Optimizing the crime career

Assume that measurable criminal activity rate (intensity) v (i.e. number of crimes during given time interval, e.g. $s/year$) takes values between 0 and v_{MAX}, i.e.

$$v = \begin{cases} v, \text{ when } 0 \le v \le v_{MAX}, \text{ and} \\ 0, \text{ when } v > v_{MAX} \text{ and } v < 0 \end{cases} . \tag{I.1}$$

It is assumed that activity is distributed uniformly between $0 \le v \le v_{MAX}$. Naturally none activity takes place when $v \le 0$ and at rate $v > v_{MAX}$ when a sure arrest happens. Higher the activity, the higher is the probability that it is stopped by the control authorities. Structure of Eq. 1) gives a linear relationship between activity and the probability of being stopped

(I.2) $$Prob[V \le v] = p(v) = kv \quad 0 \le v \le v_{MAX}, \text{ where } k = 1/v_{MAX}.$$

The person is detained for average time t_0. The random number of authorities in proximity to criminal activity has a Poisson distribution with a parameter μ. Assume that number of crimes person commits is S. Now the average or expected time $E[t]$ of committing these crimes or *criminal career*, which includes also the imprisonment time, is

(I.3) $$E[t] = S/v + \mu S p(v) t_0 = S/v + \mu S k v t_0.$$

The rational criminal tries to *minimize* the expected time devoted to number of crimes S with optimal criminal activity rate value $v^* \in (0, v_{MAX})$. Thus we have assumed that criminal has chosen a criminal career and he/she is maximizing the net benefits from it. Since high value of v gives a possibility to commit many crimes and at same time it increases the probability of arrest the rational criminal looks for a minimizing value of v^* that gives an optimal trade off value between these opposite factors. Optimizing Eq. 3) in respect to v

(I.4) $$\frac{dE[t]}{dv} = -\frac{S}{v^2} + \mu S k t_0 = 0 \quad | \quad \left(\frac{d^2E[t]}{dv^2} = \frac{2S}{v^3} > 0\right)$$

gives a solution v^* that is *independent* of S

(I.5) $$v^* = \sqrt{\frac{1}{\mu k t_0}} = \sqrt{\frac{v_{MAX}}{\mu t_0}} \qquad (v^* < v_{MAX} \text{ if } v_{MAX} > \frac{1}{\mu t_0}).$$

The first order condition for optimal crime activity

(I.6) $$-\frac{1}{v^2} + \mu k t_0 = 0$$

entails following comparative statistic results

(I.7) $$\frac{dv}{d\mu}\Big|_{v=v*} < 0, \quad \frac{dv}{dk}\Big|_{v=v*} < 0, \quad \text{and} \quad \frac{dv}{dt_0}\Big|_{v=v*} < 0.$$

Thus an increase in control activity μ, a decrease of maximal crime rate, and an increase of arrest time decreases the optimal crime number per time unit. Subsequently the expected time t of crime career increases for fixed crime intensity $v*$ and S when the control parameters increase (see Eq. I.3). Thus we run into same number of crimes as earlier but now they are committed during the longer time period. Contrary to this typically the parameters μ, k, and t_0 are lower for first time and young criminals supporting high crime intensity rates for them.

The model above is in many ways too general or abstract to analyze all salient questions concerning the crime activity and life time crime career. However it gives some support for inverted U-shaped crime-age profile. Note also that optimal result for crime intensity v was surprisingly independent of participation activity S but the optimal $E[t]_{|v=v*}$ depends positively on S. In following three modifications are introduced into the model. First we notice that v is the crime activity during the whole *expected life of criminal* $E[T]$ and $v = s'/E[T]$. Now as (fixed) expected life time $E[T]$ is the maximal total expected life time devoted to criminal activity $E[t]$ and the criminal optimize with v we must have $s' = S$. Second we assume that the probability of arrest is a convex function of crime activity during the expected life time. Thus

(I.8) $$Prob[V \le v] = p(v) = kv^{a} \quad 0 \le v \le v_{MAX}, \text{ where } k = 1/v_{MAX} \text{ with } \alpha > 1.$$

Note that with $0 < \alpha < 1$, high criminal activity entails that the probability of arrest increases slowly. However when $\alpha > 1$ the arrest probability is low for low levels of activity but the arrest probability increases rapidly with activity level. We consider only the latter case. Finally we assume that criminal activity incurs additional costs with a convex cost function depending on the number of crimes, i.e. $(\beta/2)S^2$. The minimization problem of the *average length* or *expected criminal career* $E[\tau] \le E[T]$ is

$$E[\tau] = E[T] + \mu S k v^{\alpha} t_0 - \frac{\beta}{2} S^2$$

(I.9)

$$= E[T] + \mu k t_0 \frac{S^{\alpha+1}}{E[T]^{\alpha}} - \frac{\beta}{2} S^2 .$$

As the number of life time crimes S is now the only variable of model the criminal minimizes with it the expected life time devoted to criminal career. The first order condition is

$$\frac{dE[\tau]}{dS} = (\alpha+1)\mu k E[T]^{-\alpha} t_0 S^{\alpha} - \beta S = 0$$

(I.10)

$$\Rightarrow \quad g(S) = S^{\alpha-1} - B = 0 : \quad S^* = \sqrt[\alpha-1]{B}, \text{ where } \quad B = \frac{\beta E[T]^{\alpha}}{(\alpha+1)\mu k t_0} > 0.$$

A minimum S^* exists since $\frac{d^2 E[\tau]}{dS^2}|_{S=S*} > 0$ (see Appendix I). Note that for any parameter values of $a > 1$ (see Appendix II)

(I.11) $$\frac{dE[T]}{dS}\Big|_{S=S*} > 0 , \quad \frac{dE[T]}{d\mu}\Big|_{S=S*} > 0 , \quad \frac{dE[T]}{dk}\Big|_{S=S*} > 0 , \quad \frac{dE[T]}{dt_0}\Big|_{S=S*} > 0 .$$

These results mean that if the criminal keeps his criminal participation at the optimal level $S*$ that sustains the optimal minimized expected time devoted to this activity, then an increasing number of crimes must be compensated with higher expected total life time. In similar fashion if the control activities increase ($d\mu > 0, dk > 0, dt_0 > 0$) the expected life time increases. The result is called *the paradox of crime career*. In the fight against criminality the society increases the expected life time of careerist and his/her number of crimes. The paradox lies in the existence of optimal value of number of crimes per life time once the career of criminality is chosen: higher the expected life time among the criminals, larger is the number crimes devoted to minimize the length of expected criminal career.

II.2. Age dependent crime intensity and time preference

The results obtained above support the bell-shaped age-crime curve regulatory since we can argue that young criminals have higher crime intensity than older criminals, and longer is expected life time of criminal larger is the number of crimes committed by the person. Thus we have a prediction that crime participation is distributed evenly during the life time but the intensity varies. Note however that $dS/dt_0 < 0$ at participation optimum. The crime participation decreases when the average imprisonment time increases. This is the selection effect that reduces the number of older criminals among the population.

The selection story is however deeper one in this context since $dS/dE[T] > v = S/E[T]$ with optimum S^* (see App. II). The result indicates that optimum participation level can be only sustained with expected life time with rate that is larger than life time intensity rate. Note however that if $E[T]$ refers to the expected *remaining* life time among the careerist then we observe two things.

First, as the expected life time decreases with age the number of crimes decreases also but with a higher degree, i.e. life time crime intensity $v = S/E[T]$ can still be high although the level of participation S and (remaining) expected lifetime $E[T]$ are low. A lower participation among older criminals is obtained if look in details the expected life formula, i.e.

$$E[T] = \int_0^{\infty} tf(t)dt = \int_0^{\infty} G(t)dt = \int_0^{\infty} [1 - F(t)]dt \qquad \text{(I.12)}$$

where $F(t) = Prob[T \leq t]$ is the probability to live at least t years, and $G(t) = 1 - F(t) = Prob[T > t]$ is survival probability of living past t years. A most elementary model of life time distribution that also has some relevance in human populations is the exponential distribution. As $G(t) = e^{-\phi t}$ $(t \geq 0, \phi > 0)$ for exponential distribution, we have

(I.13) $$E[T \mid t < \infty] = \int_0^t e^{-\phi x} dx = \frac{1}{\phi}[1 - \frac{1}{e^{\phi t}}].$$

Now

(I.14) $$\frac{dE[T \mid t < \infty]}{dS}\Big|_{S=S^*} = \frac{d[\frac{1}{\phi}[1 - \frac{1}{e^{\phi t}}]]}{dS} > 0$$

can happen only with increasing age $dt > 0$ if $dS > 0$ since

(I.15) $$d[\frac{1}{\phi} - \frac{1}{\phi e^{\phi t}}] = e^{-\phi t} dt > 0.$$

Thus the number of crimes decreases with less additional years of life. Note that other life time distributions can give opposite results and challenging our basic result that $\frac{dE[T]}{dS}\Big|_{S=S^*} > 0$.

Secondly, the form of time preference for the remaining life may alter the result too. In Appendix III it is shown that an age exists where dS/dt turns negative when expected life time is valued with decreasing time preference when age *increases* $(dt > 0)$, i.e. $e^{-r(t^*)}$ with $r'(t^*) < 0,$ where $t^* = T^* - t$. Thus if the criminal values his remaining planned life time highly, then the number of crimes decreases, i.e. $\frac{dS}{dt} < 0$ when $t > t_1$, and $\frac{dE[T]}{dS}\Big|_{S=S^*} < 0$.

III. Statistical models of arrest counts

Assume at the arrest time t_0 the criminal i has experienced during his lifetime T_i number of earlier arrests denoted as $y_{i|t<t_0}$. Naturally $y_{i|t<t_0} = y_i$ can take only finite number of non-negative integer value like $y_i = 0,1,2,....$ Assume that y_i measures the criminal career of person i before time t_0. This type of phenomena is called a count process $\{Y_t, t \in \mathbb{R}^+\}$, where Y_t is the number of events that have occurred before time t_0. The process is called a Poisson process if the probability of a single occurrence during a brief time interval (exposure time t) is proportional to its duration and if the occurrences in two non-overlapping intervals are independent. Now the probability function of Y has the form

(II.1) $$f(y_i;\lambda,t) = \frac{e^{-\lambda t}(\lambda t)^y}{y_i!} \quad \sim Poisson(y_i;\lambda,t).$$

The Poisson process can be characterized by exponentially distributed waiting times between consecutive events leading to a time-invariant hazard function. This observation (with $t=1$) leads to Poisson –regression model that is a natural starting point in many applications. Thus y_i, given the vector of regressors $\boldsymbol{x_i}$, is independently Poisson distributed with density

(II.2) $$f(y_i \mid \boldsymbol{x_i}) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad y_i = 0,1,2,... \ .$$

The mean parameter is

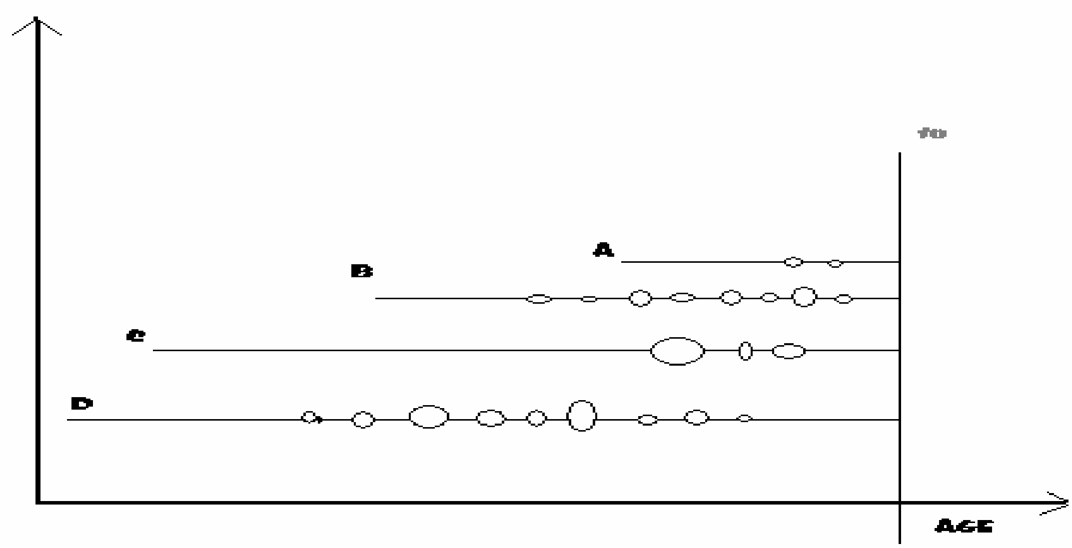
(II.3) $$\lambda_i = \exp[\boldsymbol{x_i}'\boldsymbol{\beta}], \quad i = 1,2,....,N$$

where $\boldsymbol{\beta}$ is $(kx1)$ parameter vector.

However this set-up is not applicable for crime career data analyzed in this paper. The reason is two fold. First we have data for different person with different arrest numbers y_i during his/her life time T_i (i.e. his/her age) up till time point t_0. Naturally in average older criminals have larger crime record than young criminals. Second we have argued that crime intensity is not constant but it is age dependent (i.e. the age-crime curve). This means that basically in Eq. 1) λ_i is time dependent, i.e. $\lambda_i(t)$ for $t > 0$ and the Poisson process is now non-homogenous. Thus we allow for possibility that arrests may occur more likely during certain periods of life time than during other times, i.e. the age-crime profile. We also note that exposure time $t_i \neq T_i$. Typically $t_i < T_i$ since we can not assume that all criminals devote their entire life time to criminal career.

Figure 1. depicts the situation graphically. Person A is relative young (say above 15 years) but has already experienced two arrests. Person B with age of 28 has already a large and increasing crime record. Person C is in his middle ages but his criminal career has started quite late. Person D has a criminal career motivated by our theoretical model. Most of arrests happens at younger years and they decay at later years.

Figure 1. Crime careers and ages



Note that in the Poisson regression model

(II.4) $$E[y_i \mid \boldsymbol{x}_i] = VAR[y_i \mid \boldsymbol{x}_i] = \exp[\boldsymbol{x_i}'\boldsymbol{\beta}].$$

Thus the conditional mean and variance are equal. The model assume that relevant individual factors that cause arrest counts are contained in vector $\boldsymbol{x_i}$. These are typically variables like sex, race, occupation and age. The last one is relevant in our case. Our hypothesis is that age dependency of criminal career is non-linear: first increasing in younger years and then decreasing at older ages. This can be modeled with 2nd order age polynomial like $f(T) = a_0 + a_1 T + a_2 T^2$. The functional form corresponds to our predicted criminal age profile with $a_1 > 0$ and $\mathrm{a}_2 < 0$. Thus the base model have form of

$$E[y_i \mid \boldsymbol{x}_i] = \exp[\boldsymbol{x_i}'\boldsymbol{\beta}] = \exp[\sum_{j=1}^{k} b_{j,i} x_{j,i} + f(T_i)].$$

Note that model is based on the homogenous Poisson process, i.e. λ is constant. Our theoretical model, like Leung's model, leads to an interesting notion that a criminal can "speed up" his crime activity given the time of his/her life devoted to criminal active conditioned on the arrest probability. This observation entails a case where the length of count exposure time varies across the individuals, i.e. the expected count is proportional to the length of the interval during it has occurred. Now for the individual time of exposure t_i homogeneous Poisson process entails that

(II.5) $$\lambda_i = t_i \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i)] = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i) + \ln t_i], \quad i = 1, 2, ..., N.$$

A) OLS and NLS estimation

Eq. II.5) gives an interesting possibility for OLS estimation since under assumption that $t_i = T_i$ we can use approximate model

(II.6) $$\hat{\lambda}_i = \lambda_i / T_i = \boldsymbol{x_i}'\boldsymbol{\beta} + g(T_i) + \varepsilon_i \, .$$

Thus we have a regression model that is not anymore based on counts and Poisson distribution but on exposure time age standardized "counts" with $E[\varepsilon_i] = 0$. Estimation of Eq. 6) gives us some preliminary information concerning how individual factors affect arrest intensity per age. Note also that standardization leads also to age function $g(T) = a_0 + a_1 T$ where our crime career age profile hypothesis corresponds to $a_0 > 0$ and $a_1 < 0$. An alternative estimation is based on NLS-method. The method produces less biased results compared to OLS under assumption that Poisson approach is true one.

(II.6') $$\hat{\lambda}_i = \lambda_i / T_i = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i)] + \varepsilon_i \, .$$

Note with assumption $\varepsilon_i \sim NID(0, \sigma_\varepsilon^2)$ NLS equals MLE, and it gives consistent estimator for model parameters. However the efficiency loss is evident since NLS ignores the inherent heteroskedasticity of Poisson regression.

B) Poisson Regression and unobserved heterogeneity

Holding back to integer count numbers and Poisson model

(II. 7) $$\lambda_i = t_i \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i)] = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i) + \ln t_i] \, , \quad i = 1, 2, ..., N \, ,$$

we observe that this formula is problematic since it contains both exogenous time effect, T_i = age, and endogenous time effects, the optimal criminal career t_i (= exposure time). Typically, like in this study, we do not have data for criminal's exposure time (or the period of risk) for criminal activity, i.e. waiting times between subsequent arrests or his total crime career period. One crude solution (used already above in Eq. II.6) is to assume that $t_i = T_i$ meaning that criminal's age equals the crime exposure time. Note that this result was obtained in our theory model above. Obviously they differ but they are expected to be highly correlated at least for life career criminals. This notion makes the statistical estimation of Eq. II.7) unreliable. However we can assume that T_i and t_i are related to each other randomly, i.e. $t_i = T_i \varepsilon_i$ where ε_i is some random variable defined on limits $(0,1)$. Thus part of person's life is selected randomly to the criminal career. This formulation preserves some age dependence on exposure time. Alternatively we can argue that we can not measure t_i correctly. We only observe T_i that is augmented with measurement error ε_i to correspond the unobservable t_i.

The main weakness of Poisson model is the assumption that $E[y_i \mid \boldsymbol{x}_i] = VAR[y_i \mid \boldsymbol{x}_i]$ stemming from the fact that the intensity of Poisson process is a deterministic function of the covariates. No unobserved heterogeneity is allowed for. Likewise assumption of independent random counts over time is questionable, i.e. occurrences influence the probability of future occurrences (positive count occurrence dependency or positive contagion). Unobserved heterogeneity and positive contagion lead to over-dispersion phenomena $E[y_i \mid \boldsymbol{x}_i] < VAR[y_i \mid \boldsymbol{x}_i]$. The former is easily seen if we assume that

$$\text{(II.8)} \qquad \tilde{\lambda}_i = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i\ ] = \exp[\boldsymbol{x_i}'\boldsymbol{\beta}]\mu_i$$

where u_i captures the non-modeled unobserved heterogeneity and $corr[\boldsymbol{x}_i, \varepsilon_i] = 0$. Now $E[\tilde{\lambda}_i] = \lambda_i$ and of $VAR[\tilde{\lambda}_i] = \lambda_i^{\,2}\sigma_{u_i}^2$ if we scale $E[\mu_i] = 1$. Thus

$$VAR[y_i] = E[VAR[y_i \mid \tilde{\lambda}_i]] + VAR[E[Y_i \mid \tilde{\lambda}_i]]$$

(II.9)

$$= E[\tilde{\lambda}_i] + VAR[\tilde{\lambda}_i] = \lambda_i + \lambda_i^2 \sigma_{u_i}^2 > \lambda_i .$$

Note that above we proposed the alternative that $t_i = T_i \varepsilon_i$. Now noticing this in Eq. II.8) gives

$$\text{(II.10)} \qquad \tilde{\lambda}_i = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + \ln(T_i \varepsilon_i)\,] = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + \ln T_i + \ln \varepsilon_i] = T_i \exp[\boldsymbol{x_i}'\boldsymbol{\beta}]\varepsilon_i .$$

Eq. II.10) gives the possibility to use age as exposure time variable but the price is the over-dispersed count model that is not anymore a Poisson process model. The specification in Eq. 10) entails that we can use once again function $f(T) = a_0 + a_1 T + \alpha_2 T^2$ with $a_1 > 1$ and $a_2 < 0$ to preserve the argued inverted U -shape age response to arrests counts. Note that specifications in Eq. II.7) and Eq. II.10) imply that we have to use parameter constraint $\theta = 1$ in the models

$$\text{(II. 7')} \qquad \lambda_i = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i) + \theta \ln t_i] = t_i \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i)] ,$$

$$\text{(II.10')} \qquad \tilde{\lambda}_i = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i) + \theta \ln(T_i \varepsilon_i)\,] = T_i \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + f(T_i)]\varepsilon_i .$$

These specifications support the basic assumption of count models that intensity rate λ_i is constant and preserve the proportionality to exposure time, i.e. doubling the exposure time doubles the expected number of counts. Form point of view of age-crime curve testing this means that rejecting the constraint $\theta = 1$ refers to non-homogenous Poisson- or count process where the exposure time affects the number of counts in non-linear way. This means that age and exposure time are partly endogenously selected to number of crimes, not only exogenously determining them. Thus rejecting the hypothesis $\theta = 1$ for alternative $\theta > 1$ in the presence of $f(T) = a_0 + a_1 T + a_2 T^2$ in the model is actually test for age-crime curve hypothesis augmented with indication of deeper level age dependency of crime rates. Note that it is quite easy to show that if $\lambda_i(t \mid age)$ is increasing function of age we have the case of $E[y_i \mid \boldsymbol{x}_i] < VAR[y_i \mid \boldsymbol{x}_i]$ once again.

Thus the equality of conditional expectation and variance is not realistic assumption for many applications of count data. The over-dispersion case is also evident under the presence of positive occurrence dependence.[2)] Much used Negative Binomial (NegBin) model alternative allows for *over-dispersion*, i.e

(II.11) $$VAR[y_i \mid \boldsymbol{x}_i] = E[y_i \mid \boldsymbol{x}_i] + \sigma^2 [E[y_i \mid \boldsymbol{x}_i]]^2$$

leading to model alternative also for the conditional variance

(II.12) $$\sigma_i^2 = 1 + \exp(\boldsymbol{z_i}'\boldsymbol{\gamma}),$$

where $\boldsymbol{z_i}$ is vector of some explanatory variables. When $\gamma = 1$ we get a scalar dispersion parameter, otherwise we estimate a variance function. Some more general models allow for under-dispersion (see Winkelmann & Zimmermann 1995, Winkelmann 2003, Cameron and Trivedi 1998). Negbin –model alterative can easily to show preserve positive occurrence dependence. If an arrest occurrence increases the probability of next occurrence then the Polya urn schema gives NegBin distribution.

2) Negative contagion causes underdispersion. This can be seen in following way

$$\Pr ob[Y = r] = \begin{cases} 0, \quad \text{for } r = 0 \\ Poisson(r, \alpha) - Poisson(r+1, \alpha), \text{ for } r = 1, 2, \ldots \end{cases}.$$

If $\alpha > 1$ (the parameter for positive duration dependence in Gamma distributed waiting time) then Poisson count process exhibits under-dispersion (Winkelmann 1995). Note that positive contagion (negative duration dependence) can also rise from aggregation of individuals having different waiting times (i.e. different propensity to experience an event).

We stress that above developments and comments do not imply that we have solved the statistical problems of unobserved heterogeneity and (positive) contamination - far from it. We have only shown that all assumptions of Poisson model are not fulfilled in our model specification of arrest counts and suggested specification of age effects leads to richer model alternative. One alternate to be estimated combines normal distributed unobserved heterogeneity in NegBin –model.

C) Semiparametric estimation

The Poisson equi-dispersion model can still be questioned when we face non-homogenous Poisson process $\lambda_i(t)$ for observed counts. A partly solution to this problem is to use richer functional presentation for exposure time effects in Poisson model, i.e. we have to have model like

(II.13) $$\lambda_i(t) = g(t_i)\exp[\boldsymbol{x}_i'\boldsymbol{\beta}],$$

where $g(t_i)$ is a some function that models the time (e.g. age) dependent crime event occurrence rate. The model in Eq. II.13) allows us to separate the age class effects and individual age effects on crime counts. $g(t_i)$ measures the age class effects, and $\boldsymbol{x}_i'$ - still including $f(T) = a_0 + a_1T + a_2T^2$ - measures the individual crime rate effects. Because our sample of age observations is used to model both these effects some special model and estimation alternative are used.

Assume next that non-homogenous Poisson process with time dependent intensity function fulfills the following proportional property

(II.14) $$\lambda_{xi}(t) = \lambda_{0i}(t)\exp[\boldsymbol{x}_i'\boldsymbol{\beta}],$$

where $\lambda_{0i}(t)$ is the baseline intensity function. The corresponding cumulative or integrated intensity function is

(II.15) $$\Lambda_{xi}(t) = \int_0^t \lambda_{xi}(u)du = \Lambda_{0i}(t)\exp[\boldsymbol{x}_i'\boldsymbol{\beta}],$$

where $\Lambda_{0i}(t) = \int_0^t \lambda_0(u)du$. Suppose that there are independent observations on m individuals and he/she is observed over the time interval $(0, T_i)$. Let n_i be events (arrests) observed to occur at times $t_{i1} < t_{i2} < .... < t_{in_i}$. Given the *age* information we can always order T_i's so that $0 < T_1 \le T_2 \le .... \le T_m$ with $T_l \ge t_{in_i}$. Letting $T_0 = 0$ and $n(t)$ represent the total number of events (arrest) in $(0, t)$, we can write for $i = 1, 2, ..., m$

$$E[n(t) - n(T_{i-1})] = [\Lambda_0^*(t) - \Lambda_0^*(T_{i-1})]\sum_{l=i}^m \exp(\boldsymbol{x}_k'\boldsymbol{\beta}), \quad \text{for} \quad T_{i-1} < t \le T_i,$$

where $\Lambda_0^*(t)$ corresponds to $\lambda_0^*(t) = \lambda_0 e^{\beta_0}$. Note that in this application context both t and T_i refer to different age classes, i.e. we estimate average proportional age class crime intensities.

This suggest the estimate for $i = 1, 2, ..., m$

(II.16) $$\hat{\Lambda}_0^*(t) = \sum_{n_i=1}^{i-1} \frac{n(T_{n_i}) - n(T_{n_i-1})}{\sum_{l=n_i}^m e^{x_l'\hat{\beta}}} + \frac{n(t) - n(T_{i-1})}{\sum_{l=i}^m e^{x_l'\hat{\beta}}}, \; T_{i-1} < t \le T_i.$$

Eq. II.16) entails a two step estimation routine with equations

(II.17a) $$\frac{\partial l_2}{\partial \beta_r} = \sum_{i=1}^m n_i x_{ir} - \sum_{i=1}^m \hat{\Lambda}_0^*(T_i) x_{ir} e^{x_i'\beta} = 0, \quad r = 1, 2, ..., k$$

and

(II.17b) $$\Lambda_0^*(T_i) = \sum_{n_i=1}^{i-1} \frac{n(T_{n_i}) - n(T_{n_i-1})}{\sum_{l=n_i}^m e^{x_l'\hat{\beta}}}, \quad i = 1, 2, .., m,$$

where l_2 is the loglikelihood from likelihood $L_1(\Lambda_0^*)L_2(\Lambda_0^*, \boldsymbol{\beta})$. Thus we estimate first $\Lambda_0^*(T_i)$ from Eq. II.17b) with by setting $\tilde{\boldsymbol{\beta}} = \boldsymbol{0}$ and then by using the gradient condition II.17a) a ML –estimate is obtained for $\hat{\boldsymbol{\beta}}$. Inserting this back in II.17b) provides a new estimate for $\Lambda_0^*(T_i)$. By repeating this iterative 2-step produce convergent estimates for $\Lambda_0^*(T_i)$ and $\boldsymbol{\beta}$ are obtained (for more details, see Lawless 1987). In principle this method separates and estimates the average proportional age class intensities and individual control effects on observed individual crime counts.

IV. Data and Estimation Results

IV.1. Data

Our data consists of felony defendants in large urban counties in U.S in year 1998 (USDoJ/BJS: State Court Processing Statistics, 1998). In the 1990's Bureau of Justice tracked every second year a sample of felony cases during the month of May in 75 largest counties in the US. Thus the follow up (or sample exposure) time t_0 is one month. The original sample consisted of 15878 arrested individuals with following variables

PRIARR = Number of prior arrests (0,1,2,….,N), AGE = Age of defendant.
PRICONV = Number of prior convictions.
PRIJAIL = Number of prior jail incarcerations.
PRIPRIS = Number of prior prison incarcerations
RACE = White: 1, Black: 2, American Indian or Alaskan Native: 3, Asian: 4.
SERARR = Most serious prior arrest. Misdemeanor: 1, Felony: 2, No Prior Arrest (NPA): 0.
SERCONV = Most serious prior conviction. Misdemeanor: 1, Felony: 2, No Prior Conviction (NPA): 0.
SEX = Male: 1, Female: 2.
STATE = state where the arrest took place. A qualitative index that was transformed to ascending numerical index with state population size.

After excluding missing observation from each variable the sample reduced to 6827 observations. The main interest variables are PRIARR (number of prior arrests in May 1998) and AGE (age of defendant). Table 1 gives the main summary statistics of continuous variables and Table 2 reports the distributions of discrete variables. The age range between defendants is 13 – 81 with mean of 30.6 years. Past criminal record (arrests) is found among 72% of them and earlier prison convictions are also typical. All variables are skew to right and highly peaked compared to normal distribution. The age distribution is most close to normal. Most of offenders are male and black (79.8% and 44.1%). Many of them has multiple type of criminal record.

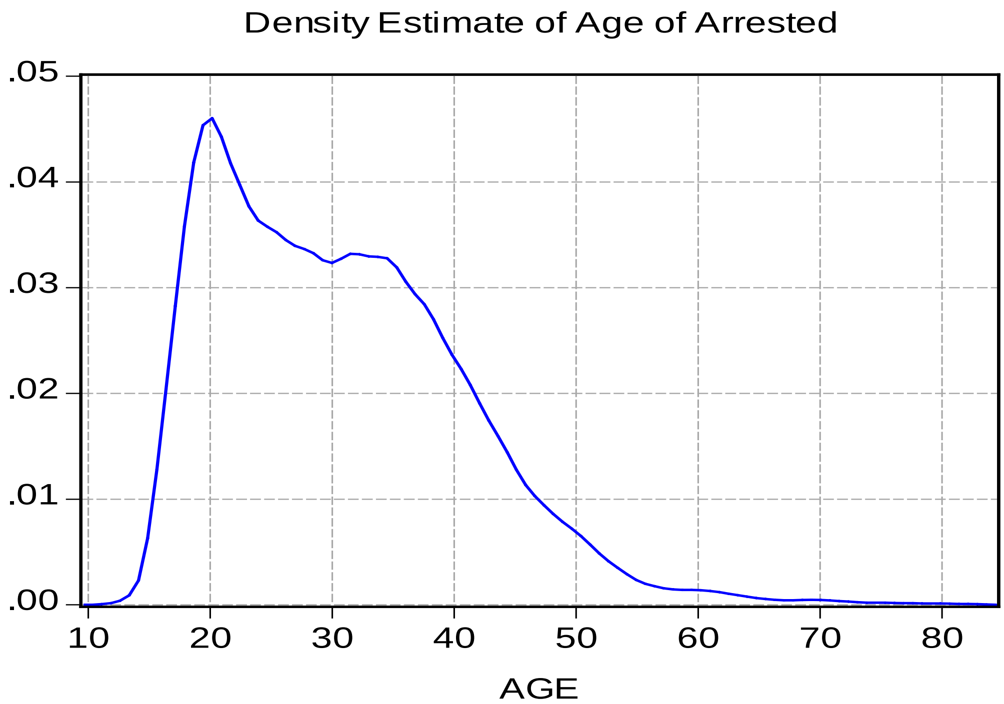
Table 1. Summary statistics of continuous variables.

	PRIARR	PRICONV	PRIJAIL	PRIPRIS	AGE
Mean	7.222	2.878	1.418	0.431	30.607
Median	3.000	1.000	0.000	0.000	29.000
Maximum	114	52	39	16	81
Minimum	0	0	0	0	13
Std. Dev.	10.497	4.559	2.833	1.208	10.043
Skewness	2.861	3.134	4.227	4.641	0.829
Kurtosis	15.123	18.356	31.185	33.567	3.768
% X > 0	72.2% (4932)	58.6% (4004)	43.1% (2931)	19.3% (1321)	

Table 2. Distributions of discrete variables.

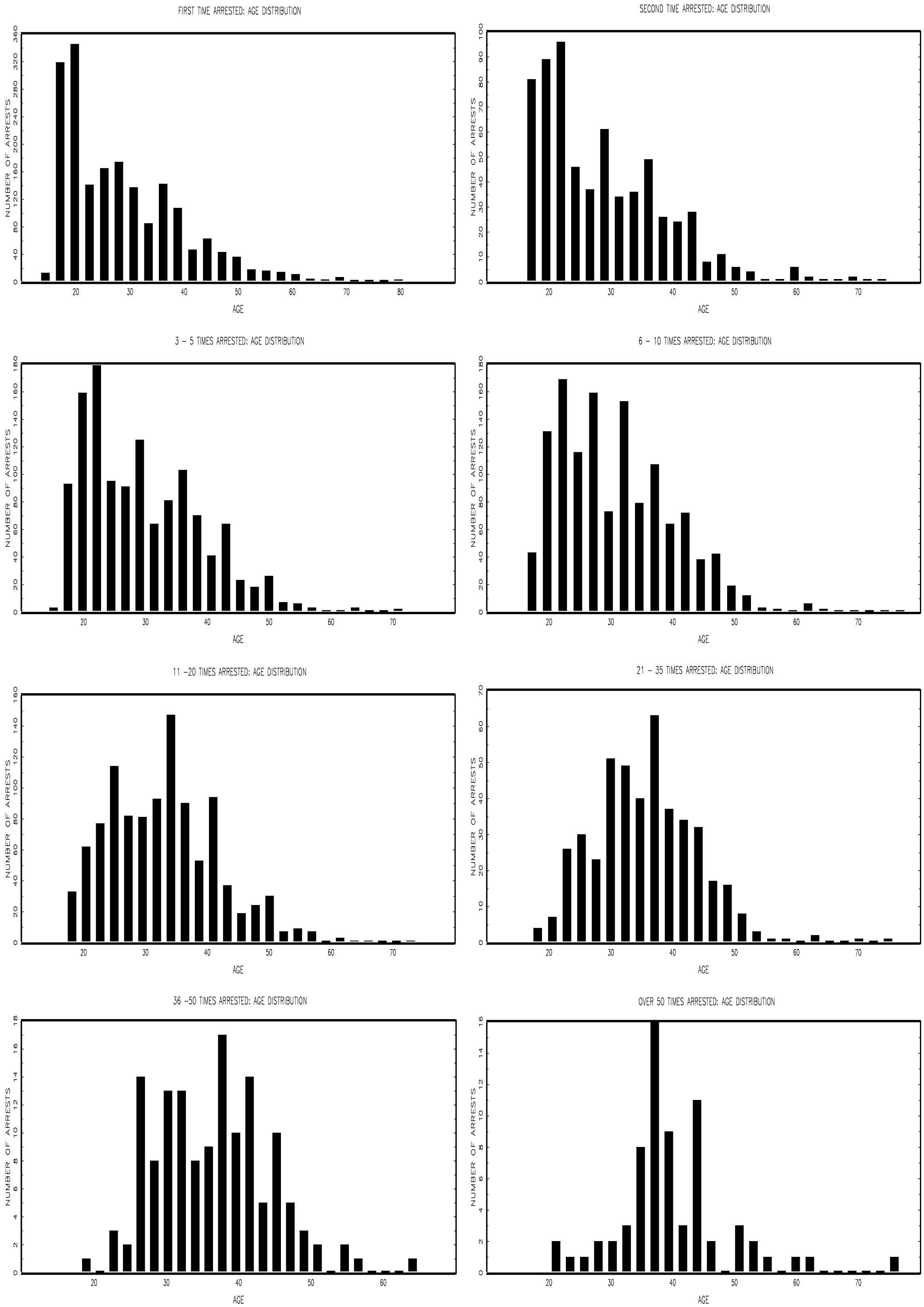
SEX	**RACE**	**SERARR**	**SRCONV**
MALE 5447 (79.8%)	WHITE 2281 (33.4%)	MISDEM. 861 (12.6%)	MISDEM 1265 (18.5%)
FEMALE 1380 (20.2%)	BLACK 3013 (44.1%)	FELONY 4071 (59.6%)	FELONY 2739 (40.4%)
	INDIAN 71 (1%)	NPA 1895 (27.7%)	NPA 2823 (41.4%)
	ASIAN 82 (1.2%)		

Figure 2. Age Distributions of Arrested



Figures 2-10 give a closer look at age distribution of analyzed sample of arrested. Figures basically tell the story of age dependency of crime participation. The peak is obtained age 20 but other lower peak is found at age 32. Thus the unimodality of aggregate age-crime profile is not valid in this sample but right skewness is obvious. Next eight pictures (age distributions of arrested in different arrest number classes) give the explanation for found density estimate. The peak number of arrests shift toward right and the distributions normalizes with age. Most importantly the number of arrested decreases fast with the age and the number of arrests.

Figures 3-10. Age Distributions of Arrested with Different Arrest Numbers



Figures 11. Distribution of Number of Arrests

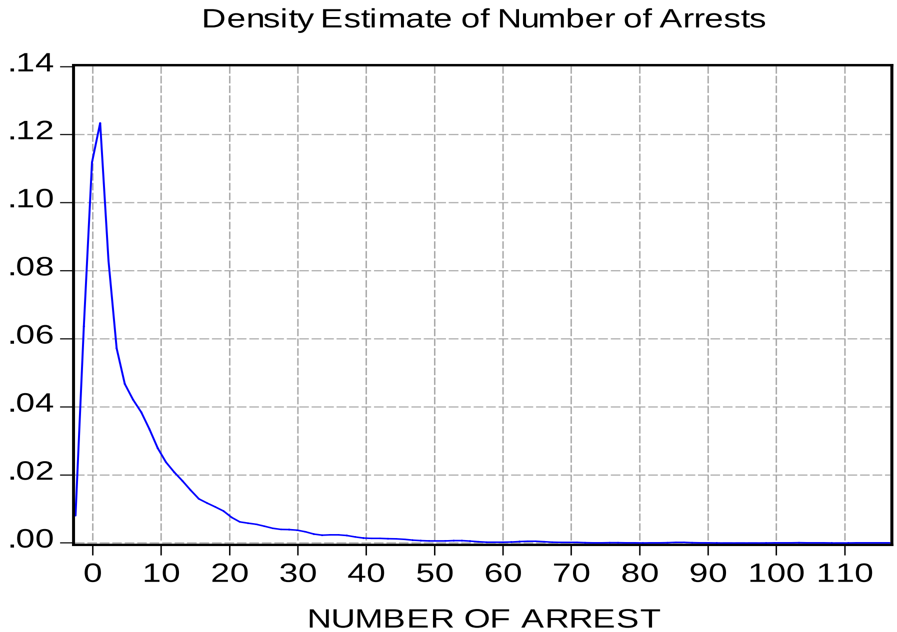


Figure 11 shows that arrest number declines rapidly after the first arrest in sample. The distributions are similar in all age classes (see Appendix IV). Thus shape of relative crime frequency in different age classes remains same but the number of arrested participation) is lower in higher age classes. Finally Box –plots in Figures 12 and 13 sum the information above. Figure 12 shows the distributions of number of arrests at different ages. Figure 13 tells the story in the opposite way: the age distributions of given arrest numbers.

Age class crime intensity distributions in Figure 12 are clearly skewed, mean values increase with the age, and in age classes above 57 years distributions are very diffuse. However after age 45 the increase ceases. The number of defendants is low in higher age classes due the voluntary and involuntary drop-out selection of crime careers. The Box-plots in Figure 13 confirm us the earlier results that crime intensity increases with age but many outliers are found with arrest number less than 20. Age distributions with high arrest numbers are also very diffuse. Expected age of defendants increase with number of arrest when number of arrests is below 25 but after that it seems to decline or stabilize.

Figure 12. Distributions of Number of Arrests in Age Classes

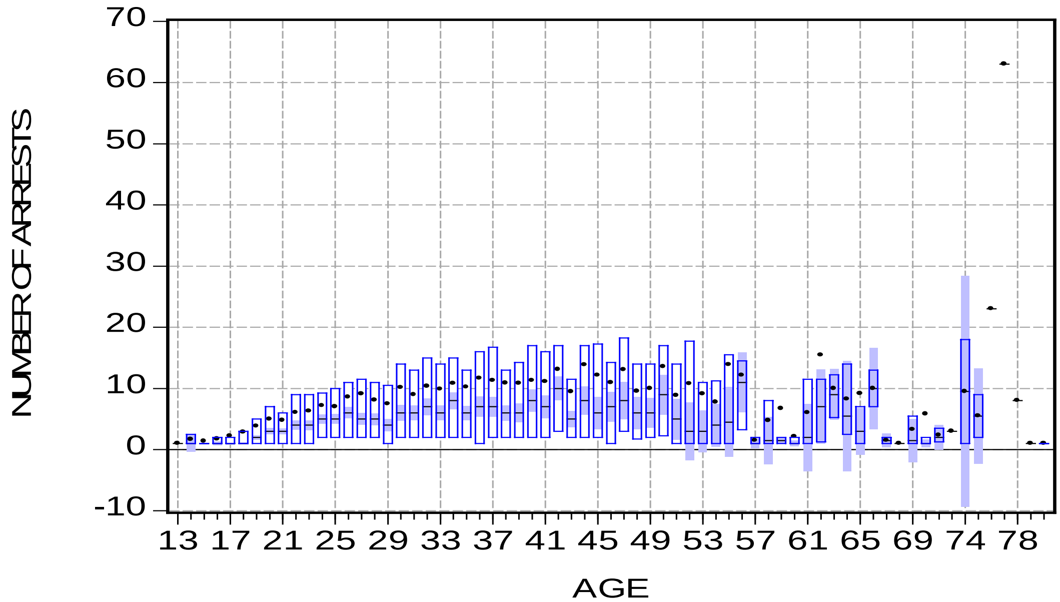
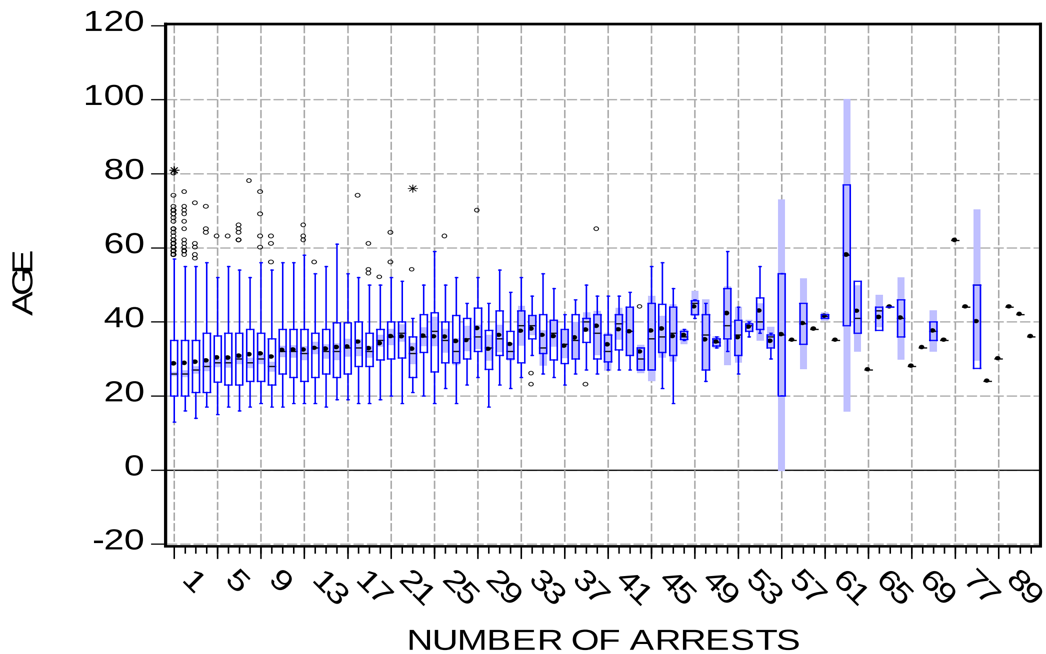
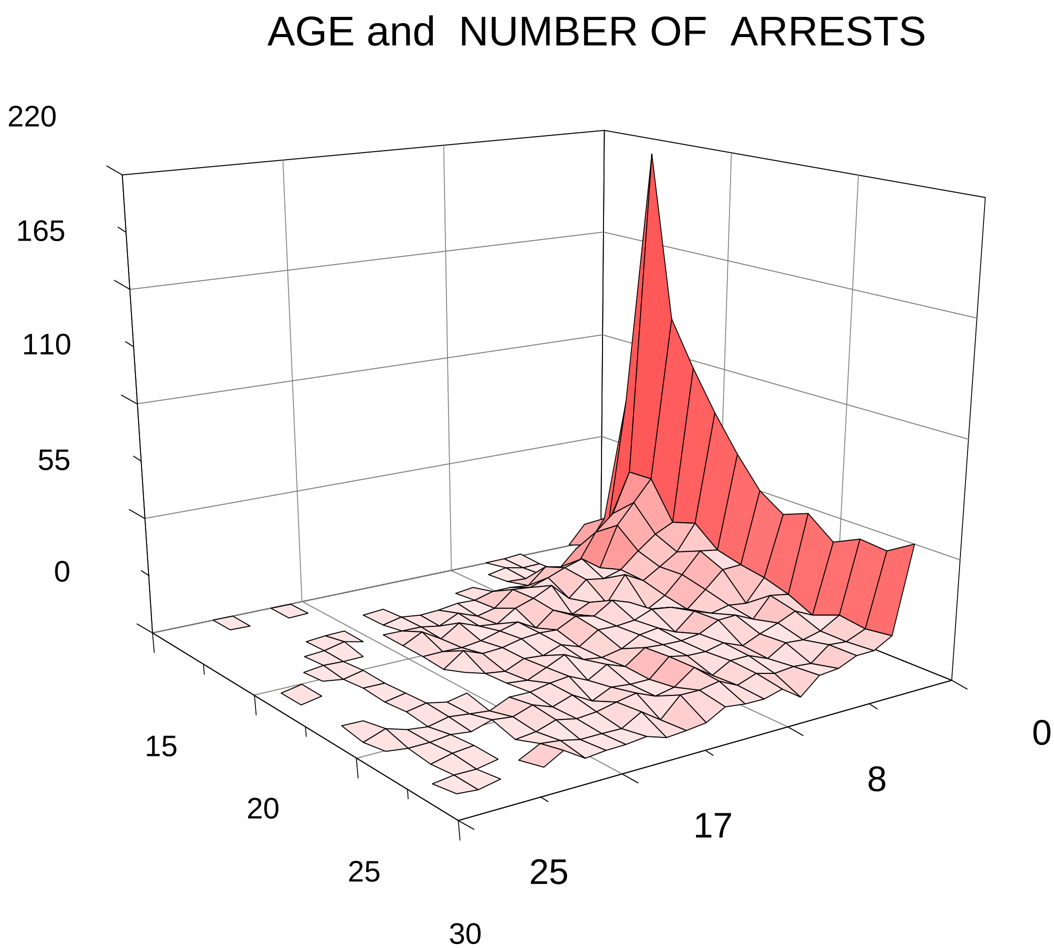


Figure 13. Distribution of Number of Arrests



Generally the figures support the view that inverted U-shape age-crime profile is valid in analyzed sample but tail behaviour of marginal distributions of two-dimensional age and arrest number distribution are diffuse (see also Figure 14). This makes the conditional modeling of number of arrests with given age of arrested, i.e. $E[Y \mid X = age]$, a demanding task.

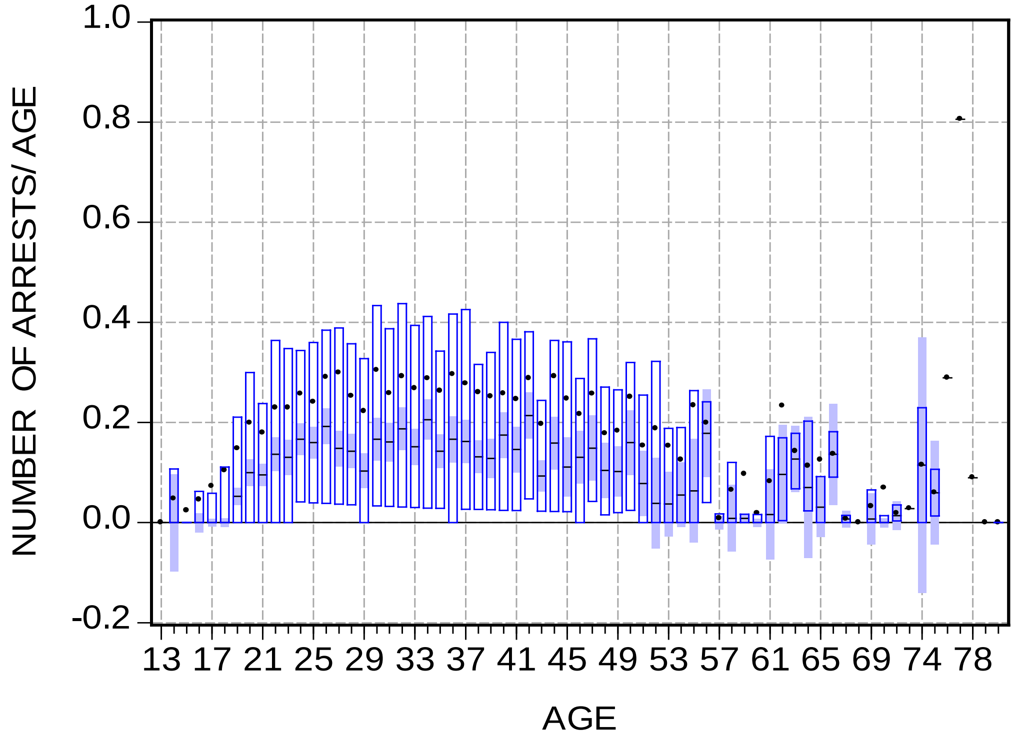
Figure 14. Age and number of arrests in 3D



V. Estimation results

We first report least square estimation results on age standardized intensities λ_i / T_i. Figure 15. below depicts the age distribution of λ_i / T_i (see also Appendix V). We notice that age response to λ_i / T_i gives the age-crime profile shape. In Figure 15 number of arrested per age increases faster than age under age class 27 and after it the converse happens. The increasing mean intensity values are in accordance with Leung's assumption of increasing age-crime intensity. The OLS estimation with different

Figure 15. Age and age standardized crime counts



covariates show that negative age effects dominates in multivariate regression setting. Number of prior convictions (PRICONV) does not halter age standardized number of arrested. Prior prison and jail incarcerations (PRIJAIL, PRIPRIS) do not affect them in statistically significant way. However the severity of prior arrests (SERARR) increases standardized arrests and prior convictions effects are non-significant. Non-white males are most often arrested and larger the state (city) less arrests happen. NLS –estimation of Poisson type model with $F(T_i) = a_0 + a_1 T_i + \alpha_2 T_i^2$ on age standardized arrest counts are close to OLS estimation results. Note negative a_2 corresponds to bell shaped age-crime profile hypothesis. However the coefficient estimates of PRIJAIL, PRIPRIS, and STATE lose their statistical significance. Coefficient for SERCONV turns positive and is statistically significant.

Table 3. OLS on PRIARRG = arrests/age: $\lambda_i / T_i = \boldsymbol{x_i}'\boldsymbol{\beta} + \alpha_0 + \alpha_1 T_i + \varepsilon_i$

Method: Least Squares, N = 6827
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.154841	0.010264	15.08573	0.0000
AGE	-0.003618	0.000234	-15.47848	0.0000
PRICONV	0.043868	0.002689	16.31571	0.0000
PRIJAIL	0.004997	0.003504	1.425968	0.1539
PRIPRIS	-0.004473	0.004193	-1.066720	0.2861
RACE	0.009234	0.003527	2.617997	0.0089
SERARR	0.083795	0.003423	24.47661	0.0000
SERCONV	-0.005002	0.005124	-0.976094	0.3291
SEX	-0.034881	0.004407	-7.915146	0.0000
STATE	-0.002290	0.000310	-7.382945	0.0000

R-squared	0.605303	Mean dependent var	0.229880
Adjusted R-squared	0.604782	S.D. dependent var	0.315579
S.E. of regression	0.198393	Akaike info criterion	-0.395666

Table 4. NLS on PRIARR =arrests/age: $\lambda_i / T_i = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + a_0 + a_1 T_i + \alpha_2 T_i^2] + \varepsilon_i$

Method: Non-Linear Least Squares, N = 6827
White Heteroskedasticity-Consistent Standard Errors & Covariance
================= ======== ========== ========== ========

	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.760519	0.227273	-20.94628	0.0000
AGE	0.026478	0.012039	2.199385	0.0279
AGE2	-0.000617	0.000181	-3.410445	0.0007
PRICONV	0.046751	0.006835	6.840137	0.0000
PRIJAIL	0.005105	0.007384	0.691324	0.4894
PRIPRIS	0.009112	0.013265	0.686907	0.4922
RACE	0.073747	0.027966	2.637073	0.0084
SERARR	1.023955	0.041260	24.81730	0.0000
SERCONV	0.217287	0.023041	9.430374	0.0000
SEX	-0.176210	0.043011	-4.096840	0.0000
STATE	-0.002308	0.002339	-0.986915	0.3237

R-squared	0.534479	Mean dependent var	0.229880
Adjusted R-squared	0.533796	S.D. dependent var	0.315579
S.E. of regression	0.215475	Akaike info criterion	-0.230334

Table 5. MLE with Poisson model on PRIARR =arrests

$$\lambda_i = \exp[\ln T_i + \boldsymbol{x_i}'\boldsymbol{\beta} + a_0 + a_1 T_i + \alpha_2 T_i^2]$$

Dependent Variable: PRIARR, N = 6827
Method: ML/QML - Poisson Count (Quadratic hill climbing)
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
ln(AGE)	1.00			
C	-5.226	0.168	-32.30	0.000
AGE	-0.0038	0.0081	-0.478	0.636
AGE2	-0.00017	0.00011	-1.578	0.115
PRICONV	0.0637	0.0072	8.851	0.000
PRIJAIL	0.0015	0.0075	0.206	0.836
PRIPRIS	-0.0007	0.0125	-0.062	0.951
RACE	0.050	0.0210	2.365	0.018
SERARR	1.365	0.0352	38.31	0.000
SERCONV	1.188	0.0244	7.731	0.000
SEX	-0.209	0.0331	-6.031	0.000
STATE	-0.0082	0.0018	-4.476	0.000
R-squared	-	Mean dependent var		7.222
Adjusted R-squared	-	S.D. dependent var		10.497
Log likelihood	-20561.27	Restr. log likelihood		-48103.79
Test of restriction θ=1 for θln(AGE)	98.55 (p-value: 0.00)			
Over-dispersion test	13.03 (0.00)			
Katz-family test	23.51 (0.00)			
Test against NegBin	998.2 (0.00)			

Results of ML–estimation of Poisson and Negative Binomial regression models with restriction $\theta = 1$ against the alternative $\theta > 1$ are confusing. The coefficient values and their statistical significances do not conflict the results for NLS–estimation but age variables (AGE and AGE2) get non-significant or opposite results compared to OLS/NLS –estimation rejecting the age-crime profile hypothesis. Over-dispersion and Katz family tests reject the Poisson alternative. Also the test against NegBin –alternative is rejected. However Vuong test favors Poisson model against NegBin -alternative with Normal unobserved heterogeneity with test value of -16.11 (see the likelihood values). Evidently the results stem from the invalid restriction θ=1 on *ln(AGE)* since Poisson and NegBin model estimates without variable *ln(AGE)* gives similar results as NLS –estimation for $F(T_i) = a_0 + a_1 T_i + \alpha_2 T_i^2$ (see Appendix VI). The intensity process is non-homogenous.

Table 6. MLE with Negative Binomial Model on arrests with Unobserved Normal heterogeneity

$$\lambda_i \mid \varepsilon_i = \exp[\boldsymbol{x_i}'\boldsymbol{\beta} + a_0 + a_1 T_i + \alpha_2 T_i^2 + \theta \ln T_i \varepsilon_i], \quad VAR[y_i \mid x_i] = \phi_i = 1 + \sigma^2$$

$$\varepsilon_i \sim N(0, \rho^2)$$

Dependent Variable: PRIARR N = 6827
Method: ML - Negative Binomial Count (Quadratic hill climbing)
QML (Huber/White) standard errors & covariance

	Coefficient	Std. Error	z-Statistic	Prob.
ln(AGE)	1.00			
C	-12.145	0.434	-27.951	0.000
AGE	-0.193	0.028	-6.754	0.000
AGE2	0.018	0.0005	37.815	0.000
PRICONV	0.064	0.0119	5.385	0.000
PRIJAIL	0.0020	0.0157	0.131	0.897
PRIPRIS	-0.0002	0.0297	-0.008	0.998
RACE	0.0492	0.0307	1.601	0.109
SERARR	1.362	0.0450	30.241	0.000
SERCONV	0.182	0.0026	7.007	0.000
SEX	-0.215	0.0652	-3.297	0.001
STATE	-0.008	0.00242	-3.447	0.000
	Over-dispersion Parameter			
1+ϕ	10.31	0.244	42.11	0.0000
	Standard dev. of heterogeneity			
ρ	0.283	0.032	8.883	0.000

R-squared	-	Mean dependent var	7.222352
Adjusted R-squared	-	S.D. dependent var	10.49760
Log likelihood	-27099.80	Restr. log likelihood	-48103.79
Test of restriction θ=1 for θln(AGE)	44.12 (p-value: 0.00)		

Results with semiparametric proportional cumulative intensity function are reasonable. The estimation is conducted for age classes 13 – 56 since above age class 56 the intensities are very unstable (see Figure 13 or 15). We lose 98 observations of full sample (1.4% of sample) due to the truncation. Most important finding, compared to earlier results, is the semi-parametric estimate for $\hat{\Lambda}_0(t)$ that controls much of variability of age class dependent intensity (see Figure 16). The outcome is more transparent when non-cumulative intensities $\lambda^*(t)$ are depicted (see Figures 17a and 17b). The semiparametric presentation is much less in scale and flatter than the non-parametric data estimate for cumulative intensity ($\boldsymbol{\beta} = \boldsymbol{0}$).

Table 7. MLE with Semiparametric Proportional Cumulative Intensity

$$\hat{\Lambda}_0^*(t) = \sum_{n_i=1}^{i-1} \frac{n(T_{n_i}) - n(T_{n_i-1})}{\sum_{l=n_i}^{m} e^{\mathbf{x}_l'\hat{\boldsymbol{\beta}}}} + \frac{n(t) - n(T_{i-1})}{\sum_{l=i}^{m} e^{\mathbf{x}_l'\hat{\boldsymbol{\beta}}}}$$

Dependent Variable: PRIARR N = 6729 AGE CLASSES 13 – 56
Method: ML (Newton-Raphson)
White Heteroskedasticity-Consistent Standard Errors & Covariance

	Coefficient	Std. Error	z-Statistic	Prob.
AGE	0.0342	0.0022	15.59	0.000
AGE2	-0.0004	0.00003	-13.61	0.000
PRICONV	0.066	0.0009	71.032	0.000
PRIJAIL	0.0007	0.0004	1.787	0.043
PRIPRIS	-0.0045	0.0007	-6.037	0.000
RACE	0.0380	0.0080	4.728	0.000
SERARR	1.3390	0.0115	116.422	0.000
SERCONV	0.1792	0.0061	29.157	0.000
SEX	-0.2232	0.0060	-36.947	0.000
STATE	-0.0082	0.0004	-20.627	0.000

R-squared	-	Mean dependent var	7.2239
Log likelihood	-12431.32	S.D. dependent var	10.4636

Note that construction of Poisson model with semiparametric intensity assumed that model does not include a constant term. Every age class a specific constant term, that is $\hat{\Lambda}_0(t)$ or $\lambda^*(t)$ for $0 < T_1 \leq T_2 \leq .... \leq T_m$, although all classes share the same parametric presentation for crime counts.

The results imply that the age-crime profile in still bell shaped when estimation controls both for individual and average age class intensity effects. The result does not support Leung's conjecture of increasing intensity with age. The results are more in line with model implications derived earlier in the paper: both the crime intensity and participation are decreasing with age. Note that we have not provided 95% confidence intervals for average age intensity estimates. Anyway we have succeeded to separate age class dependent intensities and individual control effects on crime counts. The semiparametric estimation provides more precise estimates than alternative models since the non-homogenous intensity presentation is also estimated in the presence of the individual controls, including age, for the counts.

Figure. 16. Cumulative proportional intensities

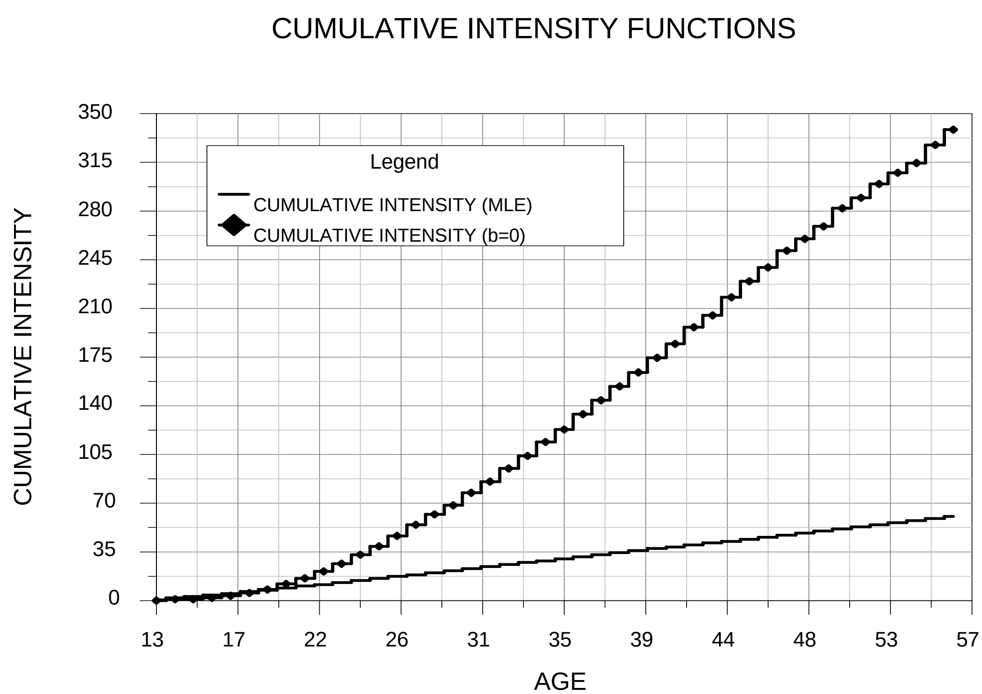


Figure 17a. Age class average intensities

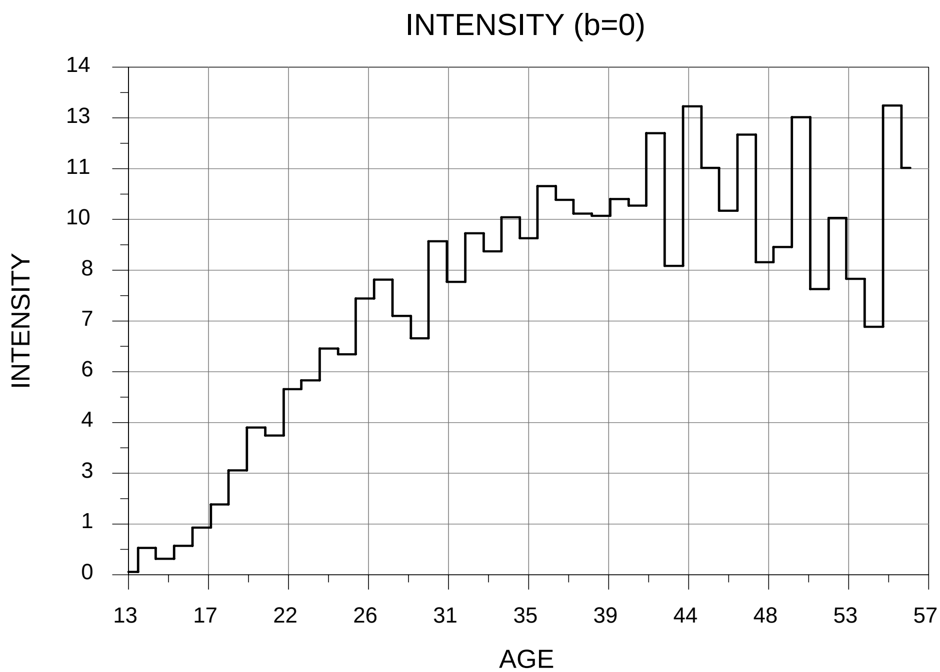
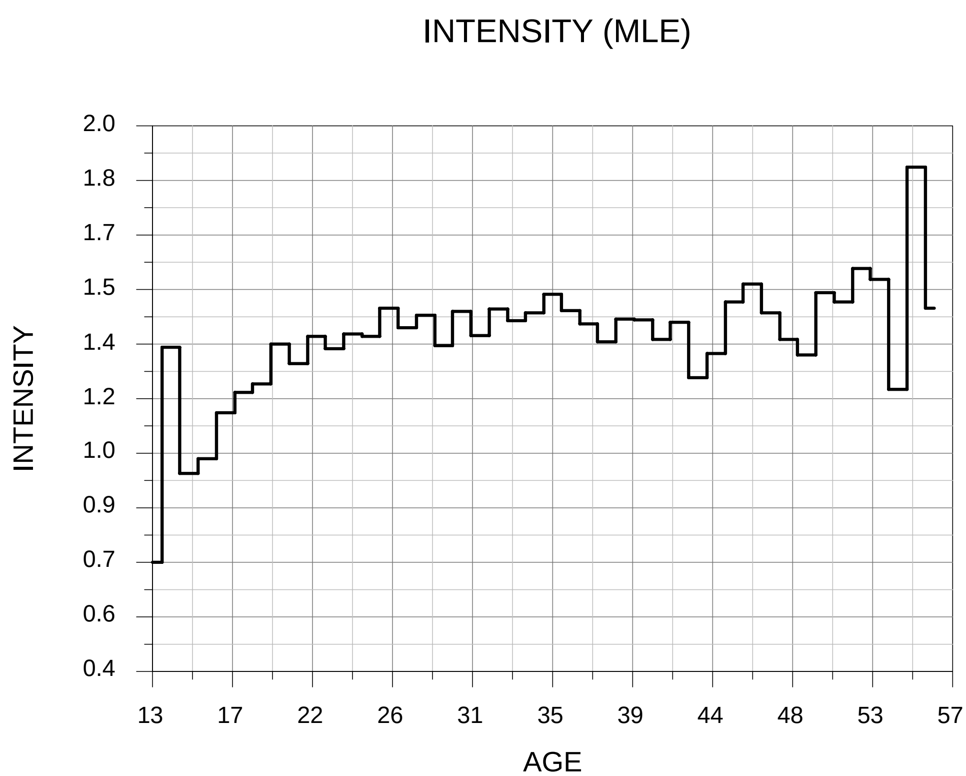


Figure 17b. Semiparametric average age class intensities



VI. Conclusions

Two classical issues in criminology, criminal career paradigm and age-crime curve (or profile) rule, were re-opened into analysis with economic analysis. Although age-crime curve and criminal career (or recidivism) are conceptually distinct they are closely interlinked empirically to each other. The career paradigm can't be put aside when analyzed data contains individuals with repeated arrests. Age-crime curve and profile sum up the crime intensity and participation aspects of individual crime behaviour during his/her lifetime. The age dependency is relevant for both the intensity and the participation. The latter is typically considered to produce the observed age-crime curve as the repeated criminal activity calms down voluntary or involuntary after age of 25.

However this answer is too simple, and partly wrong, since observed individual criminal intensity, e.g. number of crimes per year, varies substantially also with age. We have tried argue like Leung (1994) that crime intensity must be the starting point of the proper analysis, since if want to understand crime career and age-crime curve regularities with economic terms, the whole crime career is the exposition time during the crimes are committed. Thus the economic modeling takes the crime intensity as age dependent decision variable where both the number of crimes and the length of crime career generate non-homogenous intensity rate.

A simple model that included the salient features of crime career and age-crime curve was proposed. It was assumed that individuals can control their crime intensity. The criminal sets his crime intensity rate at level that minimizes his expected time devoted to criminal activity incurring some costs. By augmenting the basic model with fixed expected life time during the optimal number of crimes can be committed we were able to show that increased number of crimes increases the expected life time. The result gives an explanation to question why old criminals necessarily do not commit less crime per their age compared to young criminals. Thus their intensity rate can still be high but the number of old criminals is less among the total number of criminals due the participation selection effect. The assumption of age dependent subjective time preference and shortening remaining life emphasize the model results where crimes are committed at the beginning of life time.

In the empirical part of study the arrests count data of felony defendants in large urban counties in U.S in year 1998 was analyzed with count-data regression methods. It was shown that Poisson and NegBin models give unsatisfactory results concerning the age dependency of number of arrest counts as the models are based on constant intensity rate. Much better results were obtained with NLS–method on individual age standardized counts, and with semi-parametric Poisson estimation on individual arrest counts where average age class effects are controlled for age dependency of crime intensity. The estimates of parametric part of models showed that sex, race, and past criminal record and its seriousness had the typical, and too often observed, influences on crime counts. Only the number of prior prison incarcerations reduced the number of arrests.

The parametric part of models also included the individual age effect function $f(T_i) = a_0 + a_1 T_i + \alpha_2 T_i^2$. The function tested the bell shaped age-crime profile hypothesis in the presence of age standardized and non-standardized counts, i.e. life time intensities. The age function estimates indicated that the average age-crime profile is still a bell shaped in the sample although we control for average class intensities in semi-parametric model and NLS –approach is based on age standardized arrest counts. The results rejected the Leung (1994) approach where crime intensity increases with age but confirmed some of our model predictions. In general the results support the view the age-crime curve is still alive and well.

APPENDIX I

$$\frac{dE[\tau]}{dS} = (1+\alpha)\mu kE[T]^{-\alpha} t_0 S^{\alpha} - \beta S = 0$$

$\Rightarrow$

$$g(S) = S^{\alpha-1} - B = 0: \quad S^* = \sqrt[\alpha-1]{B},$$

where $\quad B = \dfrac{\beta E[T]^{\alpha}}{(1+\alpha)\mu k t_0} > 0$ and

$$\frac{d^2E[\tau]}{dS^2} = \alpha(1+\alpha)\mu kE[T]^{-\alpha} t_0 S^{\alpha-1} - \beta \text{ and}$$

$$\frac{d^2E[\tau]}{dS^2}\Big|_{S=S*} = \alpha(1+\alpha)\mu kE[T]^{-\alpha} t_0 B - \beta = (\alpha-1)\beta > 0.$$

APPENDIX II

$$AE[T]^{-\alpha} S^{\alpha-1} - \beta = 0, \text{ where } A = (1+\alpha)\mu k t_0 > 0.$$

Differencing totally this first order condition with respect to $E[T]$ and S

$$AE[T]^{-\alpha}(\alpha-1)S^{\alpha-2}dS - AS^{\alpha-1}\alpha E[T]^{-\alpha-1}dE[T] = A_1 dS - A_2 dE[T] = 0,$$

where $A_1 = AE[T]^{-\alpha}(\alpha-1)S^{\alpha-2} > 0$ and $A_2 = AS^{\alpha-1}\alpha E[T]^{-\alpha-1} > 0$

$$\Rightarrow \frac{dE[T]}{dS} = \frac{A_1}{A_2} = \frac{E[T](\alpha-1)}{\alpha S} = \frac{1}{v} - \frac{E[T]}{\alpha S} > 0: \quad \frac{dS}{dE[T]} = \frac{S}{E[T]}\frac{\alpha}{\alpha-1} = v\frac{\alpha}{\alpha-1} > v.$$

Next we analyze the sign of $dE[T]/d\mu$.

$$CE[T]^{-\alpha}\mu - \beta = 0, \text{ where } C = (1+\alpha)kt_0 S^{\alpha\text{-}1} > 0.$$

Differencing totally this first order condition with respect to $E[T]$ and μ

$$CE[T]^{-\alpha}d\mu - C\mu\alpha E[T]^{-\alpha-1}dE[T] = C_1 d\mu - C_2 dE[T] = 0,$$

where $C_1 = CE[T]^{-\alpha} > 0$ and $C_2 = C\mu\alpha E[T]^{-\alpha-1} > 0$

$$\Rightarrow \quad \frac{dE[T]}{d\mu} = \frac{C_1}{C_2} > 0.$$

Results for $dE[T]/dk > 0$ and $dE[T]/dt_0 > 0$ are obtained in similar way.

APPENDIX III

The discounted expected life time, when $t^* = T^* - t$ is the planned remaining life time, is

$$e^{-r(t^*)}E[T \mid t < \infty] = e^{-r(t^*)}\int_0^t e^{-\phi x}dx = e^{-r(t^*)}[\frac{1}{\phi}[1 - \frac{1}{e^{\phi t}}]].$$

Differencing this with respect to t we obtain

$$d[e^{-r(T^*-t)}[\frac{1}{\phi}(1-e^{-\phi t})]] = \frac{1}{\phi}[r'(t)e^{-r(T^*-t)}(1-e^{-\phi t}) + e^{-r(T^*-t)}\phi e^{-\phi t}]dt$$

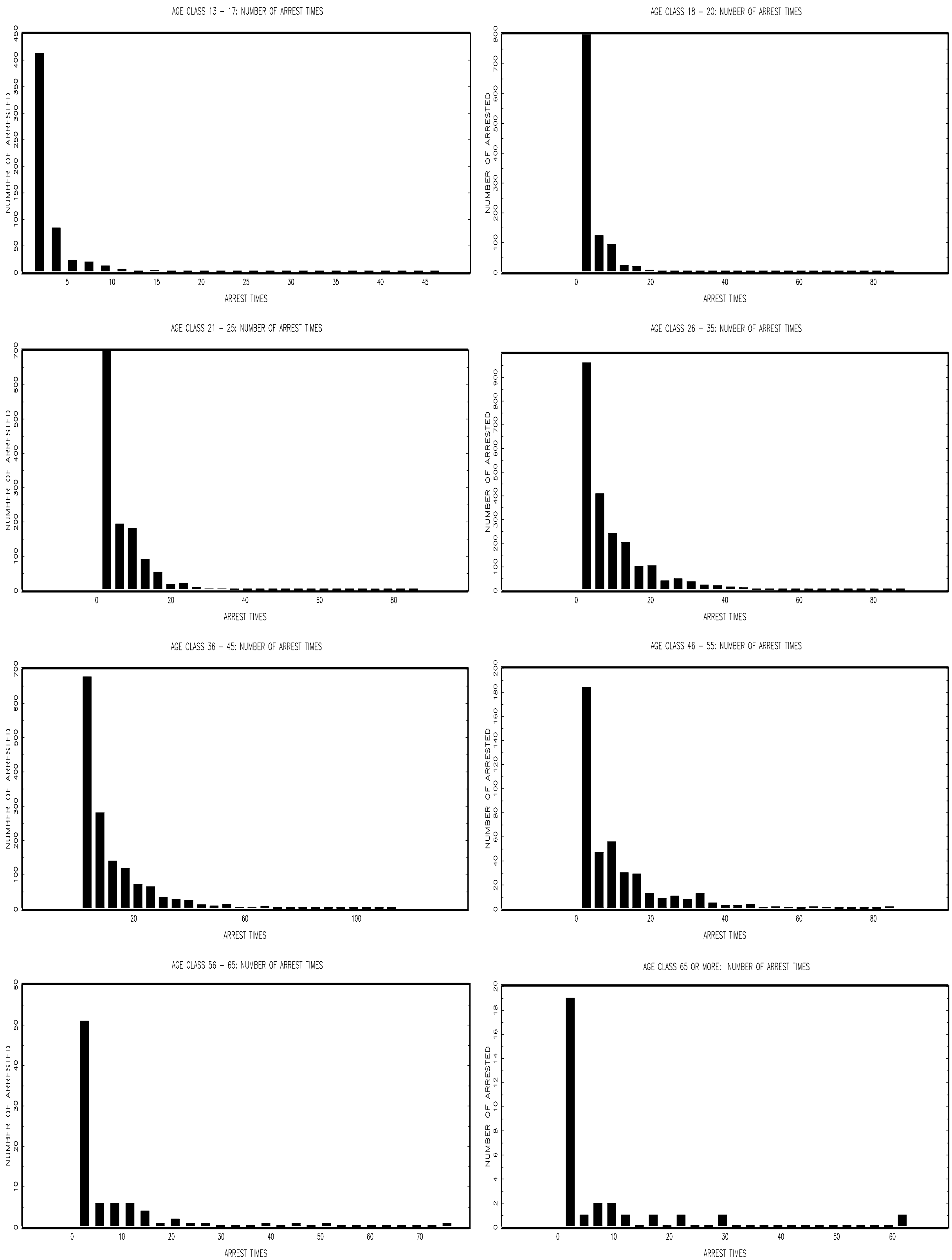
$$= \frac{1}{\phi}e^{-r(T^*-t)}[\phi e^{-\phi t} + r'(t)(1-e^{-\phi t})]dt.$$

The function $g(t) = \phi e^{-\phi t} + r'(t)(1-e^{-\phi t})$ alters its sign from positive to negative with a value $t \in (0, T^*)$, when $T^* \to \infty$ and $r'(t) < 0$.

For example with $r(t) = at - \frac{b}{2}t^2$ we obtain $g(t) = [\phi - a + bt]e^{-\phi t} - bt$ where $g(0) = \phi - a > 0$ and $g(\infty) = -\infty$.

Now, if $\phi = 0.5$, $a = 0.05$, and $b = 0.005$, then $g(t = 5.7) \approx 0$.

APPENDIX IV Number of Arrest in Different Age Classes



APPENDIX V

INTENSITY OF CRIME IN DIFFERENT AGE CLASSES

Descriptive Statistics for **PRIARRG = arrests/age**
Categorized by values of AGE
Included observations: 6827

AGE	Mean	Std. Dev.	Obs.
13	0.000000	NA	1
14	0.047619	0.082479	3
15	0.024242	0.080403	11
16	0.045673	0.102405	26
17	0.072591	0.184902	141
18	0.104238	0.227648	388
19	0.147569	0.197732	367
20	0.198899	0.351168	318
21	0.179539	0.236061	283
22	0.229521	0.284370	283
23	0.229865	0.296120	244
24	0.257037	0.362554	225
25	0.240637	0.279519	251
26	0.290617	0.344655	232
27	0.299121	0.400833	236
28	0.252723	0.335130	223
29	0.222257	0.280762	220
30	0.304538	0.396268	213
31	0.258065	0.285191	211
32	0.292271	0.326247	224
33	0.268328	0.322297	248
34	0.287701	0.310741	220
35	0.263172	0.367075	218
36	0.295930	0.409346	202
37	0.277443	0.335962	211
38	0.260249	0.342393	190
39	0.251792	0.332567	161
40	0.257292	0.296251	168
41	0.246377	0.294197	138
42	0.288132	0.342016	128
43	0.196549	0.274324	124
44	0.291924	0.421032	103
45	0.247737	0.329405	81
46	0.216498	0.288776	73
47	0.256714	0.269882	61
48	0.178066	0.212285	53
49	0.182982	0.226590	59
50	0.250980	0.362642	51

51	0.153251	0.227062	38
52	0.187965	0.272766	31
53	0.152740	0.262590	21
54	0.125220	0.155589	21
55	0.234091	0.332475	16
56	0.199176	0.212398	13
57	0.008772	0.017544	4
58	0.064655	0.105380	8
59	0.096852	0.245385	14
60	0.018333	0.041907	10
61	0.081967	0.117359	9
62	0.232719	0.443461	7
63	0.142857	0.127644	7
64	0.113281	0.138217	4
65	0.125641	0.235021	6
66	0.136364	0.064282	2
67	0.007463	0.010554	2
68	0.000000	NA	1
69	0.032609	0.055974	4
70	0.069048	0.162234	6
71	0.018779	0.021514	3
72	0.027778	NA	1
74	0.114865	0.162443	2
75	0.060000	0.065997	2
76	0.289474	NA	1
77	0.805195	NA	1
78	0.089744	NA	1
80	0.000000	NA	1
81	0.000000	0.000000	2
All	0.229880	0.315579	6827

APPENDIX VI

Dependent Variable: PRIARR,   N = 6827
Method: ML/QML - **Poisson Coun**t (Quadratic hill climbing)
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-3.206763	0.169363	-18.93422	0.0000
AGE	0.052705	0.008837	5.964434	0.0000
AGE2	-0.000552	0.000126	-4.391824	0.0000
PRICONV	0.063757	0.007224	8.825754	0.0000
PRIJAIL	0.001482	0.007544	0.196398	0.8443
PRIPRIS	-0.000899	0.012658	-0.071042	0.9434
RACE	0.050533	0.021287	2.373937	0.0176
SERARR	1.366285	0.035650	38.32450	0.0000
SERCONV	0.190435	0.024506	7.770815	0.0000
SEX	-0.207850	0.033163	-6.267530	0.0000
STATE	-0.008206	0.001856	-4.420355	0.0000
Log likelihood	-20595.94			
Restr. log likelihood	-48103.79			

Method: ML - **Negative Binomial Count** (Quadratic hill climbing)
QML (Huber/White) standard errors & covariance

	Coefficient	Std. Error	z-Statistic	Prob.
C	-3.496442	0.131476	-26.59382	0.0000
AGE	0.050857	0.006418	7.923781	0.0000
AGE2	-0.000561	8.85E-05	-6.337929	0.0000
PRICONV	0.107470	0.005979	17.97500	0.0000
PRIJAIL	0.010028	0.007565	1.325617	0.1850
PRIPRIS	-0.041233	0.008709	-4.734323	0.0000
RACE	0.026668	0.019696	1.353952	0.1758
SERARR	1.485725	0.032041	46.36942	0.0000
SERCONV	0.099808	0.024841	4.017896	0.0001
SEX	-0.222625	0.032523	-6.845219	0.0000
STATE	-0.008750	0.001758	-4.978525	0.0000
	Mixture Parameter			
$1+\phi$	1 + 0.356	0.315	11.520	0.0000
Log likelihood	-15170.91			
Restr. log likelihood	-48103.79			

References

Adda, J. & Cooper, R. (2003) *Dynamic Economics*, MIT Press.

Berkovic, J. & Stern, S. (1991) "Job Exit Behaviour of Older Men", *Econometrica* 59, 189-210.

BJS (Bureau of Justice Statistics) (1998) "Felony Defendants in Large Urban Counties USA, 1998", *State Court Processing Statistics*. U.S. Dept. of Justice.

Blumstein, A. (2005) "An Overview of the Symposium and Some Next Steps", *Annals of the American Academy of Political and Social Science* 602, 242-258.

Blumstein, A. & Cohen, J. (1979) "Estimation of Individual Crime Rates from Arrested Records", *Journal of Criminal Law and Criminology* 70, 561-585.

Campbell, A. (1995) "A Few Good Men - Evolutionary Psychology and Female Adolescent Aggression", *Ethology and Sociobiology* 16, 99-123.

Davis, M.L. (1988) "Time and Punishment: An Intertemporal Model of Crime", *Journal of Political Economy* 96, 383-390.

D'Unger, A.V., Land, K.C. & McCall, P.L. (2002) "Sex Differences in Age Patters of Delinquent/Criminal Careers: Results from Poisson Latent Class Analyses of Philadelphia Cohort Study ", *Journal of Quantitative Criminology* 18, 349-375.

Fagan, A.A. & Western, J. (2005) " Escalation and Deceleration of Offending Behaviours from Adolescence to Early Adulthood", *Australian and New Zealand Journal of Criminalogy* 38, 59-76.

Gottfredson, M. & Hirschi, T. (1990) *A General Theory of Crime*, Stanford UP.

____________________ (1983) "Age and the Explanation of Crime" *American Journal of Sociology* 96, 552-584.

Heineke, J.M. (1978) "Economic Models of Criminal Behaviour: An Overview" in *Economic Models of Criminal Behaviour*, J.M. Heineke (ed.), North-Holland.

Hopenhayen, H. & Rogerson, R. (1993) "Job Turnover and Policy Evaluation: A General Equilibrium Analysis", *Journal of Political Economy* 101, 915-38.

Jacob, B., Lefgren, L. & Moretti, E. (2004) "The Dynamics of Criminal Behaviour: Evidence from Weather Shocks", *NBER*-WP 10739.

Kanazawa, S. (2002) "Why Productivity Fades with Age: The Crime-Genius Connection", *Journal of Research in Personality* 37, 257-272.

Kyvsgaard, B. (2003) *The Criminal Career. The Danish Longitudinal Study*. Cambridge University Press.

Land, K.C. (1992) "Models of Criminal Careers: Some Suggestions for Moving Beyond the Current Debate", *Criminology* 30, 149-155.

Lawless, J.F. (1987) "Regression Methods for Poisson Process Data", *JASA* 82, 808-815.

Lee, D.S. & McCrary, J. (2005) "Crime, Punishment, and Myopia", *NBER*-WP 11491.

Leung, S. F. (1995) "Dynamic Deterrence Theory", *Economica* 62, 65-87.

__________ (1994) "An Economic Analysis of the Age-Crime Profile", *Journal of Economic Dynamics and Control* 18, 481-497.

__________ (1991) "How to Make the Fine Fit the Corporate Crime? An Analysis of Static and Dynamic Punishment Theories", *Journal of Public Economy* 45, 243-256.

Levitt, S.T. (1999) "The Limited Role of Changing Age Structure in Explaining Aggregate Crime Rates", *Criminology* 37, 581-605.

Nagin, D.S. & Land, K.C. (1993) "Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric, Mixed Poisson Model", *Criminology* 31, 327-362.

Polinsky, A.M. & Rubinfield, R. (1991) "A Model of Optimal Fines for Repeat Offenders", *Journal of Public Economy* 46, 291-306.

Rust, J. (1996) "Numerical Dynamic Programming in Economics" in *Handbook of Computational Economics* Vol. I, 619-729. North-Holland.