3

The Theory and a Simple Model

Orientation

In Chapter 2 we have seen that an analysis of a large number of detailed criminal careers of those who have at some time been convicted of a serious (standard list) offence indicates the existence of two categories of offender with constant but different reconviction probabilities. The same kind of analysis looking at the timing of offences indicates two categories of offender with constant but different rates of conviction. These categories were revealed by plotting graphs. The proportion of the population in each of these categories is essentially constant across birth cohorts. In this chapter we will describe our theory which explains these results and we will construct a three category model to predict independent criminal history data.

Introduction

Having looked at the Offenders Index (OI) we will now construct a theory which will explain the observed regularities. The theory will be ‘large scale’, in that it does not consider the psychological, social and economic causes of general offending behaviour and also in that there are almost certainly other special groups and subgroups of offenders beyond those we have discovered from our aggregate level examination of the OI.

Commonly, it has been believed that sex offenders would form one such group, generally characterized by a much greater degree of specialization than more typical offenders, although empirical evidence suggests this may be an over-simplification (Zimring, Jennings, Piquero, and Hays 2009).

A more definitive group consists of around half of life sentence prisoners who have very low probabilities of recidivism. There is also likely to be some variation in the parameters for the recidivism
probability and the rate of conviction for subsets of offenders within each category, as we saw with male and female subsets in Chapter 2. However this variation is much less than the differences in parameters between the categories.

The utility of our theory might be questioned because it does not consider individual motivations leading to crime and therefore gives us little idea how to intervene. We take the opposite view. It seems to us that without a large scale theory one cannot begin to understand the features of general criminality which need in turn to be explained by psychological, sociological, and economic criminological theories. We would have no idea of the basic parameters we were trying to measure or how they were interrelated in actual measurements. For example the often used two-year reconviction ‘rate’ (actually a probability) depends on both reconviction probabilities and offending rates. Also, different descriptions may be appropriate for differently defined categories. We shall see in Chapter 6 that in one of our categorizations offenders can be divided into those who are unusually impulsive and those who seem to have otherwise quite normal psychological features. This in turn implies the need for different interventions to reduce offending.

Another objection to our ‘large scale’ approach is that it will ignore a great deal of important detail, and we agree. However, without understanding the large scale framework one will never understand the small scale detail. A similar objection may also come from those who would argue that we are ignoring ‘statistically significant’ information and making too much use of our own judgement of what is important. We would respond by pointing out that no analysis is judgement-free and in what follows we will make our judgements explicit rather than hiding them in the underlying assumptions of particular statistical techniques.

The paradigm that we believe is most useful for developing an understanding of criminal careers is similar to that of the historical understanding of planetary motion (which is indeed the paradigm of most successful scientific research). By the time of the Renaissance it had become clear that the Ptolemaic geocentric description of planetary motion, although still empirically very effective, was philosophically unacceptable. The rival Copernican system, though superior philosophically, in its simplest form (as championed famously by Galileo) simply did not work. To make his system work, Copernicus had shown that it was necessary to build in so many epicycles that, on grounds of simplicity, it was considerably
less acceptable than the Ptolemaic approach. The answer was found by Kepler, who realized that Galileo’s simple picture of planets orbiting the sun could be made to work as accurately as Ptolemy’s by replacing the perfect circles with ellipses. In turn the attempt to explain Kepler’s ellipses led to Newton’s idea of gravity. This in turn predicted that the gravitational interaction of the planets would make them follow non-elliptical orbits, leading to the discovery of Uranus, Neptune, and Pluto. The discrepancy between the actual orbit of Mercury and that predicted by Newtonian theory led to the discovery by Einstein of General Relativity which describes the motion of the Universe. At an even smaller scale we know that the orbits of the planets are in fact chaotic and not predictable at all.

We thus have a hierarchy of descriptions, beginning with circular orbits which are still suitable for qualitative description (Galileo). These lead to slightly elliptical orbits (Kepler). In turn these lead to perturbed elliptical orbits and finally to chaotic orbits. In each case the larger scale description provides the arena within which the smaller scale structure can be identified and then understood.

Our theory is a large scale description of the process of offending, capture, conviction and eventually desistance. We fully acknowledge that it ignores many important features of offenders, offending, motivation, and criminal justice system responses. But we hope that the theory and the models can provide a framework within which these features and their associated mechanisms can be understood.

The Assumptions of our Theory

In Chapter 2 we conducted a detailed analysis of the recidivism characteristics of the (1953, 1958, 1963, 1968, and 1973) cohort samples and the 1997 sentencing sample from the Offenders Index. The statistical models fitted to the distributions of both conviction-number and inter-conviction times resulted in very high values of the correlation coefficients and were highly suggestive that the first four assumptions of our theory are correct.

The basic assumptions of the theory are:

1. The population at large can be categorized into one non-criminal and a small number of criminal categories. The criminal categories consist of individuals who will commit relatively serious criminal offences and be convicted of one or more of these offences within their lifetimes.
2. Criminality is constant: the proportion of the population in the criminal categories is approximately constant\(^1\) across cohorts.

   Within each of the criminal categories:

3. Recidivism is constant: immediately after each conviction the probability of an individual being reconvicted of one or more further offences at some time in the future is constant.

4. Rate of offending is constant: whilst active, the probability of an offender committing an offence in a given time period is constant whether that time period is now or at some arbitrary point in the future.

For individuals in the largest ‘non-criminal’ category we assume that the rate of (relatively serious) offending is very close to zero and, as they are never convicted for standard list offences, their recidivism probability is zero. This raises the possibility that there are individuals in the non-criminal category who do commit relatively serious offences; however, we believe that their numbers and the proportion of serious crime for which they are responsible is small.

These assumptions ensure that our theory will reproduce the results of Chapter 2 on recidivist behaviour as observed in the Offenders Index. It is easy to show however that alone they do not provide any hope of explaining the most well known result on offending, namely the ‘age–crime’ curve, or more accurately, in the context of the Offenders Index analysis, the ‘age–conviction’ curve.

### Explaining the Age–Crime Curve

The age–crime curve is a histogram of the numbers of convictions at each age. The curve can be constructed using data from either a cohort or a cross-sectional sample. Figure 3.1 shows the age–crime curve for the 1997 sentencing sample.\(^2\) The graph shows the count of court appearances, resulting in one or more convictions during

---

\(^1\) There will be some variation in measured criminality over time due to changes in prosecution policy concerning cautions/warnings etc and offence classifications.

\(^2\) Sentencing samples were drawn from the Offenders Index at regular intervals to investigate the impact of sentencing policy on reconvictions. The 1997 sentencing sample consists of the complete (OI) criminal histories of all offenders convicted of one or more offences at court appearances during the first weeks of alternate months from February through December 1997, including all reconvictions up to 01/01/2002.
the sample weeks, for individuals at the age indicated on the x axis, in age increments of three months. The data has been age-weighted to standardize the graph to a constant number of individuals at each age in the community.

The graph starts at age 10, the age of criminal responsibility, which is the first age at which an individual can be convicted of a criminal offence in England and Wales. As age increases the count of convictions increases until typically 17–18 years of age, and after this there is a slow decline. The small secondary peak at age 25 reflects individuals of unknown age who are coded by the courts as age 25 (with date of birth coded 01/01/1972 in the 1997 sample). This secondary peak disappears completely if offenders with a recorded date of birth of 01/01/1972 are excluded from the sample.

The assumptions we have written down so far imply that offenders appear at the age of criminal responsibility, offend and are then convicted, at which stage some drop out. The remainder then go on to be convicted again after which some drop out and so on. This would lead to an age–crime curve similar to Figure 3.2.

We can see that the fall off with age above age 20 is reproduced but not the rise until age 17–18. However, Figure 3.2 is consistent with other graphs of antisocial behaviour against age. For example we may consider the results of Nagin and Tremblay (1999) who

---

**Figure 3.1 Age–crime curve**

Source: 1997 sentencing sample, Offenders Index.

Note: Each data point represents the number of offenders in the sample convicted at the age shown on the x axis in three-month increments of age.
measured the antisocial behaviour of over 1,000 boys in Montreal from age six to age 15. Three types of antisocial behaviour were considered: physical aggression, opposition, and hyperactivity. Four trajectories of externalizing behaviour problems with age were identified for each of the types of behaviour. Figure 3.3 shows the fitted trajectories for physically aggressive antisocial behaviour at age six and then annually from age 10 to 15. Very similar trajectories were found for the other two antisocial behaviours; but the

Figure 3.2 Hypothetical age–crime curve from Assumptions (1) to (4).

Figure 3.3 Trajectories of physical aggression, Montreal longitudinal experimental study of boys.

Source: Nagin & Tremblay (1999).
groups of individuals following similar trajectories for different behaviours, although overlapping, were by no means coincident. What is important however is that none of the trajectories show antisocial behaviour increasing up to age 15. They actually show such behaviour staying constant or decreasing. There are also no late onset groups identified for any of the behaviours among the study boys (p 1189). Nagin and Tremblay (1999, p 1192) also remark that childhood physical aggression is a distinct predictor of violence and serious delinquency in adolescence and that these findings (including the non-increasing rates) are replicated in five other longitudinal data sets from around the world.

What then is the cause of the disparity between other measures of antisocial behaviour and convictions in England and Wales? There are two simple explanations. The first is society’s attitude to certain behaviours at different ages: from a legal standpoint, in England and Wales, the age of criminal responsibility precludes formal criminal conviction of children under the age of 10 (at the time of the offence). At ages just above 10, if one child hits another in the school playground or is caught stealing, these events are unlikely to result in any more than a telling-off. If juveniles are caught shoplifting, damaging property, or fighting, this will probably be dealt with within the school, by parents or informally by the police. Even if the incident is serious, there will be a high probability of the use of a formal reprimand, final warning, or formal caution by the police for younger offenders. However, if one young adult hits another young adult in a public place or steals a car, this is much more likely to lead to prosecution and conviction in the courts. The second explanation is the individual’s capacity to commit criminal acts. For example, Farrington (1997) reminds us that children must have reached a certain size before they can reach the controls of a car, and are thus incapable of stealing one. Physical aggression by children is unlikely to result in serious injury unless weapons are involved and children are unlikely to be able to purchase goods in shops with forged cheques or credit cards.

These explanations combine to give our fifth assumption:

5. The probability that similar criminal behaviours will result in conviction increases with age from age 10 to age 17.

Unless the details of informal sanctions are well understood, there are serious empirical difficulties in measuring criminality. Most studies using official records of arrest or conviction report
male participation rates between 20 per cent and 40 per cent; however, Farrington (2002) found that, although 40 per cent of his Cambridge Study male cohort had official convictions up to age 40, 96 per cent admitted to committing at least one equivalent criminal act up to age 32 in self-reports. The criminal categories defined in assumption (1) do not include any individuals who cease to offend following informal sanctions, police reprimands, warnings or cautions. Criminality as defined in assumption 2 may therefore vary between cohorts because of changes in prosecution policy.

We now need just one further assumption to enable us to model the age–crime curve:

6. The probability of capture and conviction for similar offences increases after the offender is known to the police.

Assumption 6 is necessary to resolve an inconsistency between the survival time distributions of time to first conviction and inter-conviction time. The statistical details are explained below, but the assumption is intuitively plausible \textit{a priori}.

The inconsistency becomes apparent if we construct a survival curve for time to first conviction for the 1953 cohort and compare the slope of the straight line section of the curve, Figure 3.4, with that of the reconviction survival time curve of Figure 2.7.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure3.4.png}
\caption{Survival time to first conviction}
\end{figure}

\textit{Source: 1953 cohort, Offenders Index.}
\textit{Note: The y axis is on a logarithmic scale. The solid curve on the graph shows the number of offenders in the cohort sample who remain conviction free up to the age shown on the x axis. The data is plotted at monthly intervals. The overlaid dotted straight line is the exponential fit to the curve between ages 18 and 40.}
The equation to the straight line in Figure 3.4 is:

$$y = A \cdot e^{-0.115t}$$  \hspace{1cm} (3.1)

Equation 3.1 gives a mean time to first conviction, for unconvicted offenders over the age of 18, of 8 years and 9 months.\(^3\) This can be compared with a mean time to reconviction of 4 years and 9 months derived from Equation 2.6. As we show in the Appendix, a random sample from a stream of random events, with a mean inter-event time \(T\), results in a stream of random events with a mean inter-event time \(T/p\) where \(p\) is the sampling probability. We can apply this result to the inconsistency between first and subsequent convictions by letting \(p\) equal the ratio of the mean time to the next conviction to the mean time to the first conviction. As a first approximation this gives the relative\(^4\) probability, \(p\), of a first conviction compared with subsequent convictions of 0.55.

A second inconsistency between the survival to first conviction and the survival to reconviction curves is the slope of the curves prior to the straight line sections; age less than 17 in Figure 3.4 and reconviction time less than five years in Figure 2.7. In the former, as a direct result of assumption 5, which has the effect of slowing down the rate of first conviction, the initial slope is less than that of the straight line section. In the latter, due to the rapid reconviction of the high-rate offenders, the initial slope is greater. In the survival (age) to first conviction curve the effect of the high-rate offender category is completely masked by the effect of assumption 5 and the preponderance of low-rate offenders in the criminal categories. At least 85 per cent of offenders are low-rate (see Table 2.2).

Both first conviction and reconviction curves exhibit the rounding down in the tail of the curve due to censorship of the data at age 46. This rounding down can be successfully modelled by the subtraction of a constant from the right hand side of the survival

\(^3\) For the exponential survival time distribution the mean time to failure applies from any time during the process. For example at age 18 the expected (mean) time to first conviction is 8 years 9 months \(\left(\frac{1}{0.115}\right)\) years. For all those surviving to age 25, or indeed to any other age greater than 18, the expected time to first conviction is also 8 years and 9 months.

\(^4\) If the probability of reconviction given a crime is \(q\) then the probability of first conviction given a crime is \(p*q\); ie \(p\) is the relative, or effective sampling, probability.
equations, or alternatively by adding the same constant to all of the
data points. This constant represents the number of offenders who
will be convicted for the first time after the age of 46. We are now
almost in a position to convert our descriptive statement of the
assumptions into empirically measurable parameters. But first we
need to expand on assumption 5.

The Rise in Crime from 10 to 17 Years of Age

There is very little empirical evidence relating society’s response to
similar behaviours at different offender ages except that children
under the age of 10 in England and Wales are deemed not crimi-
nally responsible. Intuitively we would not expect society’s response
to be very different at 10 years and one month compared with
9 years and 11 months. We would also expect the transition from
all acts being non-criminal to individuals being fully responsible, as
in assumption 5, to be smooth, ie a small increment in age would
not result in a large change in response. As an indicator of these
changing responses we can look at police use of reprimands, final
warnings, and cautions. Each of these police disposals is recorded
on the Police National Computer (since May 1995) but are not
criminal convictions. Reprimands and final warnings are given to
those under 18 and cautions to those over 18; we will refer to all
these disposals as cautions. Each of these disposals requires that
there is evidence linking the offender to the offence and that the
offender admits his/her guilt. These informal disposals may also
involve reparative, rehabilitative and/or punitive elements. Figure
3.5(a, b, c, and d) shows the use police made of these cautions for
offenders aged 10 to 20 in the second quarter of 2004, in age bands
of three months.

Figure 3.5a shows the number convicted and the number cau-
tioned on their first recorded police contact and Figure 3.5c shows
the number convicted and cautioned on their second or subsequent
contacts. Figures 3.5b and d show the proportions convicted for
first and subsequent contacts respectively.

It can be seen that on their first police contact the overwhelming
majority of offenders are dealt with outside the court system. Prior
to age 14, less than 6.25 per cent of offenders are charged and con-
icted, although the number of offenders steadily increases, from
around 125 per three-month age increment at age 11 and under,
to a peak of over 1,100 at age 16. The proportion convicted at
The Rise in Crime from 10 to 17 Years of Age

age 16 exceeds 10 per cent for the first time and continues to increase to 40 per cent at age 20, as shown in Figure 3.5b. The pattern is entirely different for second and subsequent police contacts. The total number of offenders with more than one recorded police contact at or below age 12 is less than 360, and the proportion convicted is 55 per cent. This proportion steadily increases to 87 per cent at age 20 (see Figure 3.5d). The increase in the proportion of offenders convicted as age increases provides some support for assumption 5 and the differences in disposals between first and subsequent police contacts provide support for assumption 6.

Figure 3.5a Recorded outcome of first police contact.
Source: PNC April 2004 sample.
Note: The y axis shows the number of offenders cautioned or convicted at the age shown on the x axis in age increments of three months.

Figure 3.5b Proportion convicted on first police contact
Source: PNC April 2004 sample.
The evidence in Figure 3.5 (a–d) does not provide the complete picture concerning changes in society’s response to criminal behaviour as age increases, only the official response after all informal actions have been exhausted. In order to model the early part of the age–crime curve we need a function which reflects society’s view, both formal and informal, of what is or is not criminal as age increases. At age 10 the probability of conviction, given a criminal act, should be close to zero. It should then increase, at first slowly but accelerating and increasing most rapidly in early to mid-teens.

**Figure 3.5c** Recorded outcome of second and subsequent police contacts

Source: PNC April 2004 sample.

Note: The large fluctuations between ages 10 and 12 are due to very small numbers of offenders with more than one police contact at these ages.

**Figure 3.5d** Proportion convicted on second and subsequent police contacts

Source: PNC April 2004 sample.

Note: The y axis shows the number of offenders cautioned or convicted at the age shown on the x axis in age increments of three months.
then slowing and levelling off at a probability of one in late teenage. Such a function is given in Equation 3.2:

\[ P(\text{convicted} \mid \text{age} - t) = 1 - \frac{1}{1 + e^{\alpha \ast (t - c)}} \] (3.2)

Where:
- \( \alpha \) Controls the slope of the transition (small values of \( \alpha \) give a shallow slope and large values >1 give increasingly steep transitions)
- \( c \) is the age at which the probability \( P \) is \( \frac{1}{2} \) (as \( e^0=1 \)).

The function is arbitrary but provides a plausible shape, which is theoretically defensible, with flexibility in the parameters to enable the initial portion of the age at first conviction curve to be modelled in a mathematically tractable way.

There are of course other functions which can be used to model this phase of the age–crime curve. Farrington (1986, pp 240–243) explores several distributions and functions describing the association between age and crime, the most successful of which used a term of the form \( \alpha \ast x^k \) to approximate the rise in crime in the early teens which is then counteracted by a negative exponential, \( e^{-c \ast x} \), which becomes dominant beyond the peak age of offending. This was offered as an empirical fit to the curve but, as Farrington points out, the Gamma distribution function is of this general form and, as we shall see in Chapter 4, the gamma distribution can be applied as a theoretically defensible approximation.

### Modelling the Age–Crime Curve

Returning to the survival time to first conviction graph of Figure 3.4, the straight line section, between ages 18 and 42, is characteristic of a proportional hazard survival process in which the number failing (that is being convicted) at a given age is a constant proportion of the number surviving to that age. This is consistent with the high- and low-rate reconviction survival processes described in Chapter 2 and with assumption 6 which asserts that the underlying survival processes are the same but with the initial failure rate reduced by the relative probability\(^5\) of a first conviction. In addition, over the

\(^5\) We use the term relative probability because we assume: that the rate of criminal behaviour is the same while offenders are active, both before and after the first conviction; and that the probability of conviction given a crime is lower for first
early part of the curve, ages 10 to 20, the failure rate is also multiplied by the right hand side of Equation 3.2.

We can describe these processes mathematically as follows. Within each of the homogeneous rate categories, the ‘survival to first conviction’ process can be described by the solution to the following differential equation:

\[
\frac{d}{dt} S_f(t) = -P_f \times \lambda \times S_f(t) \times \left(1 - \frac{1}{1 + e^{\alpha(t - c)}}\right) \tag{3.3}
\]

Giving the survival function:

\[
S_f(t) = C \times \left(1 + e^{\alpha(t - c)}\right)^{-\frac{P_f \times \lambda}{\alpha}} \tag{3.4}
\]

Where:
- \(S_f(t)\) is the number of offenders surviving (without conviction) to age \(t\),
- \(P_f\) is the relative probability of first conviction,
- \(\lambda\) is the rate parameter for reconvictions,
- \(C\) is the number of offenders in the rate category who will be convicted in their lifetime.

In order to model and fit the survival curve of Figure 3.4 we need to sum the survival functions (the right hand side of Equation 3.4) for both high- and low-rate offenders with parameter values as estimated for the 1953 cohort in Chapter 2 and listed in Table 3.1.

In Table 3.1 the numbers of offenders has been increased by 33 for the high-rate and 113 for the low-rate categories to adjust the initial value at age 9. In addition, to compensate for censorship, 310 has been added to the low-rate offender total to represent offenders who will be convicted for the first time after age 46, the limit of the observation period. The remaining parameters, \(P_f\), \(\alpha\) and \(c\), were estimated by fitting the combined survival function to the survival time data of Figure 3.4, with 310 added to each of the convictions than for subsequent convictions. The alternative assumption is that the rate of criminal behaviour increases as a direct result of the first conviction.

\(^6\) In Figure 3.6 (a and b) it can be seen that there is a data point at age 9. When the Offenders Index was first created in 1963 the age of criminal responsibility was 8 years but was increased to 10 in the 1963 Children and Young Persons Act. The revised age of criminal responsibility (which took effect on 01/02/64) was not effective during 1963 with the result that 99 individuals in the 1953 cohort received convictions before their tenth birthday.
data points. A least squares iterative fitting procedure was used, and over 99.9 per cent of the variance in the data was accounted for by the model. The parameter estimates were: $P_f = 0.51$, $\alpha = 0.56$ and $c = 14.45$. Figures 3.6a and 3.6b show the fitted curve and data points with linear and logarithmic (to base 10) $y$ axis scales respectively.

Equation 3.4 can be differentiated to give an expression for the age at first conviction curve for each of the rate categories.

$$y(t) = C \cdot \left(1 + e^{a \cdot (t - c)}\right)^{\frac{t - c}{\lambda}}$$

Figure 3.7 shows the onset age–crime curve derived from the above survival analysis based on the sum of the high and low-rate versions of Equation 3.5. All the parameter values are the same as those for the survival analysis which generated Figures 3.6a and b.

The dotted lines above and below the fitted curve in Figure 3.7 are the ±2 standard deviation (95 per cent) confidence limits assuming a Poisson distribution of convictions in each three-month interval of age. A feature of Figure 3.7 that requires some explanation is the group of convictions between age 15 and a half and 17 which fall below the $-2\sigma$ line and a second group between the ages of

![Figure 3.6a](image.png)
and 18-and-a-half which fall above the $+2\sigma$ line. In 1968 the Metropolitan Police introduced formal cautioning as an alternative to prosecution, which had the effect of diverting many juveniles away from court and reducing the numbers convicted prior to age 17. However on reaching the age of 17 (the minimum age for adult court at that time), cautioning was much less likely and individuals who offended at age 17 were more likely to be convicted. It may even have been the case that prosecutions were delayed to ensure that adult sentences were imposed. Thus the introduction of

Figure 3.6b Survival to first conviction (logarithmic y axis)

Source: 1953 cohort, Offenders Index.
Note: The data plotted on these graphs is the same as that used in Figure 3.4 above, but plotted at annual increments and with 310 added to each data point.

Figure 3.7 Age at first conviction for the 1953 cohort.

Source: Offenders Index.
Note: The points on the graph show the number of offenders with their first conviction at the age shown on the x axis, in age increments of three months.
cautioning postponed some first convictions from 1968 to 1979 (see also Farrington 1990; Farrington and Bennett 1981; Farrington and Maughan 1999). These two groups of outliers counterbalance each other making a suppressed-demand kind of explanation plausible and suggests that cautioning was less successful than hoped for. A similar, but much smaller, fluctuation in conviction numbers is apparent in the 1958 (see Figure 3.8), 1963 and 1968 cohorts.

Repeating the survival to first conviction analysis on the 1958 cohort data yields parameter values of: $P_f = 0.72$, $\alpha = 0.64$ and $c = 15.01$. Here the transition slope $\alpha$ is greater (steeper) than the 1953 cohort value and the middle age of the transition is six months later. The high- and low-rate parameters used were those estimated for the 1953 cohort. The fit to the age at first conviction curve for the 1958 cohort is shown in Figure 3.8. The suppressed demand effect around age 17 is in evidence in the data, but only the third quarter of age 16 falls outside the $\pm 2\sigma$ bounds. A suppressed demand effect could also account for the small number of above $2\sigma$ outliers in the age range 10 to 11 as the most troublesome children become eligible for prosecution. The same basic structure is seen in the remaining cohort samples but the parameter values for $P_f$, $\alpha$ and $c$ vary between cohorts. In particular the slope $\alpha$ and the mid-transition age $c$ both increase as the cohorts become more recent. These parameter changes reflect the increasing use over time of police cautions for young offenders.

![Figure 3.8 Age at first conviction for the 1958 cohort](image-url)

Source: Offenders Index.
To illustrate the consistency of the basic structure of the data over time, the survival function was also fitted to data from a sub-sample of the 1997 sentencing sample. The sub-sample was created by selecting target conviction records that were first convictions. Also, individuals with a ‘coded’ date of birth of 01/01/1972 were omitted from the sample, completely removing the spurious peak at age 25 that was observed in Figure 3.1. The rate parameters, $\lambda_1$ and $\lambda_2$, estimated for the 1953 cohort were used for the fit and the proportion of high-rate offenders was estimated from the sentencing sample. $P_f$, $\alpha$ and $c$ were estimated using a least squares procedure as before. The derived age at first conviction profile is plotted in Figure 3.9.

The fit to the sentencing sample data includes first convictions up to age 70 and overall just 13 data points lie outside the $\pm 2\sigma$ confidence limits, whereas statistical theory predicts 12. Virtually all variation in the data, in all three data sets, is explained by the statistical model. However, it should be stressed that at the individual offender level there will, no doubt, be causal explanatory factors which influence their offending behaviour. However, these individual explanations aggregate in such a way as to be consistent with our large scale theory.

The parameter estimates used to fit the age at first conviction curves for the 1953 and 1958 cohorts and the 1997 sentencing sample data are presented in Table 3.2. The relative probability of

![Figure 3.9 Age profile for first convictions during 1997](Image)

Source: 1997 sentencing sample, Offenders Index.
a first conviction is subject to greater variation than the other parameters, although changes in cautioning policy for older offenders could explain this variation. In addition, the ‘probability of conviction given age’ parameter values, $\alpha$ and $\gamma$, display a progressively increasing trend as the data becomes more recent. This trend is consistent with known policy changes; the diversion of young offenders away from formal conviction has resulted in a progressive and significant increase in the use of cautioning for juveniles. The effect of this has been that peak age of onset has been delayed slightly and the rise to that peak has become much steeper in the more recent data.

The proportions of offenders who are high-rate, estimated independently for each sample, are remarkably consistent at around 16 per cent. Although we have seen significant changes, over time, in the parameter values describing the onset phase of the criminal career, the basic structure of the model has not changed. In particular the parameters describing the ongoing criminal behaviour appear to have remained substantially constant over the 40 years covered by the available data.

The mathematical models derived above specifically describe the numbers of offenders convicted for the first time at each age (the age of onset). The full age–crime curve includes all subsequent convictions, second, third, etc. We now derive a model for the age–crime curve for each subsequent conviction number. Our theory

---

Table 3.2 Parameter values for age at first conviction fitted curves

<table>
<thead>
<tr>
<th>Data source</th>
<th>Number of offenders</th>
<th>Proportion high-rate</th>
<th>$\lambda_h$</th>
<th>$\lambda_l$</th>
<th>$P_f$</th>
<th>$\alpha$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1953 cohort</td>
<td>12,127</td>
<td>0.161</td>
<td>0.86</td>
<td>0.211</td>
<td>0.51</td>
<td>0.54</td>
<td>14.68</td>
</tr>
<tr>
<td>1958 cohort</td>
<td>13,006</td>
<td>0.159</td>
<td>0.86</td>
<td>0.211</td>
<td>0.76</td>
<td>0.61</td>
<td>15.37</td>
</tr>
<tr>
<td>1997 sentencing sample</td>
<td>14,090</td>
<td>0.160</td>
<td>0.86</td>
<td>0.211</td>
<td>0.406</td>
<td>1.00</td>
<td>15.82</td>
</tr>
</tbody>
</table>

Note: $\lambda_h$ and $\lambda_l$ are the high and low-rate parameters from Table 3.1 and $P_f$, $\alpha$ and $c$ are the values obtained by fitting the combined high and low-rate first conviction survival functions (Equation 3.4) to the specified data source.

---

7 The proportion of offenders who are high-rate differs from the estimate given in Chapter 2 for the 1953 cohort because the estimated number of offenders yet to offend has been included in the calculations here.
predicts that, within a risk/rate category, at any given age the number of offenders convicted is proportional to the size of the active offender population at that age. This relationship also holds for the subset of the active offender population at the given age with just \( i \) previous convictions. A proportion \( p \) of this subset will be reconvicted at rate \( \lambda \) moving them into the subset of the offender population aged \( t \) with \( i + 1 \) previous convictions. At the same time some of this latter subset will themselves be convicted at rate \( \lambda \) and leave the subset. This process is described mathematically by Equation 3.6:

\[
\frac{d}{dt} (y_{i+1}(t)) = p \lambda y_i(t) - \lambda y_{i+1}(t) \tag{3.6}
\]

for \( i > 0 \)

Starting with \( i = 1 \) and solving Equation 3.6 for each of the risk/rate categories and successive values of \( i \), gives us the size of the offender population in each category with just \( i \) previous convictions at age \( t \). Substituting back into Equation 3.6 and summing the results over the three risk/rate categories generates the age–conviction curve for each conviction number. This series of equations was solved numerically as no simple analytic solution exists. Over 90 per cent of data points for the number of convictions in three-monthly age increments at each previous conviction count fell within the \( \pm 2\sigma \) confidence limits for the 1953 cohort.

We can generate the age profile for all reconvictions within each risk/rate category by lumping all active offenders together into a single active convicted offender pool. A proportion \( 1 - p \) leave the pool after each conviction (convictions occur at rate \( \lambda \) in direct proportion to the size of the active convicted offender pool \( Y(t) \)) modifying Equation 3.6 as shown in Equation 3.7:

\[
\frac{d}{dt} Y(t) = p \lambda y_i(t) - (1 - p) \lambda Y(t) \tag{3.7}
\]

Where:

- \( p \) is the reconviction probability,
- \( Y(t) \) is the size of the active convicted offender pool.

Equation 3.7 was again solved numerically for each of the risk/rate categories, but with the high-risk probability increased from 0.840 to 0.855 to compensate for censorship at age 46 in the 1953 cohort and to provide a better fit to the data. To justify this change, see the graph in Figure 2.5 which shows an increasing trend in
recidivism probability estimates with the length of follow-up in the cohorts. Ideally the parameter values used in constructing the theoretical age–crime curve should be ‘whole life’ values rather than estimates determined from censored data sets.

The size of the active convicted offender pool for all convictions in each risk/rate category at age $t$ is given by:

$$Y_{at}(t) = y^*_1(t) + Y(t) \quad (3.8)$$

The overall age–crime curve is given by substituting back into Equation 3.7 and summing over the three categories, high-risk/high-rate, high-risk/low-rate and low-risk/low-rate. Figure 3.10 shows the fitted theoretical age–conviction curve with the $\pm 2\sigma$ confidence limits and the 1953 cohort data. With the exception of the outliers around age 17, explained above as a policy-induced period effect, the majority of data points fall within the $\pm 2\sigma$ (95 per cent confidence interval) of the model.

With suitable adjustment for juvenile cautioning policy (see Table 3.2 for the 1958 cohort and 1997 sentencing sample parameters) the model fits the all convictions age-crime data from the remaining cohort samples, 1958, 1963, 1968, and 1973, and the 1997 sentencing sample.

*Assumptions 1 to 6 above and their mathematical representation are thus sufficient to accurately model the age–crime (conviction)*
The Theory and a Simple Model

curve, for first convictions (onset), all subsequent convictions, second, third, fourth, etc, and overall. This is not just an exercise in curve fitting, since the equations follow directly from the assumptions and the parameters have real world interpretations relating to the process of offending and conviction.

As with any theory, it is not sufficient just to explain the observations but the theory must also be capable of making verifiable predictions and the premises of the theory must be credible and acceptable. We have already seen that the structure and parameters of the age-crime model are consistent across cohorts and even cross-sections and in Chapter 7 we will show how the theory can accurately predict the prison population for given sentencing policies. The premise of the theory is that: for individuals with a propensity for crime, criminal acts will occur at random, the convolution of inclination and opportunity. These individuals will continue to (re-)offend until they are caught and convicted at which point a life-choice decision is made either to continue as before or to modify their behaviour to avoid further conflict with the law. In this premise we have assumed that offences are committed at random, according to a Poisson process. In the Appendix we show that a random sample of events from a Poisson process is itself a Poisson process. Our analysis has shown that convictions display the characteristics of a Poisson process which in turn implies that they are a random sample of offences committed at random according to a Poisson process.

The 100,000 Active Prolific Offenders

One result of our theory became quite influential in the first few years of the twenty-first century. Among the risk/rate categories identified in the theory the high-risk/high-rate individuals are likely to have the highest number of convictions, to start their criminal careers earliest, and to commit the most crimes. Because the model separates the categories we can calculate the expected number of convictions in a year for members of this particular category. This is achieved using numerical solutions to Equation 3.6 for each conviction number with the parameters derived for the 1997 sentencing sample (Table 3.2). Summing over all conviction numbers results in some 180,000 convictions in 1997 which can be attributed

to high-risk/high-rate offenders. Not all of these offenders will become persistent, in the sense that they will accrue four or more convictions during their criminal career.

The number of high-risk/high-rate offenders who receive their fourth or higher conviction during a year can be calculated by summing over conviction numbers greater than 3. From the theory we would expect some 18 per cent of the convictions (fourth or higher) to be attributable to individuals convicted more than once during the year, but also that only 78 per cent of these active prolific offenders will actually be convicted during a 12-month period. The estimated number of four-plus convictions is 98,000, the number of convicted active prolific offenders is $98,000 \times 0.82 = 80,360$ and the total number of active prolific offenders is therefore $80,360 / 0.78 = 103,000$. This number will of course be subject to both demographic and random variation and is therefore rounded down to a ballpark estimate of 100,000. It is also implicit in the theory that this number is relatively stable over time. As individual offenders give up crime, up to 18 per cent after each conviction, the same number of high-rate offenders graduate into the active prolific offender group by being convicted for the fourth time.

Under this definition of a prolific offender, we would expect around 2 per cent of the population to become prolific offenders at some time between the ages of 10 and 35. Of these, 90 per cent will have joined the group by the age of 26 and 56 per cent will already have left by that age. Less than 5 per cent will still be active at the age of 40. The peak age for membership of the active prolific offender group is 24 when 40 per cent will be active at that age.

It is important to note that different definitions, and different models, will give rise to different numbers and that this particular calculation is only intended to provide an insight into the transient nature of the criminal population. This group of active persistent offenders has been highlighted because they are responsible for a disproportionate number of criminal convictions and, by inference, of crimes. The underlying theory suggests that the behaviour patterns that lead to conviction are consistent throughout the criminal career and predate the first conviction. Thus, for example, falling within the Home Office active persistent offender definition, by sustaining a fourth conviction, does not mark an increase in antisocial and criminal behaviour but simply confirms the status. Indeed, for about 18 per cent of these offenders, confirmation of the active persistent status marks the end of their criminal career as
they modify their behaviour to avoid criminal convictions in the future.

**Corollaries and Comments**

In many respects the theory proposed here is counter-intuitive. Crime is perceived by many as a youthful phenomenon. Certainly the peak age of offending would appear to support that contention, yet in the 1953 cohort over 50 per cent of offenders were convicted for the first time over age 19 and 50 per cent of convictions occurred over the age of 22. Our theory accurately predicts quarterly first conviction rates up to age 70 and beyond in the 1997 sentencing sample and almost 50 per cent of offenders convicted in that sample were over the age of 25.

In many theories of crime, maturation or simply getting older is thought to have a causal influence on desistance. Although the fact that crime diminishes in older generations is beyond dispute, our theory suggests that age itself is not a causal factor. More recent desistance research variously ascribe desistance to ‘turning points’ in the life course (marriage, employment, military service: Sampson and Laub 2003), developmental taxonomies (adolescent limited, life course persistent: Moffitt 1993), and the identification of life course trajectories modelled using cubic polynomials (Nagin 1999; Sampson and Laub 2003, 2005). Bottoms et al (2004, p 372) set out a list of concepts needed for the study of desistance, these were: ‘programmed potential; (social) structures; culture and habitus; situational context; and agency’. Changes in individual circumstances related to these concepts being instrumental in causing desistance. Although most of the desistance studies have found correlations between life events or personal circumstances and desistance or reductions in offending as part of the process of desistance, it is not clear whether these factors are in fact causal. As Kazemian and Farrington (2010, p 42) observed: ‘Since turning points and life events are not randomly assigned among individuals, it is difficult to assess whether these events are causes or correlates of desistance.’

In the analysis leading to the generation of our theory and in the mathematical models implementing it, we have identified strong evidence that both first convictions and reconvictions are governed by a proportional hazard survival process. These survival processes are characterized by the negative exponential inter-conviction
survival time distributions. The hazard operates on the individuals at risk of conviction, that is the active offenders. Within our risk/rate categories, after each conviction, the same proportion \( (p) \) of offenders are reconvicted and the same proportion \((1 - p)\) are never convicted again. If the offenders who were not reconvicted had continued to offend but desisted after some life event turning point at some random time after conviction, we would not expect to see the negative exponential inter-conviction survival time distribution. Unless, that is, the turning point events of the desisters occurred at precisely the right time to prevent their next conviction. In addition this precise timing would need to occur in both the high and low-rate categories and consistently across the whole life course.

A simpler and, we believe more plausible, explanation is that the proportion \((1 - p)\) of offenders do truly desist and that for them the turning point is conviction. Burnett and Maruna (2004), who interviewed prisoners just prior to release from prison, identified ‘hope’ (effectively the desire and will to desist) as a strong correlate of desistance in a ten-year follow-up. However, even among the most hopeful, social difficulties undermined their resolve and 82 per cent of the sample were reconvicted one or more times in the ten-year follow-up period. The decision to desist seems to have been made prior to release and, much as we would predict, 18 per cent managed to stick with it for at least ten years.

In our theory both desistance and the age–crime relationship are a by-product of the processes of offending, capture and conviction. We therefore suggest that the cumulative effect of the criminal justice system is the major cause of desistance. We return to this issue in Chapter 5 where we demonstrate that age-based theories are inconsistent with the evidence from the Offenders Index.

Conventional theories also suggest that the rate of offending slows down as individual offenders get older (Gottfredson and Hirschi 1990; Haapanen 1990) or that the rates of offending reduce as part of the process of desistance (Bushway et al 2001; Bushway, Thornberry, and Krohn 2003). This slowing down is apparent because samples of offenders convicted in increasing age-bands show lengthening inter-conviction times. Like Barnett et al (1989, p 347), our theory attributes this effect to the increasing proportion of low-rate offenders in the older age groups as more of the high-rate offenders have had the opportunity to desist after repeated convictions. For an individual, the rate of offending or conviction
remains constant throughout the criminal career; see Blumstein et al (1986) and Farrington (1986).

The most consistent predictor of reoffending is the number of previous convictions: the higher that number, the more likely is a reconviction. Our theory explains this apparently increasing probability as the effect of the reducing proportion of low-risk offenders with higher numbers of previous convictions. If for an individual the category membership were known, the \textit{a priori} probability of reconviction is constant and independent of his or her conviction history. We return to this issue in Chapter 6 in our discussion of the identification of risk/rate categories from psychological characteristics of offenders.

Although sentencing policy did change in the period between 1970 and 1992, the consistency of parameter estimates across cohorts indicates that these changes had little or no impact on either the rate of offending or recidivism. Further we shall see from the prison population forecasting work, discussed in Chapter 7, that after 1993, when there were major increases in the use of custody, continuing to assume no change in the recidivism probability or the offending rate accurately predicts the prison population. Thus, the hypothesis suggests that any changes in sentencing (tried on a major scale) between 1970 and the recent past, in particular the increased use of custodial rather than non-custodial sentences, were not effective in reducing conviction rates or recidivism probabilities. It might be argued that the consistency in parameter estimates simply reflects the capacity of the CJS to process offenders, but in our analysis the main driver of convictions would appear to be demographics. As we shall see in Chapter 7 demographics together with sentencing policy are sufficient to accurately model the prison population between 1970 and 1997.

As we shall see in Chapter 5, custody does not reduce recidivism probabilities compared with supervisory sentences and, in line with our theory, the rate of conviction for recidivist offenders remains the same over the entire career. Together these results imply that there is no incapacitative effect of shorter prison sentences. Custody does not reduce overall offending as crimes are effectively saved-up rather than averted whilst an offender is in prison. This is because, for active offenders in the same risk/rate category, the expected residual career length of an active released prisoner is the same as that of an active offender following a non-custodial sentence.
As has been discussed the theory does not pretend to be a complete theory. It does not begin to consider many aspects of offending or conviction. However, we show that it fits many of the known facts and is not known to contradict any particular large scale empirical finding in a way which cannot be reasonably explained by, for example, some kind of selection effect. An example of this would be when a group of offenders is selected on the basis of having committed only very serious offences, such as murder or large scale corporate fraud, which would lead to non-typical mixtures of the various offending categories. We will discover in Chapter 6 that the main offending categories identified in our theory can in part be distinguished on the basis of certain kinds of psychological information.

We know that some treatment programmes for offenders do reduce recidivism (see for example Goldblatt and Lewis 1998; Tong and Farrington 2006). These had not been used in the period of the analysis presented here (ie before 1998) to the extent required to make a significant impact on the overall rate of recidivism. However, we will show how to calculate their effects in Chapter 8 and also in more detail in the Appendix.

It may be true that improving education and employment would decrease criminality. However, any changes in these over the last 30 years of the twentieth century have not shown up as effects in our analysis, the risk and rate parameters having remained substantially constant over the entire period.

**Conclusion**

In this chapter we have shown how the results on recidivism probability and rate of offending obtained from the Offenders Index can be explained by a theory with four easily stated basic assumptions. With the addition of two further assumptions, a fifth concerning the transition from the universal informal sanctioning of

---

9 For example, it is known that the educational achievement of prisoners is typically lower than the national average (National Prison Survey, Main Findings, Home Office 1991). Of course one should not assume that improving educational standards would reduce offending from such a correlation. Low educational achievement may be a symptom of an underlying antisocial personality rather than a cause of offending.
young children to the near universal formal sanctioning of adults, and a sixth, concerning the relative probability of a first conviction, we have shown that we can explain the shape of the well known ‘age–crime’, or more correctly, ‘age–conviction’ curve. This is remarkable as the behaviour of offenders is assumed to remain the same throughout their active offending careers until such point as they decide, following a conviction, to cease offending. The rise of (recorded) offending between the age of criminal responsibility (10) and 17–18 years of age can be simply explained in terms of the increased use of formal sanctions as offenders become legally classified as adults. Additional factors may be that offenders have increased capacity for harm as they get older, and that society has run out of patience with those repeat offenders who return over and over again after informal sanctions. The decline in offending with age from 19–20 years onwards is explained by offenders being convicted and a proportion then ceasing to offend; it is not caused by an intrinsic reduction in the predilection for criminal activity with age.