Statistical Modelling of Escalation in Crime Seriousness

– through survival analysis, mixed-effects and mixture modelling approaches

Jiayi Liu

Submitted for the degree of Doctor of Philosophy

at Lancaster University,

12 January 2012.

Declaration

I declare that this thesis is my own work and has not been previously submitted for the award of any other degree here or elsewhere.

JIAYI LIU 12/01/2012

"Everything should be made as simple as possible, but not simpler."

(Albert Einstein)

Acknowledgements

I would like to take this opportunity to thank a number of people who have provided support and help to either my PhD study, or my life during this period of living in Lancaster. I have been working as a part-time research associate and a part-time PhD student. Therefore, my gratitude goes out to all my collaborators and my supervisor.

I am very grateful to my supervisor, Professor Brian Francis. He has been always provided patient guidance and training opportunities. In particular, during my study of PhD and being his RA, I gained the ability and confidence to present our work, and developed my personal research skills. His invaluable advice and support have guided me to successfully complete this thesis. Moreover, he has been a great source of advice when I am facing difficulties in my personal life. My gratitude also goes out to Professor Keith Soothill for his much needed help with understanding of criminology research, his patience in accepting my naive English writing skills, always provides valuable comments for all my presentations, and his comments on Chapter 2. I would also like to thank Professor Charlie Lewis (from the Psychology Department) for his understanding and support, in particular, for letting me to continue on completing my PhD final writing up while working as his RA.

My thanks also go to my colleagues from the Maths and Stats Department, such as Dr Juhyun Park, Dr Dennis Prangle, Dr Andrew Titman, Stuart Sharples, and Gareth McCray, who are always willing to discuss my research problems, explaining new ideas or providing help for even simple things such as coding, or basic understanding of statistical concepts.

Special thanks go to all my friends, who shared my happiness and sadness while I am living in Lancaster.

Finally, the most important people I would like to thank are my family who love me and support me with full heart.

This research was partially supported by the ESRC (ref no. RES-576-25-5020 and RES-576-25-0019).

Abstract

Escalation/de-escalation of offending is an important topic for criminal justice policy, but has been comparatively neglected in criminal careers research. This thesis introduces the Offenders Index (OI) dataset from the Home Office in England and Wales which is the preferred dataset for assessing escalation in this thesis.

Three major studies are then reported under two main research focuses. One research study focuses on 'serious offender escalation'. This study examines offenders who had been convicted of arson, blackmail, threats to kill, or kidnapping, and assesses whether they will be convicted of the most serious crime – homicide. This study suggested that 1 in 100 kidnapping offenders are likely to have a subsequent homicide conviction over a 20-year follow-up period, which doubles the risk of homicide conviction compared to the other three types of offenders. Moreover, offenders can double their risk of homicide conviction by being involved in multiple serious offences (among the four serious offences).

The second research focuses on 'general escalation'. This includes two studies: the first study examines the effects of two temporal scales, both age and order of convictions on escalation of seriousness by using a linear mixed-effects model. The results suggested that ageing is associated with de-escalation whereas the number of conviction occasions is associated with escalation, with the two processes pulling in different directions.

This is followed by the last study which examines the hypothesis that there are different types of underlying criminal development in escalation across offenders. Therefore, a combination of mixed-effects and mixture modelling methodology

has been developed to understand both individual crime growth curves and to distinguish latent types of crime development. A three-class solution has been identified by growth mixture modelling approach. The first class consists of the majority (88%) of offenders who are relatively stable in their seriousness in crimes, and have some tendency to de-escalate with age and some tendency to escalate with experience. The second class consists of 6.4% of offenders who have average high seriousness (7.6) at age 10, and have a strong de-escalation effect with age. The third class consists of 5.6% of offenders who have shown more diversity in crime seriousness, and also are involved with more high seriousness crimes.

The last study also provides a comparison framework to compare the linear mixed-effects model, group-based trajectory model, and growth mixture model through graphical investigation and proposed statistical diagnostic measures. For the particular data used in this thesis, the growth mixture model with three classes is preferred as the best fitting model compared to various other fitted models.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Crime is one of the most important subjects of social science research, because it is one of the topics with a high degree of concern by both public and government. Therefore, to study crime is essential not only for understanding crime itself but also for making major public policy decisions. Sequences of crime are committed by an offender over time, sometimes alone, and sometimes as part of a co-offending group. The study of such sequences of offending is known as "criminal careers" research.

Blumstein et al. (1986a, page 12) define criminal careers as "the longitudinal sequence of crimes committed by an individual offender". Therefore, it is important to understand criminal activity at both the population-level, and also at the level of the individual offender. Research by Farrington (1979), for example, had drawn researchers' attention to the study of crime and delinquency using longitudinal survey data. Nowadays the use of longitudinal data for studying the development of criminal activity is commonly adopted. Additionally, more longitudinal surveys of offending are available for both academic research purposes and public use.

Returning to the concept of "criminal careers", research topics on this area have been studied a great deal for the last two decades. Major works such as the two-volume book from Blumstein et al. (1986a,b), and a review paper by Piquero et al. (2003) provide classic introductions to the topic. Apart from these classic texts, there is also more recent work, such as books by Soothill et al. (2009) and Blokland

and Nieuwbeerta (2010) on the criminal career approach, and also a book which more focused on the quantitative methodology used in criminology especially in the area of criminal careers by Piquero and Weisburd (2010). The earliest use of the term 'criminal careers' appears to be in a seminal study by Glueck and Glueck (1930) whose book "500 criminal careers" is referred to later in this thesis. By 2010, there were 63 papers on the topic which specifically used the term criminal careers, and many more concerned with the topic but not specifically using the term. The reason why there is a large number of books and papers related to the study of criminal careers, is not only the importance of the study of the sequencing of crimes, but is also due to the range of topics which fall under the umbrella of criminal careers; split into what is known as the 'four dimensions' of criminal careers.

The four dimensions, first identified by Blumstein et al. (1986a) are: a) *participation*, which distinguishes someone who is an offender from someone who is not an offender given a length of observation period (Blumstein et al., 1986a, page 17); b) *individual frequency rates*, which refer to the number of crimes per year per active offender (Blumstein et al., 1986a, page 18); c) *seriousness or crime type mix*, which refers to the type of offences committed. Offenders can be 'specialists' who engage predominantly in only one offence or a group of closely related offence types, or offenders can be 'generalists' who engage in a wide variety of offence types (Blumstein et al., 1986a, page 18). Therefore, over an individual's criminal career, s/he can become either more or less specialised, or engaging in a changing mix of offence types producing escalation or de-escalation in seriousness; d) *duration*. The duration of someone's criminal career is the time between the first and the last offence. The determination of either the very first offence or the very last offence can be hardly observed in practice.

The study of criminal careers may therefore involve broad topics according to which dimension the researcher is interested in. To distinguish this thesis from many other studies, the present work focuses on the development of criminal activity since the offenders' first conviction, concentrating specifically on escalation/de-escalation in offence severity. Therefore, this thesis is firmly in the third dimension of criminal careers, which is concerned with seriousness and crime type mix over the life course. The definition of *escalation* in this thesis refers to "a tendency to move to more serious offence types" (Blumstein et al., 1986a, page 84).

Although escalation/de-escalation of offending is an important topic for policy and one way to understand criminal careers, surprisingly, there has been little research on escalation in crime seriousness in comparison with the other three dimensions in the criminological literature. This is fundamentally due to three methodological challenges in the study of escalation through offenders' criminal careers, which are how to measure the seriousness of an offence, what temporal scale to use for observing change in crime seriousness, and how best to analyse escalation through statistical approaches.

Due to the importance, but neglect, of the study of escalation in crime seriousness, this thesis will be using more modern and appropriate statistical methodologies to assess escalation, and to understand changes in crime severities both between individual offenders and among overall offenders.

This chapter introduces the area of this study by first discussing some important criminal careers concepts in the study of criminal careers, where those concepts will be assessed in relation to escalation in crime seriousness. Secondly, for each of the three research studies undertaken in this thesis, it describes the purpose of the study, its assessment of escalation, and its usage of statistical approaches. Each study is motivated to contribute new knowledge to the study of escalation. The chapter ends with a brief presentation of the structure of the rest of the thesis.

1.1 Factors relevant to the study of escalation

The purpose of this section, is to identify some of the most important factors which are likely to be associated with escalation and crime seriousness in the criminal career paradigm. These are gender, age at conviction and age at onset, incapacitation, and prior criminal history. The risk factors of age and gender on crime seriousness have been commonly examined, while incapacitation and offenders' previous criminal history have been studied less often. The focus of this thesis will therefore be concerned with the development of escalation and the identification of factors which may affect escalation and crime severity as offenders proceed in their criminal careers. In the following, each of these factors is briefly discussed in turn.

1.1.1 Gender

The differences in offending between male offenders and female offenders have been well recognised in the criminological literature. It has become a general fact that male offenders on average have a greater proportion of more serious crime-types as well as a greater level of involvement in crimes at any age, regardless of the source of data (Blumstein et al., 1986a, page 40). Other studies have also supported this statement, such as a more recent review study of gender differences by Lanctôt and Le Blanc (2002), and a study by Weiner (1989, page 67), which both concluded that gender differences in participation were particularly strong for serious crimes.

Gender differences in crime participation can also be seen through various types of data sources. For example, a study on arrest prevalence by Hamparian et al. (1978) found the gender ratio of males to females was as high as 6 : 1; one British birth cohort study which used police and juvenile court records by Ouston (1984), suggested the lifetime prevalence estimate for males was 29%, and for females was 6%. More recently, Piquero (2000) examined the Philadelphia National Collaborative Perinatal Project data, and found the proportion of the sample who had had a police contact by age eighteen for males was 31%, in contrast for females was 14%. Similarly, Piquero and Buka (2002) reported the proportions of individuals having a court contact by age eighteen for males and females were 19% and 5% respectively by using Providence National Collaborative Perinatal

Project data. There is also similar evidence from using self-report records. For example, Elliott (1994) studied the participation estimates for serious, violent offending based on the first eight waves of The National Youth Survey (NYS). This study suggested that females have a younger peak age of prevalence for serious offences compared to males. For instance, the ratio between males and females was 2 : 1 by age twelve, 3 : 1 by age eighteen, and increased to 4 : 1 by age twenty-one.

1.1.2 Age

The research interests on age and crime fall naturally into two areas. One is to understand how changing age is associated with criminal activities, such as prevalence of crimes, and seriousness of crimes. The other area is concerned more with the start age of the criminal career. This is known in the literature as the age at onset (the age at the first time of offending). Blumstein and Graddy (1982) and Blumstein et al. (1986a, page 42) state that "even though only a small fraction of youth at risk begin criminal careers at any given age, a concentration of initiation among youth is evident." Other researchers from different countries have also assented with Blumstein and his colleague via various types of data.

For example, a Swedish study from Stattin et al. (1989) studied a representative sample of both Swedish males and females from age 10 to 30. They suggested that the peak age of the first time conviction for males was between age 16 and 17, but for females was age 21 to 23. Elliott et al. (1983) used both official records and self-report records and suggested that the peak ages for males and females of initiation rates on offending were 14-18 and 13-16 respectively. Stattin et al. (1989) found that very few males had their first convictions after age 26. A study from Elliott et al. (1989) which used self-report records found that few males committed their first criminal offence after age 17. A study from the UK Cambridge study by Farrington (2002), followed school boys from age 7-8 up to the age of 40. He found that the mean age onset for those boys was 18.6 years old. Most recently, McGee and Farrington (2010) have suggested that there is no such thing as late onset offending, where offending starts after age 20.

Age effects are predominately related to maturity and opportunity. Maturity effects will be partly biological, thus, Gove (1985) identified the changing levels of testosterone in males over age. Maturation effects will also be developmental – with changing social bonds and responsibilities as an individual ages, and, specifically for sexual crime, will also reflect the ability of an individual to control a victim through physical strength and confidence. Opportunities for committing sexual crime will also change over the life course. At younger ages opportunities are high, but will decrease as an individual gains family responsibility. However, at older ages, opportunities again rise, specifically with access to children increasing both within and outside the family.

Therefore, age effects have been commonly examined in terms of its association with changing crime seriousness in the study of escalation, such as Rojek and Erikson (1982a), Blumstein et al. (1985), Datesman and Aickin (1984), and Britt (1996). A more detailed review has been provided in Chapter 2, section 2.1 for different evidence in relation to age and escalation. In this thesis, age is viewed as an effect of maturation, and therefore is examined as one of the temporal scales in crime escalation. It will be discussed further in Chapter 2, section 2.4.

1.1.3 Incapacitation

Incapacitation, according to Miles and Ludwig (2007, page 290) is "the inability of an incarcerated person to commit additional offences". Incapacitation is an important research area for crime control strategies. Ideally, the effect of incapacitation is to incarcerate active offenders and to reduce offending. However, this thesis focuses more on how the time spent in prison can affect the likelihood of escalation, rather than on crime control strategies.

In terms of the frequency of crime, a recent study from the United States by Bhati and Piquero (2008) showed evidence of crime drop among released offenders. Their large scale study of nearly 40,000 prisoners released from 15 U.S. state prisons, found that within a 3-year follow-up period after prison, 40% of released offenders showed decreased offending, in contrast just 4% indicated increased offending. Apart from this study, DeLisi and Piquero (2011) and Nagin et al. (2009) suggest that there is little other empirical research on the evidence of the effects of incapacitation and sentence length on persistence in crime, especially at the individual-level. Research on incapacitation and crime seriousness is less developed. Von Hirsch (1986) highlights some work on selective incapacitation which aims to direct incapacitation to those most at risk of committing future dangerous crimes. He states that the claims of such work are exaggerated, and there has been little good research on the effect of prison to reduce the severity of future crime.

Apart from the incapacitative effect on escalation, there is also an issue of accuracy in the measurement of time at risk of offending due to difficulties in calculating the exact length of imprisonment. Although the information of time at arrest and time at release is technically available from either police records or prison records, in practice information on parole or remand is very hard to obtain. This potential problem is also mentioned in Chapter 3 section 3.1.

The effect of time spent in prison is hard to get in practice, and has rarely been examined in statistically sound research. Therefore, the time spent in prison is examined in a relevant study ('Study 3') as the second temporal scale in the study of crime seriousness.

1.1.4 Experience of the criminal justice system

Research on the criminal justice system often focuses on policy issues in relationship to laws, such as decisions on sentencing (Piquero et al., 2003). This thesis, in contrast, considers the association between changing crime seriousness and the experience of going through the criminal justice system. Essentially, this links in with theoretical work which is loosely called 'labelling theory'. Soothill et al. (2009, page 104) state that "the labelling theory is concerned primarily with the official process of becoming a criminal, and how individuals become labelled as an offender or delinquent through involvement in the criminal justice system."

In relation to the effect of experience of the criminal justice system, the most important theorist is Edwin Lemert (Lemert, 1951) as he develops the concept of 'secondary deviance'. His idea is that there are a set of reasons which might lead to breaking the law in the first place, what he calls 'primary deviance'. Then the experience of the criminal justice system creates a set of new problems and these new problems might lead to what Lemert calls 'secondary deviance'. For instance, the stigma of being sent to prison may make it more difficult for offenders to find a job etc. Similarly, Becker (1963) argues that the effects of a court conviction might have serval outcomes. Firstly, society views the individual primarily as a deviant, and an individual may then be more likely to seek out delinquent groups for whom the social stigma of a court conviction is not relevant. Secondly, an individual's self-image may also change after being involved with the criminal justice system; they may view themselves as a criminal rather than as a normal citizen. Perhaps they begin to see themselves as thieves or sexual offenders, and therefore seek out further opportunities in a similar criminal enterprise. Thus labelling may contribute to increased specialisation or what Becker (Becker, 1963, page 32) calls the embracing of a 'master status'. Therefore, the embracing of a 'master status' as an offender is likely to lead to an increase in the seriousness of subsequent offending. Additionally, Piquero et al. (2003, page 394) state that "Labeling theory would view criminal justice intervention, especially serious criminal justice intervention, as doing more harm than good."

In terms of the empirical evidence for labelling which may lead to specialization, Soothill et al. (2009, page 104) reviewed several major studies and concluded that recent evidence on whether labelling leads to specialisation is inconclusive. For example, Sherman et al. (1992) carried out a study of domestic violence offenders who were arrested, and found a differential effect. For those who were employed, the effect of arrest (and therefore of being labelled as a domestic abuser) appeared to decrease subsequent episodes of domestic violence, while for those unemployed the episodes seemed to increase. Thus, for those with a work identity, labelling appeared not to have the serious impact suggested by the theory; in contrast, for those with no work identity the arrest appeared to increase specialisation in violence. There is also mixed evidence in studies on different types of offending. For example, Thistlethwaite et al. (1998) and Ventura and Davis (2005) studied domestic violence and found that a conviction reduced the likelihood of a subsequent reconviction in the same offence. However, a study on drink-driving convictions (Taxman and Piquero, 1998) found that a conviction increased the risk of a subsequent drink-driving offence.

It seems no quantitative study has directly examined whether experience of going through the criminal justice system can change crime seriousness in subsequent convictions. This may be due to the difficulty in measuring experience of the criminal justice system in a modelling framework. Therefore, this thesis attempts to examine the experience of the criminal justice system on the effect of escalation in crime seriousness through the order of conviction. The order of conviction is viewed as one way to reflect the effect of experience going through the criminal justice. The more conviction occasions one offender may have, the more experienced in going through the criminal justice the offender will be. Therefore, individual seriousness in offending can then be examined through order of conviction occasions. More importantly, the order of conviction is viewed as the premier temporal scale in the study of crime seriousness in this thesis, which will be discussed in Chapter 2 section 2.4 along with two other temporal scales.

1.1.5 Prior criminal history

Prior criminal history has been used commonly to predict recidivism and sentencing, but there has been relatively little work on the effect of escalation and seriousness. Offenders can commit more than one offence at the same crime occasion or at different crime occasions. Moreover, when brought to court, offences committed at different times can be aggregated together in a single court appearance. In this thesis, the research interest focuses on those offenders who had more than one conviction occasion in their criminal careers.

Normally, there are three components making up the concept of prior criminal history. The first component relates to the study of crime mix. This can include research questions like whether offenders specialise in specific types of crimes, or offenders who are generalists whether with likelihood of escalation or de-escalation during their criminal careers. In this thesis, summary measures of prior offending such as total number of different types of crimes during the observation period, the sequence of different types of crimes, and the identification of "gateway" offences which may lead to the most serious type of offending (e.g. homicide), are examined in the study of 'serious offender escalation'.

The second component relates to the study of frequency of criminal activities, therefore the amount of offences is of primary concern. Summary measures of prior offending such as total number of convictions in the offender's criminal life-span, or the total number of previous convictions, and total number of offences within each conviction occasion, will be studied through the entire thesis. The third component is duration, such as the length of time from the first target conviction (see Chapter 4) to the last target conviction, and the cumulative custodial sentence length.

Research on examining the effect of any of these three components of prior criminal history on escalation is rare. Three studies have been identified. Moitra (1981) studied adults arrested for murder, rape, robbery, aggravated assault, burglary, or auto theft in Washington D.C., during 1973. His work suggested that average seriousness is lower when there are a larger number of arrests in the offender's history. Soothill et al. (2002) found that those with a blackmail conviction were over five times as likely to become murderers as the general controls; and a previous conviction for kidnapping was shown to be a statistically significant risk factor for murder, when compared against general criminal controls and against violent controls. Rojek and Erikson (1982a) found that offenders with more than one arrest were more often arrested for serious offence types, such as 44% arrested for property offences among those offenders who had more than one arrest compared to 21% among those offenders who had only a one-time arrest. There is therefore a need for more systematic research on the effect of prior criminal history on escalation and crime seriousness.

1.2 Motivation and methodology

In this thesis, the study of escalation/de-escalation is carried out in two stages. The first stage focuses on "serious offender escalation" (Chapter 4: 'Study 1'), where those offenders who committed certain types of serious offence are analysed to estimate the risk of the most serious offence – homicide. The second stage focuses more on general escalation (Chapter 6: 'Study 2' and Chapter 7: 'Study 3'), where the sequences of crime seriousness over convictions are examined from the offender's first time conviction. Therefore, the nature of research questions at each stage is different, consequently the statistical methodologies which are applied for the two stages are different too.

1.2.1 Study 1

The first study focuses on "serious offender escalation" (Chapter 4), which compares and assesses the possible interrelationship among offenders who are convicted of arson, blackmail, threats to kill and kidnapping in terms of their risk of escalation into homicide. One motivation for this study is to identify whether there are "gateway" offences for homicide and whether those committing such serious offences need to be monitored. Therefore, the main research questions are as follows:

- For offenders with serious offences, in particular with each of the four serious offences, what proportions go on to be reconvicted for the same offence, or get convicted for one of the other three offences, or escalate their criminal activity to homicide?

- Whether certain combinations (mixed-type of crimes) and sequences of convictions on the four offences are the risk factors for subsequent homicide, in terms of both proportions and hazard of reconviction for homicide.

- Can age at conviction affect the risk of escalation to homicide among those offenders? Can age at onset and type of the first time serious offending affect the risk to homicide?

- Can the observed length of criminal career be a risk factor for escalation into homicide?

According to these research questions, firstly this study will use conventional descriptive statistical analysis, such as tables and plots of frequencies and proportions of offenders with such serious offences to explore differences among those offenders in terms of reconviction of homicide. Secondly, in order to examine and compare the time to homicide conviction from their first time conviction of a serious offence, a survival analysis approach will be adopted. This will allow the study of duration (rather than the only marginal proportions) from the serious offence of interest to the event of homicide, and can also compare the risk of reconviction of homicide (hazard rates) across the four types of offenders and difference between subgroup of offenders, in terms of gender and age.

1.2.2 Study 2 and Study 3

The focus of the second stage in the analysis of escalation is on more general escalation (Chapter 6: 'Study 2' and Chapter 7: 'Study 3'). The key research questions of interest are as follows:

- How should escalation trajectories at both population-level and individual-level be examined?

- How the age at conviction can affect escalation – effects of maturation?

- How the order of convictions can affect escalation – effects of experience of going through the criminal justice system?
- How custodial sentence (time spent in prison) can affect escalation?
- Whether there are any distinctive types of developmental trajectories in seriousness of crime among offenders?

Chapter 6 ('Study 2') therefore applies linear mixed-effects modelling, which allows the study of the sequence of crime seriousness at both individual-level and population-level through random effects and fixed effects. In 'Study 2', two temporal scales will be used to assess the effect on escalation, they are maturation (age on conviction) and experience of going through the criminal justice (the number of conviction occasions).

Chapter 7 ('Study 3') continues the work of Chapter 6 and uses the same dataset. However, this study will be seeking evidence of heterogeneity among the population of offenders. The potential problem of using the linear mixed-effects models for assessing individual sequences of seriousness is the normality assumption made on the random effects. Therefore, in terms of development of the underlying statistical methodology, 'Study 3' firstly explores the evidence of heterogeneity in the underlying distribution of random intercept and random slope. Then two alternative approaches are suggested through mixture modelling approaches. The first is the *group-based trajectory modelling* (GBTM) and which can examine whether latent types of developmental trajectories exist, the second one is the *growth mixture models* (GMM) approach which uses a combination of both mixed-effects modelling approach and the mixture modelling approach. Chapter 7, also compares results among these three methods, which are the linear mixed-effects model, GBTM model, and the GMM model.

Moreover, 'Study 3' adds a third temporal scale on the effects of escalation, which is the time spent in prison. In addition, 'Study 3' allows more flexible age-crime shape through non-parametric smoothing of age on the effects of increasing/decreasing crime seriousness.

1.3 Structure of this thesis

In brief, the remaining thesis is organised as follows. Chapter 2 provides a comprehensive literature review on escalation in crime seriousness. It summarises studies according to their measurement of seriousness, methodology which is currently applied to assess escalation, and their temporal scales used on observing sequences of crime seriousness.

Chapter 3 focuses on available data sources for the study of offending. Firstly it provides a systematic summary on the advantages and disadvantages of using either self-reports or official records for the study of offending. Secondly, it discusses three major longitudinal data sources in England and Wales on the study of offending. Then outlines the reason why one of the three sources – the Offenders Index dataset – was chosen in this thesis for studying escalation. Finally, it describes two distinct datasets which were extracted from the Offenders Index, one is for 'Study 1' on serious offender escalation and the other is for 'Study 2' and 'Study 3' on general escalation.

Chapter 4 to Chapter 7 will carry out three major studies on assessing escalation in this thesis. Chapter 4 ('Study 1') will examine 'serious offender escalation' through survival analysis. Then, Chapter 5 to Chapter 7 will focus on assessing general escalation in crime seriousness. Chapter 5 is a methodological chapter and describes relevant statistical methods which will be used in the subsequent two studies. The empirical results of studying general escalation are described in two sequence studies: Chapter 6 ('Study 2') and Chapter 7 ('Study 3').

Finally, Chapter 8 concludes the thesis by summarising the results from the three studies, and suggests some possible policy implications of the research. Additionally, it discusses potential development and future work from this thesis.

Chapter 2

The Concept of Escalation

The term *escalation* is used in the criminological literature for describing both increasing offence frequency and increasing offence seriousness. For example, some authors such as Sherman et al. (1991), Fagan and Western (2005) and Piquero et al. (2006) use the term escalation in the context of an increasing frequency of domestic assaults; whereas Blumstein et al. (1986a, page 84) define escalation as a tendency to move to more serious offence types. Similarly, de-escalation refers to the tendency to move to less serious offences. This thesis limits its attention to escalation in offence seriousness, and uses the type and nature of offending to assess escalation.

As just mentioned, this thesis is concerned with escalation in offence seriousness over the criminal lifespan. However, the study of escalation is rarely examined in the literature due to three methodological challenges. They are firstly how to measure the seriousness of an offence, secondly what temporal scale to use for observing change in crime seriousness, and thirdly how best to analyse escalation through statistical approaches.

Therefore, this chapter, firstly provides a review of the studies of escalation which will provide a clear picture of the major research on escalation. Then based on the review of the previous work, the three main methodological challenges in assessing escalation in offence seriousness are taken in order. The firstly challenge which is needed to consider is how best to measure the seriousness of offending;

the second challenge is to identify a suitable methodology to assess change in seriousness, and the final challenge is to engage with the temporal scale on which escalation is assessed. Each of these is considered in turn in relation to the previous work.

2.1 Review of the escalation literature

Escalation is an important concept in criminal careers, but, in contrast to other topics such as onset or specialisation, relatively little work has been carried out. It is crucial because the assumption of escalation in offending seriousness underlies much criminal justice policy. For example, Rojek and Erikson (1982b) report on a diversionary scheme for juvenile offenders which aims to divert them away from "a natural progression from lower to higher stages of delinquency"; more recently Beck et al. (2006) report on a diversionary scheme in Ohio which has the aim of "restraining the escalation of delinquency". Assumptions of escalation are particularly common with regard to sex offending. Firestone et al. (2006), in discussing the risk of escalation of exhibitionists to more serious sexual crime, state that "exhibitionism is not a benign act and should be dealt with seriously"; however Tappan (1950) is sceptical, reporting that one of the six myths of sex offending is that "most (sex) offenders have an escalation in the seriousness of their behavior". Non-sexual crimes can also be seen as precursors to sexual offending - Sample and Bray (2003) report that burglary is being considered by legislators as the gateway offence to sex offending and that nonsexual crimes are seen by many as precursors to sexual offending. There is also evidence that stalking is seen by legislators as a gateway offence to violence - the victim organization "Safe Horizons" states on its website that "The law (in New York State) helps victims by recognising that stalking is a crime of escalation that can result in physical injury and even death".

While many legislators seem convinced that escalation exists, previous work on escalation has come to widely differing conclusions. Blumstein et al. (1986a) provide one of the first reviews of the escalation literature. They identify that there is evidence from studies by Wolfgang et al. (1972) and Rojek and Erikson (1982a) that escalation is present for juvenile offenders; in contrast, studies by Moitra (1981) and Blumstein et al. (1985) show the presence of de-escalation for adults. A more recent review by Piquero et al. (2003), which refers to escalation by the term 'aggravation', identifies that the evidence for escalation is less certain. On the one hand, many studies in their review have found no evidence for escalation (e.g. Datesman and Aickin, 1984; Sheldon et al., 1987). On the other hand, more recent studies using different methodology have found evidence of escalation. Le Blanc and Frèchette (1989), in particular, found five developmental stages of escalation from age 10 to age 25, with offenders gradually increasing the severity of their offending. Yet further studies were ambivalent in their conclusions, finding some evidence of escalation among subgroups of the sample but not others (Britt, 1996; Loeber et al., 1998), or evidence using one method of assessment but not others (Blumstein et al., 1988). The time is therefore ripe for a reappraisal of escalation focusing on using recent statistical advances to critically appraise and assess the concept.

2.2 Measuring crime seriousness

The first methodological issue which relates to the assessment of escalation is how to measure crime seriousness. In fact, measuring the seriousness of a specific offence can be carried out by various means. The recent work by Stylianou (2003) and Ramchand et al. (2009) together provide an excellent review of the literature. Views of crime seriousness can be gathered either from members of the general population, or from criminal justice professionals. These two groups may well have different views on seriousness, and a choice needs to be made on which target group to use. Once this choice is made, there are varying methodologies - surveys can be used to assess public views of crime seriousness, and either expert panels, surveys or examination of official sentencing records can be used to assess professional views. Measures of seriousness will also vary across jurisdiction and will change over time (Francis et al., 2001).

Focusing first on how the general public view crime, Wolfgang et al. (1985) surveyed a large sample of the US population by presenting them with a series of 204 crime vignettes, and asking them to rate each vignette against a standard crime of stealing a bicycle.

More commonly, crime seriousness scales developed for or devised by criminal justice professionals are used. *Expert judgement* is used by some countries and states to construct crime seriousness scales that are then used as part of the criminal justice sentencing process. Thus the state of Florida has a ten category offence severity ranking chart for non-capital felonies (Florida Department of Corrections, 2009), ranging from the least serious level 1 (e.g. possessing a still, operating an aircraft under the influence) through to level 10 (e.g. unpremeditated homicide). Each level is used to determine a score which differs according to whether it represents the primary offence at conviction, an additional offence, or a prior convicted offence. The scores for current and past offences are then used to produce a 'total sentencing guidelines' score, which also takes into account a range of other factors such as victim injury, whether a firearm was used, and if there was a prior serious felony. This final score provides a guideline recommendation for sentencing; the judge can choose to follow the guideline or to mitigate or aggravate the sentence.

A less common alternative is to estimate a crime seriousness scale from *official court data.* A common approach is to use average length of prison sentences. Thus, Carrington et al. (2005) used a measure of seriousness developed by the Canadian Centre for Justice Statistics which was computed using average lengths of prison sentence. Reilly and Witt (1996) also used average length of sentence in a study of the effect of economic variables on crime.

Finally, one methodology not mentioned previously for measuring crime seriousness from criminal justice professional, is the *paired comparison* approach. Theoretically it can also be used to build such scales. Thurstone (1927) first applied the methodology to an experimental study of students who were presented with all combinations of pairs of 19 distinct offences and asked to determine which offence of each pair was more serious. Francis et al. (2001), in contrast, used court conviction data, taking pairs of offences coming before the court at the same court appearance, and using sentencing information to determine preference. A closely related methodology is that of conjoint analysis – Brocke et al. (2004) give an application applied to crime seriousness. In contrast, Ramchand et al. (2009) used yet another approach, taking the temporal ordering of offences to determine seriousness order. This method relies on an assumption that more serious crimes occur after less serious crimes in the criminal life course, which is supported by evidence from the developmental criminology literature (Farrington, 1986b; Le Blanc and Loeber, 1998). Ramchand et al's work is however unsatisfactory if the crime seriousness scale is being used to investigate escalation, as their methodology makes a fundamental assumption that escalation is present over the life course in order to develop the crime seriousness scale. To avoid circularity, methods of measuring crime seriousness need to be independent of any assumption of escalation.

Previous work which is specifically on escalation has in the main adopted two approaches to the assessment of crime seriousness. The most common approach is to categorise offences directly, using the expert knowledge of the researcher. So, Rojek and Erikson (1982b) used a categorised scale and looked at transitions from runaways and status offences to other crimes; Blumstein et al. (1988) constructed an ordered ten-category scale. An alternative, adopted by the Danish study of Kyvsgaard (2003), is to categorise offences either according to the maximum sentence which could be imposed by law, or by the imposed sanction given in the court proceedings. The US Crime Severity score developed by Wolfgang et al. (1985) has also been used in some escalation studies (e.g. Loeber et al., 1998).

In this thesis, an innovative measurement is used to measure the crime seriousness in 'Study 2' and 'Study 3'. This is a continuous measure of the seriousness of crime (Francis et al., 2005), which was developed using sentencing data in England and Wales (detailed methodology see Chapter 5 Section 5.2.1). However, in 'Study

1', there is not a specific approach used to measure the seriousness of crimes, since it is a study of four particular types of crimes – arson, blackmail, kidnapping, and threats to kill – aggravated to the most serious crime of homicide (murder and manslaughter). As there is general agreement that homicide is the most serious crime, the four other types of crime will rank below homicide on any measure of crime seriousness.

2.3 Methodological approaches to assessing escalation

The second methodological issue is the choice of quantitative method which has been used to assess crime seriousness for the study of escalation. In the escalation literature, there are three main statistical methodologies that have been developed. The first is based on *crime-type switching tables* which extend specialisation measures into escalation. The second is *analysis of means* where graphs of average escalation over time are examined visually. The third is a *regression* approach.

Crime-type switching tables

The construction and analysis of crime-type switching tables is the most common method of studying how crime changes from one occasion to the next. The criminal occasions are taken to be arrests or convictions and are categorised into distinct offending groups with the researchers using their own judgment as to the severity of each category. An example for a crime-type switching table between successive convictions is showed in Table 2.1 (Blumstein et al., 1988).

The crime seriousness at conviction k for crime type 1 to crime type j are in the columns to the left, and are decreasing seriousness from crime type 1 to crime type j. In contrast, type of crimes in the next conviction $k+1$ are represented at the top of this table. The transition matrices are therefore made up of individual transition probabilities, p_{ij}, which reflects the proportion that crime type i at

Table 2.1: Crime-type switches between successive convictions.

			Crime type of $k+1^{st}$ conviction → Decreasing seriousness						
			1	·	·	i	·	·	j
Crime type of k^{th} conviction	↓ Decreasing seriousness	1	p_{11}						
		·		·					
		·			·				
		i				p_{ii}			
		·					·		
		·						·	
		j							p_{jj}

conviction k is followed by a crime type j at conviction $k+1$. The p_{ij} is estimated as follows:

$$p_{ij}(k) = n_{ij}(k)/n_{i.}(k), \tag{2.1}$$

where $n_{ij}(k)$ is the number of convictions (arrests) for type i at occasion k and conviction type j at occasion $k+1$; and simply $n_{i.}(k) = \sum_{j=1}^{N} n_{ij}(k)$; and N is the total number of crime types examined. Therefore, the diagonal elements in this table, say the p_{ii} to p_{jj}, represent the observed proportions of specialisation. The upper-right corner of the table indicates movement to less serious offence types (de-escalation), while the lower-left corner indicates movement to more serious offence types (escalation).

Movement in the crime switching table is assessed by various means in the literature. Using four sequential transition matrices ($k = 1, 2, 3, 4$) on juvenile offending, Rojek and Erikson (1982a) applied a chi-squared test to examine whether the four transition matrixes are equal and constant over time. They suggested that, as there was no evidence of change in the transition matrixes, then this implied that there was no evidence of escalation. Furthermore they tested the relationship of age at first offence with offence escalation, and suggested that there was no evidence of offence escalation for any of the three age-of-onset groups.

Blumstein et al. (1988) proposed a crime-specific measure of escalation (E(k))

which is based on the Farrington (1986a)'s F(k) index (the forward specialisation index). The basic idea of measuring escalation through the transition matrices is to capture the pattern of switches from less serious crime type to the more serious crime type (the lower-left corner of Table 2.1). Similarly, measuring de-escalation is to capture the pattern of switches from the more serious type to the less serious type (the upper-right corner). The index of escalation for crime type i at conviction number k is defined as follows:

$$\mathrm{Esc}_i(k) = \frac{\mathrm{Obs}_i(k) - \mathrm{Exp}_i(k)}{\mid \mathrm{Max}_i(k) - \mathrm{Exp}_i(k) \mid}, \quad i = 1, ..., N; \tag{2.2}$$

where

$$\begin{aligned}
\mathrm{Obs}_i(k) &= \sum_{j=1}^{i-1} p_{ij}(k) \\
\mathrm{Exp}_i(k) &= \sum_{j=1}^{i-1} p_{.j}(k) \\
\mathrm{Max}_i(k) &= \begin{cases} 1 - p_{ii}(k) & \text{if } \mathrm{Obs}_i(k) > \mathrm{Exp}_i(k) \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

In this measure, only the elements from below the diagonal ($j < i$) are considered (escalation). The calculation of $\mathrm{Obs}_i(k)$, for instance, is the sum of observed proportions below the diagonal for each row i, which relates to escalation movements from stage k to stage $k+1$ for each crime type i. The $p_{.j}(k)$ is the sum of jth column (at $k+1$). Basically, it describes the proportion of each crime-type at the $k+1^{st}$ conviction. The $\mathrm{Exp}_i(k)$ hence is the sum of the marginal proportions of $p_{.j}(k)$. The expected proportion describes switching which is independent of prior crime type. Therefore, if the observed proportion to escalation for each type i at k^{th} conviction, $\mathrm{Obs}_i(k)$, is greater than the expected proportion – $\mathrm{Exp}_i(k)$, this indicates evidence of escalation. The measure of escalation is standardised with respect of the maximum possible switching for transition below the diagonal.

Therefore, for a specific crime-type, this measure takes values in the range -1 (indicating complete de-escalation) to 1 (indicating complete escalation). A value of 0 means that there is no evidence of either escalation or de-escalation for a specific crime type i.

Similarly, the index of de-escalation is given:

$$\mathrm{Desc}_i(k) = \frac{\mathrm{Obs}_i(k) - \mathrm{Exp}_i(k)}{\mid \mathrm{Max}_i(k) - \mathrm{Exp}_i(k) \mid}, \quad i = 1, ..., N; \tag{2.3}$$

where

$$\begin{aligned}
\mathrm{Obs}_i(k) &= \sum_{j=i+1}^{N} p_{ij}(k) \\
\mathrm{Exp}_i(k) &= \sum_{j=i+1}^{N} p_{.j}(k) \\
\mathrm{Max}_i(k) &= \begin{cases} 1 - p_{ii}(k) & \text{if } \mathrm{Obs}_i(k) > \mathrm{Exp}_i(k) \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

The elements *above* the diagonal ($j > i$) are considered only in the calculation of de-escalation. Then the measure of $\mathrm{Desc}_i(k)$ also takes values from -1 to +1, indicating complete escalation or complete de-escalation, respectively.

Then a single measure of escalation (E) which is based on both indexes is:

$$\mathrm{E}_i(k) = \frac{\mathrm{Esc}_i(k) - \mathrm{Desc}_i(k)}{2}. \tag{2.4}$$

Therefore, when there is an evidence of escalation in both indexes of $\mathrm{Esc}_i(k)$ and $\mathrm{Desc}_i(k)$, then the combined measure $\mathrm{E}_i(k)$ will take a high positive value, and $max(\mathrm{E}_i(k)) = 1$. In contrast, the $min(\mathrm{E}_i(k) = -1)$ indicates complete de-escalation. The value of 0 again shows no evidence of either de-escalation or escalation.

Blumstein et al. (1988) applied this method to a series of arrest histories of adults who were aged from 17 onwards in the Detroit SMSA and in the remaining

Southern Michigan region in the period 1974 to 1977, Blumstein and his co-workers found that there was no evidence of either escalation or de-escalation by using the E measure.

Nevertheless, there are three main disadvantages of using Blumstein's escalation measure E. Firstly the interpretation of E lacks a clear meaning with no formal statistical test available to test difference from zero. Secondly, E is unable to test for significant differences across subgroups, such as age and gender difference in terms of escalation patterns. Finally, there is an assumption that conviction/arrest at kth and subsequent conviction/arrest at $k+1$st are independent, which is very unlikely to be true for criminal activity.

In contrast to Blumstein et al. (1988)'s model-free approach, Britt (1996) reviewed two modeling approaches on offence escalation based on crime-type switching tables; these were originally introduced in the area of social mobility. One approach was proposed by Sobel et al. (1985). He reparameterised the quasi-symmetry (QS) log-linear model, which was introduced by Caussinus (1966) and popularised in the analysis of social mobility tables, constructing parameters which assessed exchange (or specialisation) and structural mobility (or escalation). The QS model is defined as follows:

$$F_{ij} = \beta_i \beta_j \alpha_j \delta_{ij} \gamma_{ij} \tag{2.5}$$

or the alternative formulation is the log-linear regression, and is given:

$$F^*_{ij} = \beta^*_i + \beta^*_j + \alpha^*_j + \delta^*_{ij} + \gamma^*_{ij} \tag{2.6}$$

where the F_{ij} is the observed frequency of say offenders with crime-type i at the kth conviction/arrest and were convicted of crime-type j at the $k+1$st conviction/arrest. Then $F^*_{ij} = log(F_{ij})$. The F^*_{ij} is modelled by four parameters, which are decomposed into symmetric marginal β^* and association δ^* parameters as well as asymmetric marginal α^* and association γ^* parameters.

The β_i^* and β_j^* parameters are the estimates of the marginal means for row of i and column of j. They are both constrained to be equal. The parameter of α_j^* basically examines the equality assumption on β_i^* and β_j^*. If $\alpha_j^* < 0$, it indicates the category j at $k+1$st conviction holds proportionally fewer cases in the marginal distribution of the origin variable (the kth conviction). Similarly, $\alpha_j^* > 0$ indicates that cases from destination category j increases its proportion of cases in the marginal distribution.

Apart from estimations of the marginal distribution, the QS model also assesses associations among the off-diagonal cells (de-escalation/escalation) through the δ^* and γ^* parameters. The parameter of γ_{ij}^* is the symmetric term and is restricted by making $\gamma_{ij}^* = \gamma_{ji}^*$. This basically, implies equality assumption between upper-right corner and bottom-left corner. In contrast, the parameter of δ_{ij}^* relaxes such assumption on γ and is basically testing the equality of the upper-right corner and bottom-left corner. If $\delta_{ij}^* < 0$, it indicates more cases moves to cell F_{ji} from cell F_{ij} than moves from F_{ji} to F_{ij}, in other word indication of de-escalation (move to upper-right corner). In contrast, $\delta_{ij}^* > 0$ suggests more cases move to bottom-left corner – escalation. Therefore with combination of the four parameters, QS model is able to assess marginal distribution of the structure and switching movement away from the diagonal.

In addition, a second approach extended the above model to what was named the conditional quasi-symmetry (CQS) approach (Bishop et al., 1975:299-300, Sobel, 1988:172-176), which allowed the testing of equality of the escalation parameter across groups of contingency tables. Britt applied these two methods to the data used by Blumstein et al. (1988), and examined four sequential arrest transition tables. However his findings suggested no strong evidence of escalation either within or between racial groupings.

The advantages of using QS and CQS approaches are firstly the coefficients of escalation can be tested as to whether they are statistically different from zero. Secondly, tests for differences among subgroups can be provided statistically, such as assessing differences in escalation among ethnic groups, or age groups. Thirdly, association among two marginal means, say at conviction k and conviction $k+1$, can be tested. There is no need for assumption on independence between successive convictions or arrests.

However, the disadvantages of the crime-type switching table approach are now well known. Spelman (1994) pointed out that the ecological fallacy holds - that measures based on aggregates of individuals do not necessarily provide information on individuals. Osgood and Schreck (2007) identified another disadvantage - that escalation is measured only on offences which are temporally adjacent and ignores information on other adjacent offences. Thus the sequence of theft-robbery-theft would give a pattern of escalation followed by de-escalation, whereas the similar sequence theft-theft-robbery would identify stability followed by escalation. Although methods could be devised for analysing three- and four-way tables of transitions, cell counts would soon become too small for an efficient analysis. Thirdly, the number of crime-type categories for a given size of sample of offenders affects the size and accuracy of the resulting measure. Finally, the researcher has to devise some ordering of the crime types in order of seriousness. Blumstein et al. (1988), for example, used a number of different underlying crime scales for ordering their ten categories, and identified that changes of ordering occur in moving from one scale to another.

Analysis of means

A few studies have explored escalation by using what are essentially summary descriptive statistics, calculating mean levels of crime seriousness over the lifecourse.

A chapter in Kyvsgaard (2003) examined escalation in Denmark by using such methods. Using a six category measure of seriousness derived from standard sanctions available from offence legislation, she explored possible variables which are associated with escalation, such as age, gender and length of criminal careers. She found evidence of de-escalation within all age groups apart from the 20- to

24-year-olds; that males on average committed more serious crimes, and that the average seriousness is higher for longer criminal careers. While shorter criminal careers showed evidence of de-escalation, longer careers showed an initial escalation period followed by subsequent de-escalation.

Similarly, Carrington et al. (2005) studied court careers for Canadians who were born in 1979/80 and followed up to their 22nd birthday. Measuring seriousness by the Canadian Centre for Justice Statistics scale (developed using average lengths of prison sentence), they measured escalation by the difference in seriousness from the first to last convictions. They found an overall de-escalation pattern for offenders with ages of onset up to 18, and an escalation pattern for ages of onset of 19 and 20 years. They also examined the differences among three types of offender (adolescent-limited offenders, persistent offenders, and adult onset offenders) in terms of life-course typology of court careers and found little difference between these groups.

Such approach is informative and essentially visual. However, it provides no statistical testing as to whether an observed increase or decrease is important or significant. It also fails to control for other variables or to take account of the multi-level structure of longitudinal data.

Regression

The Blumstein et al. (1988) study used not only transition matrixes based on crime-type switching table to examine escalation, but also proposed a regression approach to detect multiple-step trends in seriousness (long-term trend). They firstly defined a measure of average seriousness of the kth arrest for offenders with a complete sequence of p total arrests, say $S_p(k)$, and is given:

$$S_p(k) = \sum_{i=1}^{I}[s_i * N_i(p,k)]/N(p,k) \text{ for } k = 1, .., p \tag{2.7}$$

where I is the total number of crime types, whereas s_i is the seriousness score of the ith crime type; $N_i(p,k)$ is the number of offenders with p total arrests whose

kth arrest was for crime type i; then $N(p,k) = \sum_i^I N_i(p,k)$[1] is the total number of offenders with p arrests at the kth arrest.

Therefore, they expected that the averaged seriousness scores depended on the order of arrests (k), while also controlling the effect of the total number of arrests p (the career length). They then applied a linear regression to these average seriousness measures, modeling them as a linear function of k and p. The regression coefficient of order of arrests (k) indicates whether escalation or de-escalation to offending. Their results suggested that white offenders are more likely to increase in average seriousness over successive arrests, but the average seriousness varies across offenders. In contrast, they suggested that people who had longer observed career length are more likely have lower seriousness on average.

Regression overcomes much of the criticism of the crime switching tables and exploratory approaches described above. However, the work of Blumstein et al. fitted models at the aggregate level rather than the individual level. It also fails to control for other important covariates affecting crime seriousness. In Chapter 5 section 5.4 three statistical approaches will be introduced such as the linear mixed-effects model and mixture modelling approach which will be able to address these shortcomings.

2.4 Temporal scales in crime escalation

The final methodological issue is how best to measure and assess the temporal ordering of offences. Researchers have taken two approaches in choosing a temporal scale to assess escalation over time. Some studies have used arrest number (e.g. Blumstein et al., 1988) or conviction number (e.g. Kyvsgaard, 2003); other studies have tended to use age as the temporal scale. The choice of temporal scale is independent of methodology. For example, with crime switching tables, some researchers have summarised seriousness measures into age groups and looked at

[1]The original notation for $N(p,k)$ given in Blumstein et al. (1988, page 337) is $N(p)$, which I suggest is an incorrect notation.

transitions between age groups (e.g. Rojek and Erikson, 1982b); others have looked at transitions between adjacent arrests or convictions (e.g. Britt, 1996).

An important point to notice is that these scales measure different criminological processes. Change in escalation by age can be thought of as a developmental or maturational process, whereas change by arrest number or conviction number can be thought of more as an experiential process, where offenders gain expertise and knowledge of criminality as the number of previous contacts with the criminal justice system increases. The latter process is a form of state dependence.

Additional to these two temporal scales, a third temporal scale is proposed in this thesis, which can also affect escalation over time. This is the *time spent in prison*. As expected, offenders who are convicted of certain types of crimes, normally for more serious crimes, are likely to be awarded a custodial sentence and the offender will spend time in prison.

Criminal career researchers have primarily been concerned with time spent in prison through the concept of incapacitation. Incapacitation, according to Miles and Ludwig (2007, page 290) is "the inability of an incarcerated person to commit additional offenses". Therefore, in reconviction studies interested in the risk of offenders once released from custody, the time between any two conviction occasions should be corrected by subtracting the length of any time spent in prison, giving as estimate of 'street time' or 'time at risk'.

However, for studies on general escalation, the issue of street time is less of an issue. Research in the United Sates over the past two decades showed evidence that imprisonment has contributed to the crime drop (Bhati and Piquero, 2008). However, according to DeLisi and Piquero (2011) and Nagin et al. (2009), there is little empirical research on the evidence of the effects of incapacitation and sentence length on persistence/desitance in crime, especially at the individual-level.

Two possible effects of imprisonment are conceptualised in this thesis, one positive and one negative. For instance, the positive effect can be that while offenders spend time in prison, they will reappraise their life and value to society, and will potentially gain new and practical skills for living after returning to society. In such a case, the effect of imprisonment on future offending can be positive and offenders can be deterred from committing future crimes or will commit less serious crimes on release.

However, every coin has two sides. In contrast to the deterrence idea, others have proposed that prison is a "university of crime for the young", especially for those in overcrowded jails (Prison Reform Trust, 2008). In this scenario, offenders will meet other offenders in prison who are more experienced in crime and with a more serious conviction history, and they can learn new criminal skills and build up enhanced networks of other criminals. Therefore, the experience of prison in such a case can possibly enhance the offender's knowledge of crime and the skills needed to commit crime, with the potential of the offender moving on to more serious offences after returning to society.

Thus, whether the overall effect of prison is positive or negative, time spent in prison needs to be considered as a third temporal scale, and it is separate from the experiential scale of 'number of convictions' discussed earlier.

As already hinted, the neglected area of focus in these earlier studies has been the temporal scale on which escalation has been measured. Researchers have looked either at changing escalation as an offender age, or changing escalation as the number of offences, number of arrests or number of convictions increases. However effect of time spent in prison has rarely examined on escalation in future criminal activities in quantitative research. Moreover, previous work has not considered that these three effects may act together and that observed escalation might be a combination of these processes. This thesis argues that these are distinct criminological processes. The first one – age effect – can be thought of as change due to maturation, and the second one – number of conviction occasions – as change due to increasing offending experience. The third one – time spent in prison – can be though of as a custodial effect.

Therefore, it is possible to conceive of offenders where experience is gained rapidly with very little change in age; it is also possible to conceive of other offenders who have only an occasional conviction widely separated in years, who will gain experience slowly over time. Therefore, in contrast to other work on this topic, this thesis proposes that there are three types of escalation process in crime seriousness over the criminal lifecourse - escalation due to age, escalation due to experience, and escalation due to time spent in prison. One focus of this thesis will be how to assess the magnitude of these different types of escalation through the following three studies.

2.5 Summary

This chapter has firstly addressed the definitional issues of what is escalation in the study of crime seriousness. While different authors have used the term in different ways, the working definition for this thesis was determined to be the tendency to commit more serious crime. Therefore, the types of offences and their seriousness matter in this thesis on the study of escalation. Then major qualitative studies on crime seriousness escalation in the literature were reviewed. Those major work include studies from Wolfgang et al. (1972), Moitra (1981), Rojek and Erikson (1982a), Datesman and Aickin (1984), Blumstein et al. (1985), Blumstein et al. (1986a), Sheldon et al. (1987), Blumstein et al. (1988), Le Blanc and Frèchette (1989), Britt (1996), Loeber et al. (1998), and Piquero et al. (2003).

The second part of this chapter summarised three methodological challenges in the literature of studying escalation. They are (a) measuring crime seriousness; (b) methodological approaches to assessing escalation; (c) two temporal scales in crime escalation.

Therefore, this chapter has provided a clear insight of what is being studied when it refers to escalation and what issues matter for assessing escalation in the literature. Then in the following chapters, it will be possible to introduce improved methods into the study of escalation and make contributions to the quantitative research on escalation. However, in the next chapter, the datasets which are used in this thesis will be introduced.

Chapter 3

Data Sources on Offending Behaviour

This chapter will introduce and then discuss two types of data sources on studying offending in general – official records and self-reports. Both types of sources are very useful to study the longitudinal patterning of criminal activity. However, there are limitations of using either self-reports or official records. Therefore, the advantages and disadvantages of both data sources will be discussed.

In the second section, three datasets collected from England and Wales for studying offending – the Offending, Crime and Justice Survey, the Police National Computer, and Offenders Index. Each dataset will be described, and then the reason will be provided on why the Offenders Index (OI) is chosen in the thesis for the examination of escalation . In the final section, two different datasets which are both extracted from the OI will be described for the studies on escalation in the subsequent chapters.

3.1 Type of crime data source on offending

To study criminal careers on offending, both *official records* and *self-reports* are commonly used in criminological research. Official records and self-reports are different measures of criminal activity. Official records can be the records of arrests, court referrals, or convictions collected by the police or court system. Self-reports in contrast are usually in the form of a survey, which individuals are asked to report their own criminal activity. Sometimes, the sample is of the general population; other times it consists of individuals of specific age and locality.

There are a large number of datasets used in Criminal Career research. Farrington (1979, page 291, 294) and Piquero et al. (2003, page 364) together probably provides the most comprehensive list of major surveys which contain the major data sources on longitudinal studies of criminal careers. Three of these sources are heavily used in the study of criminal careers. The first one is the Philadelphia Birth Cohort Studies (Wolfgang et al., 1994; Figlio et al., 1994), which is an American dataset of official records. This study consists of two birth cohorts – the 1945 birth cohort with $9,945$ boys and the 1958 birth cohort with $27,160$ boys and girls. All cohort members were born in Philadelphia and were also living there at age 17. The second source is the Cambridge Study in Delinquent Development (Farrington, 1989), which sampled 411 almost entirely white working class boys aged $7-8$ (born in 1953-54) living in Camberwell, London in 1961. The cohort was followed up regularly to age 50. This study uses both official and self-report records. The third source is known as the Glueck Study (Glueck and Glueck, 1930, 1950), and later is used by Laub and Sampson (1994) in their work on life course criminology. Basically, the Gluecks studied the developmental life course criminology from 500 delinquent and 500 non-delinquent males (seven to eleven years in age) from Massachussets. Data was originally collected in 1940 through psychiatric interviews with subjects, parent and teacher reports, and matched up with official records obtained from police, court, and correctional files. The subjects were followed up and subsequently interviewed again in 1949 and 1965.

Both official records and self-report data are widely used for both quantitative and qualitative research. Research has highlighted the advantages and disadvantages of these two data sources. For example, Piquero et al. (2003) and Blumstein et al. (1986a, Chap. 4) both summarised the potential benefits and drawbacks of using either official records or self-reports from the viewpoint of quantitative research in criminal careers. In contrast, Merton (1957, Chap.4), and Kitsuse and Cicourel (1963) both argued whether the use of official statistics for studying deviance is appropriate from the aspect of sociologists. More recently, an empirical study by Kirk (2006) examined the differences in using data of self-reports and official records on inferences about adolescent life-course of crime. Following this discussion, the advantages and disadvantages of each data source can be summarised into four issues: reliability, sampling, street time, and sociological relevance.

First of all, there is the issue of *reliability*. For self-report data, research from Hindelang et al. (1981), Blumstein et al. (1986a, Chap.4), and Weis (1986) have pointed out that problems such as response errors, analytical problems from questionnaire responses, and the design of survey instruments can lead to distorted results. First in terms of the response errors, it arises mainly from problems in respondents misclassifying or recalling an event, or intentional misrepresentation by respondents. For example, recall questions are commonly used in self-reports, so respondents may be unable or unwilling to give a reliable answer. Thus, respondents may be unable to recall the frequency and timing of their criminal activities accurately, especially for those high-rate offenders and those with a history of heavy drug and alcohol use; or respondents may lie about their involvement in crime; or respondents may be uncertain about which events are to be counted as police contacts, arrests, or convictions.

Secondly, there are some analytical problems which relate to subjective interpretation towards survey items or ambiguous responses by respondents. For example, people's interpretations of self-report items may change as they age (Lauritsen, 1998). The final problem relates to the design of survey instruments. Blumstein et al. (1986a, page 98) mentioned that various survey methods such as whether questionnaires or interviews are used, whether responses are anonymous or nonanonymous, and the effects of different interviewer attributes can all potentially bias or limit the validity of responses.

For official records, the issue of *reliability* results from two main structural sources of recording errors: misclassification of events and nonrecording (Piquero et al., 2003; Blumstein et al., 1986a, Chap. 4). Classification errors can result from differences among local courts or police forces in their classification of offence types, for instance. Nonrecording errors may occur because the event does not meet reporting standards, such as the requirement for a fingerprint or disposition data, which may not be available (Michigan State Police, 1983a,b). The nonrecording of some events can obviously understate the number of arrests.

For both self-reports and official records, there is a common potential analytical issue affecting the reliability, which is that the reported time of an event may not reflect the real time of the event. Thus the date of conviction will not reflect the true date of offence, but is often taken to represent date of offence. This issue is related to the nature of using recall questions in self-reports – inaccurate memory of time at offence, or using conviction occasions in official records – arrests or conviction may be discovered, for example after a series of crimes have already happened.

The second issue relates to *sampling*. Using self-reports, the sampling problem relates to the researchers' ability to obtain a sufficient number of cases of serious offending. In addition high-rate offenders who are involved in self-report studies are more likely to drop out over time or to lie about their involvements in crimes (Brame and Piquero, 2003). On the other hand, using official records, selection bias occurs through the use of the official records of arrests, or convictions, which are offender-based samples. Only offences that come to the attention of officials are reported in the Official records. Therefore, using official data can understate the total number of offences compared with using self-report data. Moreover, official records may also include more of the worst offenders and the worst offences compared to the equivalent self-report data. Therefore, official records capture a small fraction of the total number of crimes but more serious offences; in contrast, self-reports capture a larger fraction of the total number of crimes but fewer serious offences.

The third issue is to do with *street time*. For any accurate estimates of crime, studies should calculate the "time at risk", "street time", or "free time" and correct for any time spent in custody. For most research using official data a common unspoken assumption is that the offender is completely free to commit crimes at all times. However, some researchers have recognised the importance of this problem (Weis, 1986; Francis et al., 2007; Liu et al., 2008) and have attempted to correct for it. The information of time at arrest and time at release is technically available from police records and prison records, but information on parole or remand is very hard to obtain. Also, this information needs to be matched with either official conviction data or self-reports. However, in practice this requires access to different authorities, and complex matching tasks, as name or date of birth may disagree across sources, for instance.

The last issue is to do with whether data from official records has *sociological relevance*. There are debates on the uses of official statistics from the conventional sociological point of view. Merton (1957, Chap. 4) argues against the use of official statistics on studying deviant behavior, since official statistics are not appropriately designed to represent his need for data to be 'sociologically relevant'. Merton states that "Before social facts can be explained it is advisable to ensure that they actually are fact". His view is that categories of deviance such as theft, criminal damage etc are set up to suit administrators and legal officials, and not sociologists. However, Kitsuse and Cicourel (1963) expand on this view, stating that official data can be viewed as telling the researcher about organizational processes as much as social process.

In summary, criminal careers research may benefit from either official records, or self-reports, or even the combination of both, and will depend on the purpose of the study. In addition, the researcher should always bear in mind the limitations of using either type of data source on studying offending, and interpret any results with great caution.

3.2 Datasets on offending in England and Wales

There are three major longitudinal studies on offending behaviour in England and Wales, which are the *Offending, Crime and Justice Survey*, the *Police National Computer*, and the *Offenders Index*. In the following sections, a description of each study in the above is provided. Then the reason why the Offenders Index is used in this thesis on the study of escalation in criminal careers is summarised.

3.2.1 Offending, Crime and Justice Survey

The Offending, Crime and Justice Survey (OCJS) [1] is the first national longitudinal, self-report offending survey from England and Wales, and is commissioned by the Home Office for the study of attitudes towards and experiences of offending.

The purposes of this survey are to measure prevalence of offending, to examine anti-social behaviour and drug use among the household population, in particular among young people aged from 10 to 25 in England and Wales. This survey provides information on the prevalence of offending (eg. fraud and technology crime); the prevalence and frequency of drug/alcohol use, victimisation, anti-social behaviour (covers non-criminal acts that cause offence); motivations and consequences of offending.

The survey has completed four annual sweeps (2003, 2004, 2005 and 2006). The first sweep of the OCJS in 2003 collected information from around 12,000 people aged from 10 to 65 living in private households in England and Wales. The subsequent annual sweeps from 2004 to 2006 were limited to survey young people aged from 10 to 25. For those young people who have previously been interviewed and have agreed to further contact, there is a follow-up and re-interview. In addition a 'fresh sample' also randomly drawn from the population at each subsequent

[1] Information is available at Economic and Social Data Service (ESDS): http://www.esds.ac.uk/findingData/snDescription.asp?sn=6345#doc.

sweep to ensure a cross-sectional representative sample of young people.

3.2.2 Police National Computer

The Police National Computer (PNC) is the main operational system for criminal records for police in Scotland, England and Wales and is a source of official records on offending [2]. There are two computerised systems, the *operational* PNC used by the police for operational tasks, and the *Ministry of Justice* (MOJ) research download. They are updated together, but the MOJ data contains fewer fields of information (Francis et al., 2002; Francis and Crosland, 2002).

The MOJ research download of the PNC is commonly used by researchers to study offending. It contains information on all proven offending for offenders in England, Wales and Scotland but not Northern Ireland. This includes cautions, warnings etc where guilt has been admitted as well as court convictions. It contains details of date of offence, date of charge, date of conviction, the offence details, court and police authority codes, and disposals for the offence. It can be searched by name or by other characteristics. However, it does not contain personal information apart from ethnicity and gender, and does not contain information on crimes yet to be proven. It has no information on immigration, or emigration. Hence an individual might have left the country, but this would be viewed as a period of not offending. Those who have died may not have been removed from the database. The convictions are added to existing convictions by confirming identity through fingerprints. Therefore, changes of names, address, or genders (rarely) should not result in mismatched records. There is an issue of weeding which a weeding policy is (or was) in force whereby minor records were erased. It is unclear how widespread this was.

The MOJ database was started in 1996, and before that time there was a paper system run by the Criminal Records Office (CRO). If an offender offends after 1996, then the offender's entire criminal history is back-record, converted and

[2]http://www.npia.police.uk/en/10508.htm (NPIA, 2011)

placed on the PNC, but the earlier conviction history is unknown to the PNC.

3.2.3 Offenders Index

The Home Office Offenders Index (OI) is an official records dataset and is provided by the Home Office for the purpose of research. The OI contains details of all Standard List offences convicted in a crown court or magistrates court in England and Wales from 1963 onwards. The complete data set consists of over 6 million offenders and is not publicly available. However, the Offenders Index Cohort sample is publicly released[3] and consists of an approximate 1 in 13 sample of all offenders born in four selected weeks (one week was selected in each quarter) in each of the years 1953, 1958, 1963, 1968, 1973 and 1978. The cohort data provides information on details of court appearances and convictions at each appearance for individual anonymised offenders. Dates of appearance, date of birth, gender, offence and disposals are recorded on the file. To prevent identification of individual offenders, information such as geographical markers, police force, and Criminal Record Office number have been deleted in the publicly available version.

The OI has no information on death, or immigration, or emigration. In contrast with the PNC, there is no weeding and no removal of people who have died in the OI. The OI dataset is formed by record matching, taking court records and matching them on name and data of birth to form criminal histories. Although this procedure compares well with police records (Francis et al., 2002) it can introduce inaccuracies, particularly for females. The OI does not contain all offences, but only standard list offences – minor offending such as speeding and public order offences are omitted. A particular issue relevant to long follow-up time studies is that the definition of standard list offences changes slightly over time. For example, new offences are passed into law, or become viewed as more or less serious. This can be dealt with by removing all offences which become standard list or stop being standard list over the period.

[3]The data set is SN 3935 at the UK data archive (http://www.esds.ac.uk/findingData/snDescription.asp?sn=3935)

3.2.4 Discussion on the use of the three data sources

Three main data sources in England and Wales have been introduced previously. The Offending, Crime and Justice Survey is a dataset of self-report offending, on the other hand, the Ministry of Justice research download of the Police National Computer and the Offenders Index are both sources of official records either recorded by the police or by the courts. As mentioned in the perivous section, the advantages and disadvantages of the use of either official records or self-reports primarily depend on the motivation for each research.

The focus of this thesis is on studying escalation in crime seriousness. Therefore, offenders are needed to be followed over a relatively long period of time in order to follow up their criminal activities. Importantly, this thesis is also interested in the pathway of escalation from less serious crimes into more serious crime such as murder. Therefore, OCJS is not suitable for such purpose, since it picks up relatively few serious offenders. Moreover, OCJS only consists of 4 years follow-up time. Compared to the computerised PNC (which started in 1996) and the Offenders Index, which traces back offending to 1963. Therefore, the OI provides a very long follow-up time since 1963 (compared to the PNC and the OCJS) and contains a large number of conviction occasions from England and Wales. Then it is possible to study the pathway of offenders' criminal activity, in particular escalation into more serious crimes through both cross-sectional and longitudinal analysis. Therefore, the datasets will use for all analyses are generated from the Home Office Offenders Index (OI) and will be described in the coming section.

3.3 Datasets are used in this thesis

Datasets for all three studies in the following chapters are extracted from the Offenders Index, but according to the purposes of each research topic each data were constructed from different subsets of the OI data. Therefore, the dataset used for each of the three studies is described in the following.

3.3.1 Study 1

The aim of this study is to explore the interrelationships between four types of serious crimes – arson, blackmail, kidnapping, and threats to kill and the likelihood of escalation to subsequent homicide (murder or manslaughter). Therefore, for this study, a special run was downloaded from the OI which allowed the development of four separate datasets for the four types of offenders. Each dataset contained those offenders with conviction histories of arson, blackmail, kidnapping, or threats to kill respectively between 1963 and 2001. For instance, the arson dataset contains complete criminal histories of offenders who had a conviction for arson in this period of time.

The datasets were reduced further in two stages. Kidnapping was only coded as a separate offence in the Offenders Index in 1979 – previously, it was included in the 'Others' category. Therefore, in the first stage, it needs to ensure that offenders were convicted of their first focus offence (one of the four types above) between 1979 and 2001. This procedure was possible for the offences of arson, blackmail, and threats to kill, but not for kidnapping. In theory, the kidnapping dataset could include persons who were convicted of kidnapping prior to 1979. However, those offenders cannot be identified from this data. Thus, those offenders who were convicted of any of these serious focus crimes of arson, blackmail or threats to kill, or of homicide between 1963 and 1978 (inclusive) were discarded. For example, a person who had been convicted of arson *for the first time* in 1977, and then convicted of blackmail in 1980 would be removed from both the arson dataset and the blackmail dataset.

Then a second restriction was that all those known to have been convicted of homicide prior to or as a co-conviction to the target conviction were eliminated from the relevant dataset. The reason for this additional restriction is that 'Study 1' is interested in first-time subsequent homicide following a particular serious offence. As it will become clear, the purpose of the study was to identify whether a conviction for one of the first time serious offences with no prior history of

Table 3.1: Follow-up periods and total number of offenders for subset of each of the Offenders Index cohort (for an explanation of the codes see text).

Dataset		No. of offenders (%)	Total
	A (1)	41,375 (90.1%)	
Arson	A (2+)	3,680 (8.0%)	45,915
	A + others	860 (1.9%)	
	B (1)	4,905 (84.9%)	
Blackmail	B (2+)	242 (4.2%)	5,774
	B + others	627 (10.9%)	
	K (1)	6,097 (83.6%)	
Kidnapping	K (2+)	218 (3.0%)	7,291
	K + others	976 (13.4%)	
	Th (1)	8,438 (86.0%)	
Threats to kill	Th (2+)	408 (4.2%)	
	Th + others	970 (9.8%)	9,816
	Combined		67,052

homicide increases the risk of first time homicide compared to those without such a conviction.

Table 3.1 [4] shows the total number of offenders in each of the resulting four datasets, which are arson $n = 45,915$, blackmail $n = 5,774$, kidnapping $n = 7,291$ and threats to kill $n = 9,816$. As they are separate datasets, a person could contribute to more than one dataset. In fact, there is overlap between the datasets; of the $67,052$ persons in the four datasets (eliminating overlaps), $1,689$ persons (2.5%) were in two or more datasets. An example of overlapping is illustrated by using the arson dataset. The label of 'A (1)' refers those offenders with only one conviction occasion[5] of arson ($41,375$ offenders) but not involved with any of the three other focus offences in their complete criminal histories. The label of 'A (2+)' refers offenders (8.0%) who had more than 1 conviction occasions of arson but not three other types. Finally, there are 1.9% of offenders in the arson dataset which are labeled as 'A + others' who had either co-conviction[6] between arson and

[4]The numbers of offenders in each subset data in Table 3.1 are slight different than the original paper by Soothill et al. (2008) due to computing errors.

[5]A conviction occasion is a court appearance where the offender was found guilty of one or more offences.

[6]A co-conviction is a conviction with multiple offences of interest at the same conviction

any other three focus offence types, or subsequent conviction for any other threes following an arson conviction.

One thing to point out is that the majority of offenders are more specialised with one type of conviction on offences of interest. For example, only 1.9% of arson offenders had a conviction for arson and one of the other three types of focus offence. The more detailed descriptions are in the following chapter which describes the 'study 1'. This table, also shows the existence of a combined single dataset from the four datasets ($n = 67,052$). Both the combined dataset and the four separate datasets are used in the analysis.

3.3.2 Study 2 and Study 3

The remaining two studies are on a general study of escalation in crime seriousness between offenders and within offenders over conviction occasions. 'Study 2' and 'Study 3' are two sequenced studies in terms of the development of statistical methodologies. To provide the opportunity to follow up a set of individuals' criminal careers for the longest time period, the 1953 birth Offenders Index cohort is used and followed through to 1999 (Prime et al., 2001).

From the 11,068 offenders in the initial dataset, $5,711$ (56%) offenders who had only a single conviction occasion were removed – escalation cannot be measured for these with only one conviction occasion. An additional 91 offenders whose first conviction was made after age 37 were also removed in order not to bias the analysis with a relatively small number of older criminal starters who had short follow-up time. Then a further 435 invalid offenders were discarded as unable to match each offence type in their criminal history to a seriousness score. The reason for this was a mismatching problem due to Home Office offence coding changes over time. The seriousness score was developed in 2005 based on all convictions on the PNC in a three month period. Thus the scores were based on offences as they existed in 2005. Old offence codes which did not exist in 2005 thus presented a problem.

occasion.

Table 3.2: The number of offenders in each subset of the data in terms of number of convictions from 1963 to 1999.

No. of convictions	Male	Female	Total offenders Frequency	%
2 convictions	1,368	292	1,660	34%
3 convictions	736	88	824	17%
4 convictions	493	62	555	12%
5 convictions	350	23	373	8%
6 convictions	218	23	241	5%
7 convictions	187	14	201	4%
8+ convictions	936	41	977	20%
Total (%)	4,288 (89%)	543 (11%)	4,831	100%

For example, some old offence types may be merged into new offence categories, or have disappeared from the Offenders Index Code book (detailed methods on the measurement of crime seriousness see Chapter 5 section 5.2.1). Therefore, the final dataset consists of $4,831$ [7] offenders with $4,288$ males (89%) and 543 females (11%).

Therefore, this dataset allows us to observe each offender's criminal activities with a follow up time of at least 10 years and at most 36 years, and with at least two conviction occasions. The age of offenders at conviction ranges from age 10 (the age of criminal responsibility in England and Wales) to age 46.

The distribution of the number of distinct convictions (over the period from 1963 to 1999) for the offenders in the dataset is shown in Table 3.2. Unsurprisingly, the number of offenders declines as the number of convictions increases. In addition, the proportion of male offenders increases with the total number of convictions. Above eight convictions, numbers become small, so a combined number of all such offenders is illustrated into a single group. The maximum number of convictions for an offender was 137.

[7] The total valid number of offenders is slightly different to the published paper (Liu et al., 2011) due to a better matching of the old offence codes with the current Home Office Standard List offences codings (2005). This thesis uses the improved matching throughout.

Table 3.3: A list of information from OI is used in each of the three studies in this thesis for studying escalation in crime seriousness.

Variables	Study 1	Study 2	Study 3
Age	✓	✓	✓
Gender	✓	✓	✓
Conviction date	✓	✓	✓
Offence type	✓		
Length of sentence	✓		✓
Other disposals	✓		✓
Prior criminal history	✓		

3.3.3 List of variables

As mentioned before, the Offenders Index can only provide certain items of information. In general, the dataset is limited to criminal history variables. Table 3.3 summarises the information available from the OI data, which can be simply used or are needed to develop new variables for the research purposes of each study. The more detailed description of variables which were extracted from the OI for each of the three studies will be introduced in the relevant chapter.

3.4 Summary

This chapter has been focused firstly on two types of data sources – official records and self-reports which are used in the research of offending. Then three available longitudinal datasets for studying offending in England and Wales were described as well as the reason for choosing the Offenders Index as an empirical dataset for studies of escalation in this thesis.

However, in this chapter the limitations of using the OI data was also recognised, such as an underestimation of offending, especially minor crimes; lack of information on the offenders' social background – social factors for escalation. However, the OI contains a large number of offenders from England and Wales and with conviction occasions from minor to serious offences, and follows them up for a very long period of time (since 1963).

The next chapter will be introduce one of the three major studies on escalation in crime seriousness – 'Study 1'. 'Study 1', in particular, will be focusing on interrelationships among serious crimes – arson, blackmail, kidnapping and threats to kill and escalation from such crimes to even more serious crimes – homicide.

Chapter 4

Study 1: Escalation to Homicide after Serious Crime – which patterns and sequences are important?

4.1 Introduction

The earlier chapters have introduced the background to the study of escalation in crime seriousness. This study is now introducing three new research studies on escalation. The first study, which is described in this chapter, will look at the concept of "serious offender escalation". More precisely, four specific serious crimes will be examined which are arson, blackmail, kidnapping, and threats to kill and escalation from these four crimes into the most serious crime – homicide (murder and manslaughter) will be examined. The remaining two studies will look at general escalation (Chapter 6: Study 2 and Chapter 7: Study 3).

In the criminological literature, the common methodology for studying escalation is to study the existence of a "gateway" offence to more serious crimes. Thus, within sexual offending there is the idea that some offences, such as indecent exposure or burglary, can be gateway offences to more serious sexual offending. In drugs offending, there is some evidence that police consider cannabis possession to be a gateway offence to other more serious drugs offending, and so they will be less likely to ignore such possession (Warburton et al., 2005). Therefore, in particular, research in the past decade has shown increased interest in gateway offending. A review study by Sample and Bray (2003) discussed two respects in terms of sexual offending: one is sex offenders' recidivism rates in comparison of nonsexual offenders, the other one is "gateway" offences to sexual offending. They conclude that there is no clear evidence to support either assumption of unusually high rates of recidivism among sex offenders, or some crime types, such as burglary, are "gateway" offences to sexual offending. For example, in relation to studies on recidivism rates among sexual offenders, few studies suggested that lower rates of reoffending among sexual offenders than nonsexual offenders (Langan and Levin, 2002; Hanson et al., 1995; Sapsford, 1998; Sipe et al., 1998); other study such as Blumstein et al. (1988) showed weak evidence of escalation following arrests for aggravated assault, especially for white offenders.

The present study will examine escalation in crime seriousness in low-frequency but high-tariff offences, since there has been much less criminological focus on such crime types. The concept of *serious offender escalation* refers to escalation for an offender who has committed a serious or dangerous offence to an even more serious offence. In this study, therefore, arson, blackmail, kidnapping, and threats to kill are taken as the *focus offences* and the risk of escalation from the four focus offences to homicide is assessed. In other words, this work is to identify whether arson, blackmail, threats to kill and kidnapping are gateway offences to homicide, either considered on their own or together in some combination. Therefore, the motivation of this work is to identify whether there are "gateway" offences for homicide and whether those committing such serious offences need to be monitored.

The structure of this chapter is: this section firstly reviews two studies which attempted to examine the inter-relationships among serious crimes, and sequencing of serious crime convictions; secondly providing the purpose of this study in more details. Section 4.2 describes statistical methods which are used to assess escalation among the four specific types of serious offenders, including both descriptive statistical analysis and survival analysis. Survival analysis is useful to examine the time from the first focus crime to subsequent homicide conviction. A detailed statistical definition and description of the usage of survival analysis are provided in section 4.2. Results from both the descriptive analyses and the regression approach are shown in Section 4.3. In the end, results from previous analysis are concluded and potential applications for policy making are considered.

4.1.1 Literature review

In both qualitative and quantitative criminological research, it seems that little work has been done in the area of interrelationship and sequencing of serious crime convictions as predictors for later more serious offending. One exception is the study by Soothill et al. (2002). Using a matched case-control study of murderers, and examining the prior criminal conviction careers of the murderers and the non-murderer controls, they found that 'a previous conviction for kidnapping was shown to be a statistically significant risk factor for murder, when compared against general criminal controls and against violent controls' (Soothill et al., 2002, page. 33). Similarly, for blackmail, 'those with a blackmail conviction were over five times as likely to become murderers as the general controls' (Soothill et al., 2002, page. 34).

More recently, Liu et al. (2008) examined the time from the first conviction for kidnapping to some specific subsequent serious crimes: a subsequent kidnapping, murder, manslaughter and rape of a female. Using survival analysis procedures, this work estimated that five out of every 100 kidnap offenders who were convicted of kidnapping will be reconvicted for this offence. In contrast, one in every 100 kidnap offenders will be convicted of homicide after 20 years and close to two out of every 100 will be convicted of rape of a female in 20 years. It was further demonstrated that kidnappers are over 30 times more likely than males in the general population to be convicted of homicide and four times more likely than sex offenders.

Further, it seemed from the earlier research (Soothill et al., 2002) that those who were subsequently convicted of murder following a kidnapping conviction seemed to be much closer in time to the earlier kidnapping conviction compared, for instance, to those who were convicted of blackmail, where the subsequent killing often seemed to be more distant in time.

In brief, earlier work certainly suggests that those who were convicted of kidnapping or blackmail are at greater risk of being subsequently convicted of murder, but they fall short of comparing systematically various kinds of serious offences and probing the possible interrelationships between them.

4.1.2 Purpose of this study

As little work has been done in comparing various kinds of serious offences on escalation in offending in the literature, in this study, there are four types of offences which arson, blackmail, kidnapping, and threats to kill are considered. Additionally, the four types of offences are referred as the *focus offences* in the rest of this work. The purpose of this study is to compare and assess possible interrelationships among the four focus offences in terms of the risk of escalation into the most serious crime – homicide (murder or manslaughter).

According to the introduction and literature in this section earlier, there are two *questions* needed to be answered in this study. (1) For each of the four focus offences what *proportions* go on to be reconvicted for the same offence, or get convicted for one of the other three offences, or for homicide (murder or manslaughter)? Essentially, this question is focusing on offence *specialisation* within each offence, and among the four focus offences. It is also concerned with *escalation* - how many of those offenders escalate their criminal activity to homicide? (2) Are certain combinations and sequences of convictions on the four focus offences risk factors for subsequent homicide (escalation in crime seriousness)? This question focuses more on the mix of serious offending and the diversity of serious offending as a predictor of homicide.

Before shifting attention to the actual methodology of assessing the two questions, the reason why the four particular offences are chosen needs to be explained. The choice of these four offences was determined by three factors. First, earlier studies (Soothill et al., 2002; Liu et al., 2008) identified that among general offenders with a previous criminal history or history of serious crimes, such offenders displayed an increased risk of conviction of murder. Second, these four offences are seemingly very different types of serious offences. *Kidnapping* seems essentially a 'hands-on' potentially violent offence with face-to-face interaction; *blackmail* seems a more 'hands-off' offence with little or no face-to-face interaction; *threats to kill* seems a more verbal or written type of aggression that may or may not involve face-to-face interaction; and *arson* is a serious property crime that may or may not endanger human life. The final factor is the availability of the data for these four offences.

4.2 Methods in survival analysis

The investigation on the two previous questions is explored in two ways. The first is using descriptive statistics including basic cross tabulation, and a plot of Kaplan-Meier survival curves on subsequent convictions of certain serious crimes following the four focused offences. The second way is using a survival regression approach – the Cox proportional-hazards model which various risk factors, follow-up time and time at risk are considered. Following, a few fundamental terminologies in *survival analysis* are needed to be defined first. These terminologies include observation time, failure time, censoring, survival function, and hazard function. Firstly, the definition of those fundamental concepts are described. Then the survival function and hazard function are defined for the use of Kaplan-Meier survival curve and

Cox proportional-hazards model to study escalation from the four focus offences to homicide.

4.2.1 Basic concepts

For each of the four datasets (see Chapter 3.3.1), the target conviction for an offender is defined to be the first conviction of the offence of interest; thus, the target conviction for an individual in the blackmail dataset, for instance, would be the first blackmail conviction. However, for the combined dataset, the target conviction is the very last conviction of any of the four serious focus offences prior to a homicide conviction. The reason is that the combined single dataset is used for examining the time from one last of the focus offence to a conviction of homicide.

Therefore, this study focuses on examining the time from the target conviction until the event is of interest – the first subsequent conviction of one of the four offences or of homicide (murder and manslaughter). In survival analysis, the response variable is the time until that event and is often called a *failure time*, *survival time*, or *event time* (Harrell, 2001). If there is no successive conviction occurred (failure-free) until the end of follow-up period under the investigation (12/31/2001), it is called *censoring*. The two concepts of failure and censoring, and one potential problem relates to time at risk (street time) are explained through Figure 4.1.

Firstly, the concepts of failure and failure time are explained through four artificial offenders which are illustrated in Figure 4.1. Plot (a) shows these four offenders in real time (year at conviction) from 1963 to 2001. The starting point of each line is the target conviction year of each offender. Then the sign of '$\times$' indicates the first subsequent conviction occurred, which is referred as an event of *failure* in survival data analysis. The failure can be a reconviction on one of the four focus offences or a subsequent conviction of homicide according to the purpose of each analysis. Offender A, for instance, had a first-time arson conviction in 1964 and had a reconviction for arson in 1979 (in 15 years time). In contrast, the

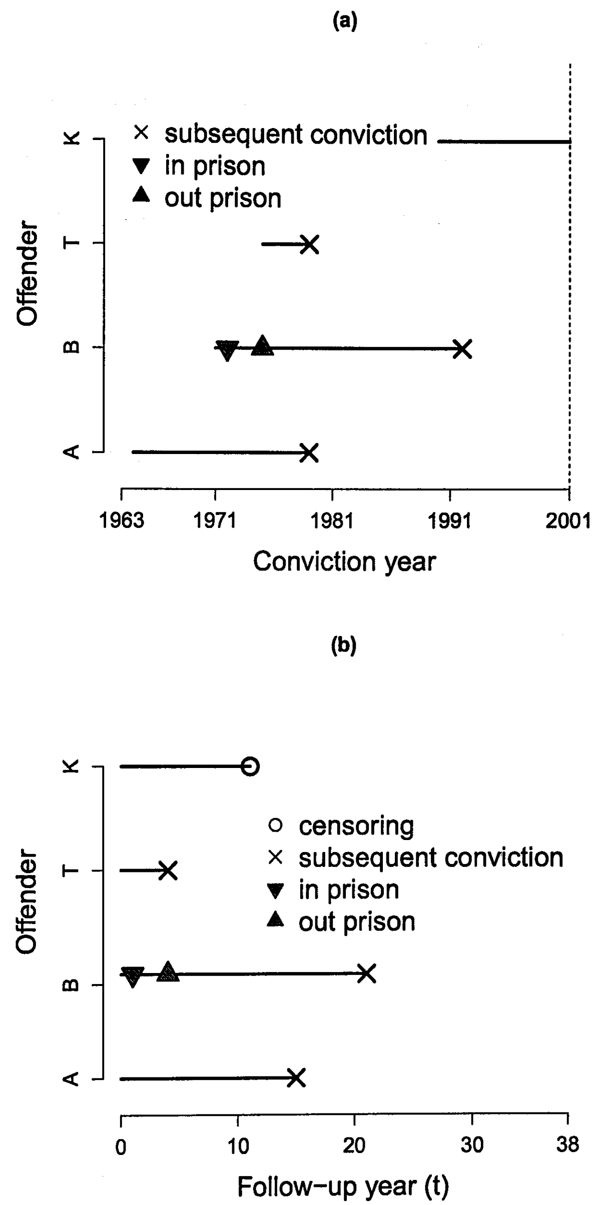

Figure 4.1: Four artificial examples of offenders' criminal activity from 1963 to 2001. (a): Observed on real time scale; (b): Follow-up time from target conviction.

offender T was convicted of threats to kill firstly in year 1975, then within a very short follow-up (4 years) had a subsequent conviction of homicide in 1979.

Therefore, the time between the target conviction and failure is defined as the *failure time.* More specifically, failure time refers to the time from the target conviction to the first failure but not further sequence of failures. For example, there could be more than one arson reconviction following the first-time arson conviction. However, only the time to the first arson reconviction is considered. Therefore, this offender after the very first arson reconviction, s/he by assumption is not in the population of risk of being reconvicted of arson. This assumption is essential for definition of survival function and calculation of the Kaplan-Meier estimator later in this section.

Secondly, the concepts of observation time and censoring are introduced in the following. In Figure 4.1 Plot (b), the same four offenders which are shown in Plot (a) are observed from the time when their target conviction occurred, say $t = 0$. Then their criminal activity is observed until either the event of failure, or until the end of this study (year 2001) if they are failure-free. The *observation time* t hence can take values from 0 to $max(t) = 38$ (if the target conviction in 1963 and follow up until 2001). The observation time is used as primary time scale in the survival analysis.

Then the concept of *censoring* refers to an individual who is observed, but failure-free under the period investigation – also named as right censoring. Offender K, for instance, is censored at time $t = 11$ with a symbol of circle at the end of this line, since there is no event of failure observed after the first-time conviction, up until the end of the observation period at $t = 11$.

In this study, there are three possible sources of censoring. The first is offenders who have been given a very long sentence or a life sentence for their convictions on serious offences. Therefore, they are unable to be convicted of any crimes before the end of the study. The second is where there are no observed convictions until the end of study period, either desistance from crimes or no official records of their crimes. The final source of censoring is death or emigration of offenders, but there are no such information available in the Offenders Index (OI) dataset (detailed description of OI see Chapter 3 section 3.2.3). Therefore, the point of censoring in this study is known and fixed at 31/12/2001.

Finally, one potential criticism of this study pertains to time at risk, alterative names are "street times", or time at large. An example shows in Figure 4.1, offender B was given a four-year sentence after the first conviction for blackmail in year 1971, then during 1972 to 1975 this offender was in prison. Therefore, this four year period is a period of no risk of offending. In theory, this period of imprisonment should be deducted from the total failure time for this offender, where the interest of the study is in time at risk rather than calendar time.

Adjustment for time spent in custody is not straightforward. The reason is although a record of sentence awarded is available in the OI dataset, which includes the length of any custodial sentence, there is no indication of the actual time served for each conviction. Again, certain assumptions must be made. For the target conviction and for every conviction after that date that had a custodial sentence, the accurate time spent in prison should be the sum of these times spent in custody. However, there are various unpredictable circumstances that can change the imprisonment time, such as whether the offenders will get remission or not, whether they get parole, and whether it is at the first, second or later opportunity. On account of that fact, an estimation of the time served as some fraction of the total sentence length awarded, is taken to be 0.3 in this study [1].

By making this estimate for time at risk, there will be some individuals who are convicted of a homicide while the estimate says that they will still be serving

[1] The estimate of the time from conviction to release being around 30 per cent of sentence awarded was made as follows. 1995 Prison Statistics in England and Wales (Home Office 1996) contain an estimate that, for adult males, between 40 and 50 per cent of sentence awarded is actually served (Home Office 1996 : Table 4.14). However, this includes time spent on remand before conviction. Forty-eight per cent of the prison population spent time on remand, with an average of around 60 days increasing to around a year for some cases (Home Office 1996: Chapter 2). This gave a reduction of between 10 and 15 per cent, depending on sentence awarded, giving the final result of 30 per cent. Data for other years are similar and 1995 figures are representative of the period under study.

time in custody. Such convictions are considered as pseudo-reconvictions. They are assumed not to be true reconvictions, but will relate to offences committed before the appropriate conviction, but discovered or admitted to later when the offender is in custody.

4.2.2 Notation for survival analysis

This section firstly defines the *survival function* in survival analysis, then defines the concept of *hazard function*. These two functions are fundamental for understanding survival analysis, and further regression approaches can be applied based on them.

Let T denote the response variable, which is the time until an event and is regarded as a random variable. The failure time t_i for each offender i ($i = 1, .., n$) is one random realisation from the distribution of T.

Then, the cumulative distribution function of T is given as:

$$F(t) = Pr(T \leq t), \tag{4.1}$$

which describes the cumulative probabilities of offenders' survival time T which are smaller or equal to a value of time t. In other words, the probability of all those offenders having an event of failure (a subsequent conviction occurs) before or at time t.

Therefore, the survival function is then defined as:

$$S(t) = Pr(T > t) = 1 - F(t). \tag{4.2}$$

It captures the probability that offenders will survive beyond a specific time point t. The value of $S(t)$ is always 1 at $t = 0$. The random variable T denotes a random failure time from the survival distribution $S(t)$. Then additional notation is needed to define for the response if censoring occurs for the ith subject. Let Y_i denote the response for the ith subject; T_i is the survival time for the ith subject;

and C_i denote the censoring time for the offender i, then the censoring indicator is defined as:

$$\begin{aligned} e_i &= 1 \text{ if the event was observed } (T_i \leq C_i), \\ &= 0 \text{ if the event was censored } (T_i > C_i). \end{aligned} \tag{4.3}$$

The observed response is

$$Y_i = min(T_i, C_i), \tag{4.4}$$

which is the time that occurred first, the failure time or the censoring time. The response is hence a pair of values (Y_i, e_i) in the survival analysis.

Secondly, the concept of *hazard function* needs to be defined. It is another representation of the distribution of survival times – $S(t)$. The probability density function is:

$$f(t) = Pr(T = t) = \frac{\partial F(t)}{\partial t}, \tag{4.5}$$

which the probability of failure at time t. Then the *hazard function*, which also called the *force of mortality*, or the *instantaneous event (failure) rate*, and is given by:

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{Pr[(t \leq T < t + \Delta t)|T \geq t]}{\Delta t} \\ &= \frac{f(t)}{S(t)}. \end{aligned} \tag{4.6}$$

Therefore the hazard at time t is related to the probability that the event will occur in a small interval around t, given that the event has not occurred before time t. Then the cumulative hazard function is:

$$H(t) = \int_0^t h(u)du. \tag{4.7}$$

The cumulative hazard function describes the accumulated risk up until time t, and is the negative of the log of the survival function. $H(t)$ is nondecreasing as t increases. The survival function $F(t)$ can be represented by the cumulative hazard function, and is:

$$S(t) = \exp\{-H(t)\}. \tag{4.8}$$

4.2.3 Kaplan-Meier estimator and the Cox model

Two approaches in survival analysis are used for assessing escalation from the four serious offences to subsequent conviction for homicide. The first approach is known as the Kaplan-Meier estimator – also known as Kaplan-Meier survival curves – which is essentially a descriptive statistical approach (non-parametric). The second approach is the Cox proportional-hazards model (semi-parametric) which is used to estimate the survival function and also to control various risk factors, such as the type of first-time conviction, gender, and age at conviction. In the following development, both the Kaplan-Meier estimator and the Cox proportional-hazards model are defined respectively.

Firstly, let $n(t)$ be the number of offenders who are at risk of a first homicide conviction subsequent to a conviction for one of the four focus offences at time $t = k$, and range of k can take from 0 (the date of the first focus offence) to K (the end of study); and $d(t)$ is number of those who have subsequent convictions on homicide (failure) at time $t = k$. The Kaplan-Meier estimator is defined as:

$$\hat{S}(t) = \prod_{t:0 \leq K} (1 - \frac{d(t)}{n(t)}). \tag{4.9}$$

Basically, the Kaplan-Meier estimator is a descriptive survival function, which is the product of survival rates at each time t. Therefore, the plot of the Kaplan-Meier estimator against time $t = 0, .., K$ will illustrate the speed of subsequent conviction for each of the four focus offences. More formal testing of the difference between different types of serious offences can be obtained through the Cox proportional-hazards model which is described below.

The Cox proportional-hazards regression model was introduced in a seminal paper by Cox (1972). This approach has been widely used and broadly applicable in the survival analysis. All the analyses in the later section used the *survival* library in R.

Modelling survival data usually employs the hazard function or the log hazard. For example, assuming a constant hazard, $h(t) = \nu$, implies an exponential distribution of survival times. The *Cox* model, in contrast, defines the log hazard function as following:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}, \tag{4.10}$$

or equivalently,

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \cdots + \beta_k x_{ik}). \tag{4.11}$$

The $h_0(t)$ is the baseline hazard function, which is the hazard when all of the x's are zero. The function of $\alpha(t)$ thus represents a log-baseline hazard. The notation of this model may vary across different text books or articles, such as the classic text book on survival analysis by Cox and Oakes (1984) and a more recent book by Therneau and Grambsch (2000) – the first author of which is also the author of the *survival* library in R.

This model is *semi-parametric* because the baseline hazard is left unspecified. One feature of this model is that the hazard ratio from two observations is independent of time t. For example, two observations i (male) and j (female) that their hazard ratio for the two observations is:

$$\begin{aligned}\frac{h_i(t)}{h_j(t)} &= \frac{h_0(t)e^{\beta_1 x_{i1}}}{h_0(t)e^{\beta_1 x_{j1}}} \\ &= \frac{e^{\beta_1 x_{i1}}}{e^{\beta_1 x_{j1}}}\end{aligned} \tag{4.12}$$

Therefore, the Cox model is a proportional-hazards model.

In the next section, both descriptive statistical analysis and survival analysis were applied to assess escalation from the four focus offences to homicide. There were four separate datasets, which contained offenders' complete conviction histories for arson offenders, blackmail offenders, kidnapping offenders, or threats to kill offenders from 1979 to 2001 from England and Wales. In a further analysis, also a merged dataset (combined four datasets) was used in order to carry out a complete analysis on time to homicide from any of the four focus offences, where as also to asses the number and sequencing of the focus offences as additional risk factors. The detailed description of the data is provided in Chapter 3 section 3.3.1.

4.3 Results

4.3.1 Descriptive statistics

Using the four separate datasets (details see Chapter 3 section 3.3.1) constructed as the first stage, Table 4.1 shows the number of offenders with one or more subsequent convictions of a given type: arson, blackmail, kidnapping, threats to kill, or homicide. Research questions, which relate to proportions, specialisation and the likelihood of a subsequent homicide conviction, have some preliminary answers:

(1) *Proportions.* For each of the four separate offences (or datasets), an important minority go on to be convicted of at least one or more of the five serious offences of interest. The cumulative numbers of persons who were subsequently convicted of at least one of the five serious offences for each of the four focus

Table 4.1: The number of offenders with subsequent convictions for serious offences following the target (first-time) conviction.

	No. of offenders	No. of offenders with subsequent convictions of given type: arson (%)	blackmail (%)	kidnapping (%)	threats (%)	homicide (%)	combined total[a]	Average follow-up time in years (SD)
Arson	45,915	3,819 (8.32)	79 (0.17)	160 (0.35)	280 (0.61)	241 (0.52)	4,394 (9.57)	12.89 (6.49)
Blackmail	5,774	52 (0.90)	279 (4.83)	65 (1.13)	51 (0.88)	34 (0.59)	438 (7.59)	12.40 (6.46)
Kidnapping	7,291	44 (0.60)	21 (0.29)	273 (3.74)	61 (0.84)	48 (0.66)	398 (5.46)	8.18 (5.61)
Threats to kill	9,816	110 (1.12)	27 (0.28)	63 (0.64)	473 (4.82)	60 (0.61)	682 (6.95)	9.07 (5.62)

[a] 'Combined total' refers to the total number of persons who were convicted subsequently of at least one of the five offences (arson, blackmail, kidnapping, threats to kill or homicide).

offences, showed in the penultimate column in Table 4.1. The range of its proportion is from 9.57% for those who were convicted of arson for the first time to 5.46% for those who were convicted of kidnapping for the first time. Moreover, for the first-time arson offenders their proportions of subsequent convictions for the other four types of serious offences are relatively smaller than the first-time blackmail, or kidnapping, or threats to kill offenders. For example, the proportion of subsequent blackmail following arson is 0.17%, in contrast, the proportions of subsequent blackmail following kidnapping and threats to kill are 0.29% and 0.28% respectively.

(2) *Specialised serious offending.* The shaded diagonal in Table 4.1 shows the numbers and proportions who are reconvicted of the same type of offence as their original target conviction. Table 4.1 demonstrates that, within the four focus offences, the offenders tend to specialise in the type of subsequent offence they commit, by being more likely to be reconvicted of the same offence than one of the other three focus offences. Arsonists seem the most specialised, with around one in 12 of those who are convicted of arson being later reconvicted for arson. For the other three offences (blackmail, kidnapping and threats to kill), around one in 25 are reconvicted for the same offence.

(3) *Escalation to homicide.* Homicide can be regarded as the most serious reconviction and the four focus offences have quite similar proportions (from 0.52% to 0.66%) of subsequent conviction for homicide. The figures are uncontrolled for the speed of subsequent conviction of homicide. However, the final column in Table 4.1 provides some extra information on the average follow-up time (in years), which is the averaged individuals' follow-up time from their *first-time* convictions on one of the four focus offences to the end of study. Arson and blackmail offenders, for instance, have longer average follow-up time of around 12 years. In contrast, threats to kill and kidnapping offenders have average follow-ups of nine and eight years, respectively. The explanations of the variation of follow-up times can be investigated through the distributions of the numbers of the first-time convictions

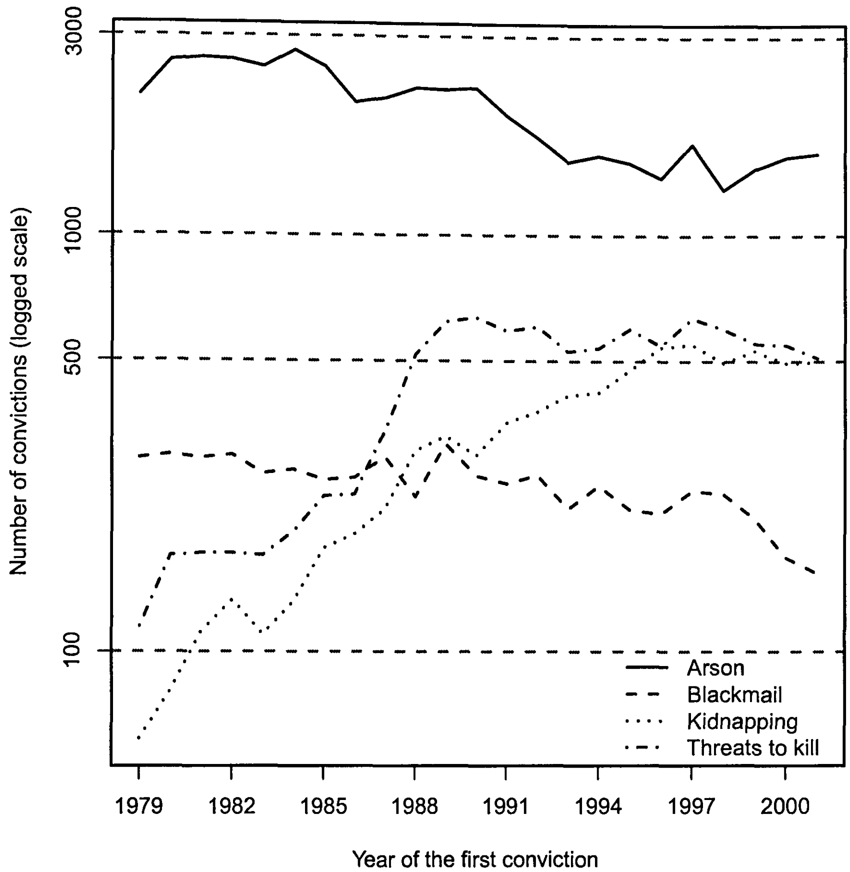

Figure 4.2: The numbers of first-time convictions by year for each of the four focus offences.

by year for each of the four focus offences.

Therefore, Figure 4.2 shows the number of the first-time convictions for the four offences during the period under investigation. Firstly, the number of first-time arson convictions are remarkably larger than the other three serious offences, that is why the number of convictions are scaled on a logarithmic axis in Figure 4.2. In proportion, the first-time arson and blackmail convictions have larger proportions at the earlier year (say before 1990) than the first-time kidnapping and threats to kill convictions. In other word, the areas under the two lines before 1990 are proportionally larger than the areas after year 1990. Therefore, on average, arson and blackmail offenders have longer follow-up time than kidnapping and threats to kill offenders.

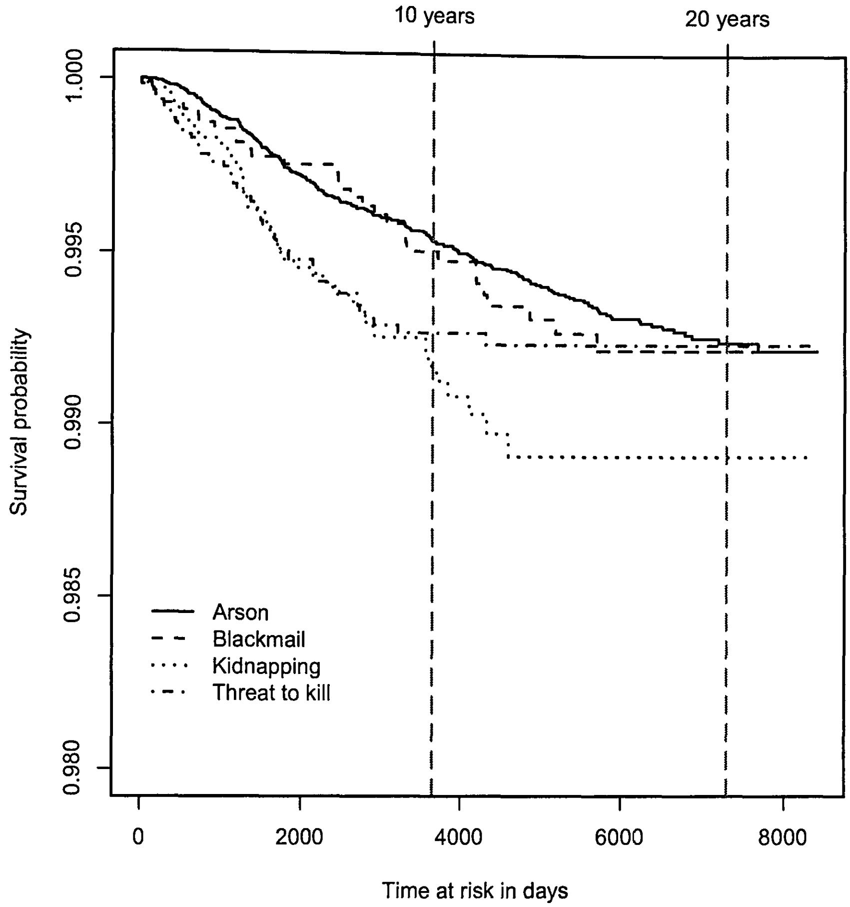

Figure 4.3: Kaplan-Meier survival curves for risk of homicide (adjusted for time at risk) following four serious crimes.

Secondly, the number of first-time kidnapping offenders increases over time and peaks at about 1998, so the averaged follow-up time of kidnapping (in Table 4.1) is even shorter (8.18 years) than threats to kill (9.07). The limitations of Table 4.1 and Figure 4.2 are that there is no control for length of follow-up time and time at risk (street time). These potential criticisms are answered by applying survival analysis through firstly plotting of Kaplan-Meier survival curves (nonparametric) and secondly using the Cox proportional-hazards regression model which is a semi-parametric approach (section 4.3.2).

According to the exploratory results showed in Table 4.1, the risk of subsequent homicide following one of the four target convictions might be expected to be similar. However, by looking at the Kaplan-Meier survival curves on risk of subsequent homicide conviction in Figure 4.3, it is clearly shown that the different offences have different trajectories. Firstly, at the 20-year point, arson, blackmail and threats to kill all eventually reach very similar estimated proportions, namely around 0.8%. However, kidnapping shows a different estimated rate of subsequent homicide conviction at 1% (roughly, one in a 100) in a 20-year period at risk.

Secondly, both kidnapping and threats to kill have a more rapid gradient in the first 10 years of the follow-up period on subsequent homicide conviction rate. After that time, they both flatten off; this is partially due to the lack of cases with more than ten years of follow-up in the previous analyses. In contrast, arson has a rather constant slope throughout the follow-up. Therefore, the gradients provide some clues as to whether the risk of homicide remains constant, falls or rises over periods of time. This feature has relevance for the length of supervision that may be required for different type of offences.

However, despite using the survival analysis technique and controlling for actual time at risk, there is still a concern about comparing the four datasets. After all, there are 1,689 persons who are in two or more of the datasets. Perhaps these offenders contribute disproportionately to the subsequent homicide conviction rate and may be the reason why the subsequent conviction rates proportionally for homicide are similar for the four offences.

Therefore, at stage two, this study is trying to answer whether being involved with more than one focus offence can increase the risk of escalation to homicide. For the purpose the combined dataset of 67,052 persons for whom the offence of arson, blackmail, kidnapping or threats to kill first occurred in 1979 or after is considered. In Table 4.2, shows the total number of offenders and number of offenders who are subsequently convicted of homicide for specific subsets of the Offenders Index cohort data. The subsets are defined by letters in the first column. Thus, the 'A' refers to arson convictions; the 'Th' refers to threats to kill convictions; the 'B' refers to blackmail convictions; and the 'K' refers to kidnapping convictions. The number '(1)' indicates only one conviction occasion occurred; and

'(2+)' two and more conviction occasions occurred. Then the '+' is for combining multiple types of convictions. For example, 'A (1)' stands for those offenders who only involved with one arson conviction but not the other three types, and 'A (2+)' stands for those arson offenders who have two or more conviction occasions on arson, but not on other three types. 'B + K' then means those offenders who have multiple conviction occasions on blackmail and kidnapping, or co-convictions of both.

Table 4.2 also presents the outcome in terms of subsequent homicide and reveals that the overall homicide rate is 0.52% (roughly one in 200). However, the important feature of Table 4.2 is that it shows the effects of (1) being convicted of one of these focus offences on two or more occasions (but without being convicted of one of the other serious offences) and of (2) being convicted of two or more different kinds of serious focus offences.

First, considering those who were convicted for one of these serious offences on two or more occasions, the patterns are different among the four offences. With arson and kidnapping, the chances of being convicted of homicide appear to be at least double when the person has two or more two occasions – A (2+) and K (2+) – compare with either an arson conviction ('A (1)') or a kidnapping conviction ('K (1)'). With the threats to kill, the proportion subsequently convicted for homicide actually declines from 0.52% to 0.25% when a person is convicted of one of these offences on two or more occasions. In contrast, the two proportions are similar (0.47% and 0.41%) for blackmail offenders.

Now moving on to the offenders who have been convicted between 1979 and 2001 of a mix of two or more of the four offences being considered. For every combination pair shown in Table 4.2, there is at least a doubling of the likelihood of being subsequently convicted of homicide compared with the figures when the offender has been convicted of just one of these serious offences. Moreover, the proportions of subsequent homicide convictions from those involved with multiple types of convictions are quite different among the four focus offences. For example,

Table 4.2: Total number of offenders and number of offenders who are subsequently convicted of homicide in each subset of the Offenders Index cohort (for an explanation of the codes see text).

Type of offender	No. of offenders	No. of subsequent homicides	%	Average follow-up time in years (SD)
A (1)	41,375	183	0.44	12.82 (6.52)
A (2+)	3,680	42	1.14	11.19 (6.08)
Th (1)	8,438	44	0.52	9.09 (5.64)
Th (2+)	408	1	0.25	7.34 (5.62)
B (1)	4,905	23	0.47	12.59 (6.43)
B (2+)	242	1	0.41	11.10 (6.48)
K (1)	6,097	29	0.48	8.11 (5.60)
K (2+)	218	2	0.92	6.46 (4.85)
A + Th	450	6	1.33	7.30 (4.90)
A + B	155	3	1.94	11.9 (5.43)
A + K	218	6	2.75	7.86 (5.00)
Th + B	101	2	1.98	8.64 (5.09)
Th + K	371	5	1.35	7.78 (5.49)
B + K	341	4	1.17	7.86 (5.94)
A + Th + B	7	0	0	7.94 (6.06)
A + Th + K	23	1	4.35	6.46 (4.09)
A + B + K	5	0	0	5.18 (3.72)
Th + B + K	16	1	6.25	8.19 (5.34)
A + Th + B + K	2	0	0	5.39 (4.35)
Total	67,052	353	0.52	-

there are a total of 241 arson offenders who had subsequent homicide convictions (this value can be computed from Table 4.2 or read from Table 4.1), 16 (6.64%) of those are contributed from arson offenders who had been involved in multiple types of convictions. The proportion of multiple types of convictions who contribute a subsequent homicide for the other three focus offences are: threats to kill 25% (15 out of 60); blackmail 29.41% (10 out of 34), and kidnapping 35.42% (17 out of 48). Therefore, kidnappers who were involved with more than one type of serious crime appear to be more likely to escalate to more serious crime, such as homicide.

The last thing to point out is the averaged follow-up time in this table has a different meaning compared with the definition in Table 4.1. The follow-up time here refers to the time between the *last conviction* on one of the four serious offences to the end of the study. The length of the follow-up will on average be much shorter for an offender who has been convicted for a number of serious offences in sequence compared to an offender who has been convicted of only one serious offence. The varying length of average follow-up will affect the probability of observing a later homicide conviction. This feature is shown in the last column of Table 4.2. For each offence, those with two or more convictions have on average a shorter follow-up period than those with just one relevant conviction.

4.3.2 Cox proportional-hazards model

The previous descriptive statistical analysis suggested that those offenders who were specialised in their serious crime behaviour, involving only one type of crime of interest, have a lower risk of being subsequently convicted of homicide than those who were involving two or more types of serious crimes. Other results such as the observed differences in subsequent homicide conviction rates among the four focus offences are all needed to formulate a more appropriate statistical approach. Such an approach can take into consideration of other risk factors which may impact on failure time (escalation to homicide conviction). Moreover, this approach can explore whether any types of sequencing are related to a higher risk of subsequent homicide conviction. Therefore, the Cox proportional-hazards model is applied for such purposes and to look for risk factors for a subsequent homicide conviction following one or more serious crimes of interest.

By doing so, this study will also be able to control for other important risk factors such as gender, age at conviction and previous convictions; it is recognised that these three risk factors are important in the prediction of recidivism of all types.

The sequencing of the serious focus crimes starts with the first focus conviction, and ends with the last focus conviction prior to homicide. Therefore, the last of these convictions are needed to take into consideration and are needed to examined the time from that conviction to homicide, while adjusting for periods spent in custody (as described earlier in section 4.2.1). To summarise the sequencing of an offender's serious criminal history, five measures were taken as the potential risk factors in the Cox analysis:

1. The type of the first focus conviction in the sequence (i.e. arson, blackmail, kidnapping, threats to kill or a mix of more than one of these offences – a co-conviction).

2. The type of the last conviction in the sequence.

3. The number of different types of focus offences involved in the sequence. This measures the diversity of serious offending.

4. The number of focus offences, giving a measure of serious offence frequency.

5. The duration of the sequence, namely the time between the first and last focus offence in the sequence.

A series of Cox proportional hazard models were fitted, and a final model was chosen by examining the AIC (Lindsey and Jones, 1998) for each model, and removing insignificant risk factors. The final model with the lowest AIC included age, gender, logged previous convictions, the type of the first serious offence and

Table 4.3: Cox proportional-hazards model (adjusted for time at risk) for subsequent homicide conviction following one or more sample offences.

	Coef.	S.E.	Relative risk	z	p
Log(no. previous convictions)	0.7949	0.0622	2.214	12.781	< 0.001
Age at conviction	0.0493	0.0079	0.952	6.238	< 0.001
Gender: Female vs. Male	0.5798	0.2846	0.560	2.037	0.042
First serious conviction:					
Arson					
Blackmail	0.1267	0.2043	0.881	0.620	0.540
Kidnapping	0.3940	0.1890	1.483	2.084	0.037
Threats to kill	0.4387	0.1699	1.551	2.582	< 0.001
Co-conviction	0.2872	0.4669	1.333	0.615	0.540
No. different types of serious sample offences	0.6015	0.2946	1.825	2.041	0.041

the number of different types of serious offence. All other potential risk factors were not significant.

The Cox analysis models the underlying hazard rate (rather than survival time), and the results are shown in Table 4.3. As expected, males are more at risk of a homicide conviction than females ($p = 0.042$), with the risk doubling compared to females. Also, as expected, the risk of homicide declines with age ($p < 0.001$), with around a 5% decline in risk for each year of age and increases with the number of prior convictions ($p < 0.001$).

Results on the risk factors which are summarizing the sequencing of the four focus offences suggest that two of five measures of interest were important. First, the type of the first serious offence is a significant risk factor for subsequent homicide. An offender who began their serious offending with kidnapping has about 48% higher risk of homicide compared with one who started with arson; similarly, an offender who began with threats to kill has around 55% higher risk over the arson offender. Other types of starting offences show no significant difference from arson. Second, the number of distinct types of serious offences was also a significant risk factor, with increasing homicide risk as the number of types increases. Thus, the offender who had two types of serious offence has nearly double (1.82) the risk of subsequent homicide conviction compared with an offender with only a single type of serious offence. This is also consistent with Table 4.2.

4.4 Conclusion and discussion

4.4.1 Conclusion

This chapter (Study 1) has been an attempt to open up a discussion about the interrelationships between serious types of crime and escalation from those serious crimes into the most serious crime – homicide. This feature of criminal careers has been neglected in the criminological literature. The previous results can be summarised in the following points:

1. There is evidence that those convicted of one of the four focus offences are often specialised in the types of serious offences for which they get convicted. However, they vary in this respect, with those who are convicted of arson the most likely to specialise.

2. Both the descriptive analysis and the Cox proportional-hazards model suggest that those convicted of arson and blackmail all have very similar risk of subsequent convictions for homicide. Basically, the Kaplan-Meier survival curves showed that, for both arson and blackmail, the risk of subsequent convictions for homicide is 0.8% in a 20-year follow-up period – that is, around one in 125. Moreover, the hazard is not significantly different between offenders whose first focus offence was arson compared with those whose first focus offence was blackmail.

3. Those offenders who were convicted of threats to kill are similar to arson offenders and blackmail offenders in terms of their risk of escalating to homicide (0.8%) in a 20-year period. However, both threats to kill offenders and kidnapping offenders have a quicker speed of homicide reconvictions than arson and blackmail offenders in a 10-year follow-up period. In particular,

kidnapping offenders seem to have a higher risk of subsequent homicide, with a higher proportion (1.0%) – around one in 100 – being subsequently convicted of homicide. This compares with the likelihood of around one in 3,000 male members of the general population being convicted for homicide over a 20-year follow-up period (Francis and Soothill, 2000).

4. The speed of subsequent conviction for homicide varies for these four offences. In particular, kidnapping shows the most rapid gradient, with a heightened risk of homicide in the early years after conviction. In contrast, arson offenders have a fairly constant rate over the 20-year period.

5. Having two or more of the focus offences can increase the risk of subsequent homicide. The effect becomes clearer when incorporated into a statistical analysis of risk factors for homicide. The involvement of one more serious offence nearly doubles the chance to escalate to homicide.

6. The risk of a homicide conviction decreases with increasing age, and is significantly lower for females compared to males.

4.4.2 Discussion

The motivation of this study is to identify whether there are "gateway" offences for homicide among a collection of serious offences such as arson, kidnapping, blackmail, and threats to kill. It has been found that the risk of homicide is high for kidnapping offenders (one in 100), arson, blackmail and threats to kill offenders (around one in 160) and the risk doubles for offenders with a mix of such gateway offences.

The findings from this study have some important policy implications. Firstly, previous work has identified that the likelihood of a male member of the general population being convicted for homicide over a 20-year period is one in 3,000 (Francis and Soothill, 2000). This risk substantially increases to around one in 100 particularly for kidnapping offenders who have the highest risk of subsequent homicide among the four serious offences. Secondly, if someone is convicted of a serious offence, such as arson, blackmail, kidnapping or threats to kill, then this study has given evidence that those convicted of more than one serious offence are at higher risk of homicide. For instance, the offender who had two types of serious offence has nearly double (relative risk of 1.82) the risk of subsequent homicide conviction compared with an offender with only a single type of serious offence.

It has always been an issue whether criminal justice professionals can correctly identify and selectively target a small group of offenders who have committed certain types of serious offences and who are likely to escalate to even more serious offences. This study has indeed identified small groups of such individuals and it is worth considering whether increased monitoring of such offenders after release is worthwhile. Such monitoring could take the form of additional resources being directed towards rehabilitation on release from prison.

Alternatively, offenders with such serious offences and also with multiple types of serious offences could be registered in a serious offenders database, which would be similar to the existing Violent and Sex Offender Register (ViSOR) database in the United Kingdom [2]. This is a national database designed to enable probation, police and prison services to share information, risk assessments and intelligence about high risk offenders. Arguments against such a policy might be that 99 individuals (out of 100 kidnapping offenders) would be needlessly added to the register to identify the one case which will go on to commit homicide; and this may be a violation of rights. However, Soothill and Francis (1997) in talking about the registration requirements of the Sex Offenders Act 1997, state that it is "a laudable aim to try to keep those most likely to be serious sexual recidivists under surveillance" as long as at risk groups are identified in a systematic manner. The extension of the ViSOR database to cover those at risk of serious escalation to homicide may be a sensible strategy towards the reduction of homicide in future years.

[2]http://www.npia.police.uk/en/10510.htm

The next chapter will now describe one of the two studies on general escalation in crime seriousness. The study of general escalation will be carried out in two stages. The first stage which is 'Study 2' is focusing on the use of linear mixed-effects modelling to examine sequence of seriousness scores from each individual offender and considering two temporal scales – age and experiences of going through criminal justice system – on the effects of escalation over conviction occasions.

Chapter 5

Methodology for Escalation in General Crime Seriousness

5.1 Introduction

The previous study ('Study 1') has attempted to assess escalation from four specific types of serious crimes into the most serious crime of homicide. From this chapter onwards, escalation in crime seriousness is studied through a more general crime population. Therefore, the nature of this study is different from 'Study 1'. The failure time to a specific event is no longer the research interest of this study. In contrast, this work focuses on how offenders' seriousness of crimes change through conviction occasions from their first conviction over their entire criminal careers. This study assesses whether offenders have different developmental trajectories over conviction occasions – escalation, de-escalation, or remaining constant in seriousness.

As mentioned previously, escalation in crime seriousness over the criminal life-course continues to be an important issue of study in criminal careers. Quantitative research in this area has not yet been well developed owing to the difficulty of measuring crime seriousness and the complexity of escalation trajectories. Therefore, this thesis attempts to overcome these challenges by identifying and considering three main methodological issues. The three methodological challenges are measurement of crime seriousness, temporal scales, and methodological approaches to assess escalation in crime seriousness. A previous chapter (Chapter 2 on 'Concepts of Escalation') has provided a detailed summary of various means of measurement of seriousness, reviewed existing methodological approaches for studying escalation in the literature, and discussed three temporal scales in crime escalation.

This chapter focuses on the methodological side of the study of general escalation. It describes methods which are used for assessing crime seriousness (section 5.2), three temporal scales (section 5.3), and three competing approaches for modelling trajectories of crime seriousness (section 5.4). The empirical results based on applying these methods will then be presented in the subsequent two chapters.

5.2 Measuring crime seriousness

For the study of general escalation in crime seriousness, a recently developed measure of crime seriousness (Francis et al., 2005) which was based on court sentencing is used. This method used sentencing data taken from the Police National Computer for 76,699 offenders who had a court conviction or a police caution in the month of January 2001. This sample consisted of 126,790 separate offences after excluding convictions from the British Transport Police and Scottish and Northern Irish police forces. The resulting seriousness scale therefore relates only to offences in England and Wales. The final dataset consisted of 405 separate offence codes and subcodes defined by the Home Office – see Home Office (1998) for an earlier version of the codes used.

Francis et al. categorised offence sentences into 74 disposal categories – these were either a single disposal (such as a fine, custody in prison up to one month, drug treatment and testing order under one year) or a multiple disposal (such as a curfew order under three months combined with a community punishment under one year). They then formed a two-way cross-classified table of offences by disposals, and applied *correspondence analysis* to scale both offences in order of seriousness and disposals in order of severity. Finally, the resulting correspondence analysis offence score was rescaled and log-transformed so that the most serious offence had the score of 10 and the least serious a score of zero, which produced an approximately normal distribution on the transformed scale. For instance, the score for murder is 10.0, the rape of female aged 16 or over is 8.1, robbery is 5.7, petty theft is 3.9 and selling food not complying with food safety is 1.1. Full details are given in Francis et al. (2005).

The major strength of the correspondence analysis approach is that it allows all forms of sentencing to be taken into account in the production of a crime seriousness scale. This is particularly important in jurisdictions such as England and Wales where a wide range of non-custodial disposals are used.

Compared with using a more traditional approach of grouping offences into a limited number of crime categories, the advantage of using this continuous score is that it is more sensitive to change over time. In addition, it allows the evaluation of the seriousness of each offence regardless of the broad category of crime. For example, possession of a firearm with intent to cause fear of violence (a violence offence) and stealing from another person (a property offence) both have a score of 4.2; in contrast, in a more traditional approach of measuring seriousness, the violent offence might be viewed as more serious compared to the non-violent property offence.

There are also criticisms which can be made. One is that the correspondence analysis methodology combined with the transformation does not produce a scale that can be interpreted as a ratio scale. Thus it is not possible to say that murder is twice as serious as robbery. There has been a lively debate about whether seriousness scales should be additive (Wagner and Pease, 1978). Some of this literature has been concerned with whether crime seriousness of two individual crimes committed separately would, when added together, produce the same seriousness score as one for the two crimes committed jointly. In order to avoid this problem, the maximum seriousness score is taken over all offences brought to court on a particular occasion.

Another criticism which has been made about crime seriousness scales measured from court conviction data is that of media bias. This says that a judge or magistrate might be unduly influenced by media reporting or campaigning and thus will impose a higher sentence than would otherwise be the case. Bruschke and Loges (2004), for example, made the claim that in the US, the more publicity a trial receives, then the stiffer the sentence given. If this is repeated over many such offences of the same type, then this would generate a higher seriousness score. In this study it is recognised that crime seriousness can change over time for a variety of reasons, with certain offences increasing in seriousness and reflecting changing public and media views on the ordering of crimes. Media influence will therefore be part of this process of changing norms. Hence, any particular study will reflect crime seriousness at a moment in time.

One potential problem is changes to legislation over time. Since the Francis et al. score was computed based on 2001 convictions, some crimes convicted in earlier years have disappeared from the statute book and thus do not have seriousness scores. In order to deal with such offence types, the appropriate seriousness score for an equivalent modern offence was assigned to them. However, for some offences, such matching could not be undertaken, and 200 offenders with such missing information were omitted from the analysis. The final dataset contains 4,831 offenders (see Chapter 3 section 3.3.2 for more details).

The above discussion is concerned with the seriousness of an individual offence; however regression analysis below is concerned with conviction occasions (court appearances), which consist of one or more offences brought to court at the same time. In this work, the definition of the seriousness of a *conviction occasion* is the maximum seriousness score of the convicted offences at that court appearance. Thus the measurement of the conviction occasion seriousness refers to the seriousness of the worst convicted offence rather than the total seriousness over all convicted offences in the court appearance – this latter method can be thought of as the total damage caused by the offender at that court appearance. The rationale for taking the seriousness of the worst convicted offence is that this work conceptually views an offender as committing major offences around their personal mean seriousness level together with other minor offences, and taking the worst offence provides a better measure of an individual's mean seriousness level. Moreover, the assumption that all offences brought to court at a conviction occasion are committed at the same time is rarely true in practice, making the 'total damage' approach therefore is problematic.

By taking the maximum seriousness score, it is necessary to note that the maximum of n independent samples from a common distribution increases with n [1]. To account for this effect, the number of offences at the conviction will be taken as a covariate in the analysis.

5.3 Temporal scales in crime escalation

As mentioned in a previous chapter (Chapter 2, section 2.4), there are two approaches in choosing a temporal scale to assess escalation in crime seriousness over time. One approach is to assess crime seriousness over age, the other is to assess by the order of conviction occasions. In the study of general escalation, both changing crime seriousness over age and over conviction occasions are considered together.

Additionally, the *conviction occasion* is preferred as the primary temporal scale in the study of general escalation, whereas the *age* is treated as an explanatory variable. There is no essential difference between taking age or conviction occasion as the primary temporal scale in terms of statistical modelling. The rationale for observing the sequence of individual's crime seriousness by conviction occasions, is to follow up every individual offender from the same starting point, which is the time of the first conviction occasion. Therefore, the interpretation of escalation

[1] Ewens and Grant (2005, p 92) highlight the result that if X_{max} is the maximum of n independent and identically distributed random variables $(X_1, X_2, \cdots, X_n)$ with cumulative distribution function $F(X)$ then for any fixed x, $P(X_{max} \leq x) = P(X \leq x)^n = (F_X(x))^n$. Thus the mean of X_{max} increases with n.

over time is the development of crime seriousness since an offender' s first time of conviction.

As also mentioned in section 2.4, a third temporal scale is also considered in the study of general escalation, that is the *time spent in prison.* There is no previous quantitative research on the effect of escalation. Therefore, it is interesting to examine this from the respect of criminological research. The time spent in prison is not include in 'Study 2' but is examined in the subsequent study – 'Study 3'. More details of the purpose of each study is given in the relevant chapters.

5.4 Competing approaches for modelling trajectories

For the analysis of longitudinal data, the *linear mixed-effects* (LME) model is a well developed and popular statistical approach, which is well described in many texts such as Pinheiro and Bates (2000) and Verbeke and Molenberghs (2000). It is also common for this model or closely related to models to be referred to as a growth curve model (GCM).

Alternatively, in the distinct subject areas of psychology, medicine, and criminology, the work of Nagin and Land (1993) has popularised the use of *group-based trajectory modelling* (GBTM) or *latent class growth analysis* (LCGA) for the study of developmental trajectories. Both approaches assume that there are a number of latent subpopulations with different temporal trajectories present in the data. While there has been a great deal of interest in assessing changes in the frequency of drug use or offending, there has been little interest in assessing change in the level of crime severity over conviction occasions or age. This is nevertheless an important research area for criminologists to understand how offenders develop their criminal careers in terms of the seriousness of crimes.

Despite the popularity of Nagin's model for understanding trajectories, there are a number of alternative statistical methods that have also been used. In psychological and sociological applications, the term *growth mixture modelling* (GMM) is also commonly found in the literature. In the statistical literature, alternative terms such as the *heterogeneity model*, and *latent class linear mixed model* (LCLMM) can be found.

Therefore, there are different terminologies which are commonly used by researchers in the areas of studying developmental trajectories and longitudinal data. They all are designed to study and model repeated observations over time with many of these approaches taking account of within individual and between individual variation. However, it seems very confusing for researchers from different disciplines to understand the meaning of each method clearly.

The different statistical methods can be categorised into three broad methodologies: the *mixed-effects modelling* approach, the *mixture modelling* approach, and the *mixtures of mixed models* approach. Individual methods within each group are conceptually very similar and only differ with respect to technical details. In the following, the statistical properties of each approach are summarised and different approaches are grouped together according to the type of methodology. Additionally, two major comparison studies on applications by using two of these methods are described in this section.

For the convenience of showing definitions of each method mathematically, this study uses a capital letter for a random variable, and bold case for a vector or a matrix. Let Y_{it} represent the response variable, for observations $i = 1, ..., m$ at the time points $t = 1, ..., n_i$. Here m is the total number of cases, and n_i is the number of observations for each case i. The mean and variance of Y_{it} are represented by $E(Y_{it}) = \mu_{it}$ and $Var(Y_{it}) = \sigma^2_{it}$. The y_{it} is then the observed response from the random variable Y_{it}.

The repeated outcomes for each i can be gathered into an vector of length n_i, $\boldsymbol{y}_i = (y_{i1}, ..., y_{in_i})$. The responses for all i are stacked into a long vector of length n, thus $\boldsymbol{y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_m)$, with $n = \sum_{i=1}^{m} n_i$.

5.4.1 Mixed-effects modelling approach

The terminologies of the *linear mixed-effects* model (Laird and Ware, 1982; Diggle et al., 2002) and the *growth curve model* (Rao, 1965; Fearn, 1977; Verbyla, 1986; Verbyla and Venables, 1988) are commonly used in the study of longitudinal data analysis. Plewis (1996) has shown the link between growth curve modelling and multi-level modelling in a psychological context. These two terminologies essentially refer to the same approach but differ with more specific setups.

The terminology of *growth curve model* is commonly used in the disciplines of sociology, psychology, and criminology through the statistical package MPLUS. The conventional growth curve model is often expressed in the form of an intercept plus variables representing the time effect (slopes), such as time and time squared which are referred as *growth factors*. Individual-level variability in the intercept and the polynomial time parameters are represented by random effects which are multivariate normally distributed with means of 0 and an estimated variance-covariance matrix (Diggle et al., 2002, Chap. 5). It is commonly used in conjunction with time-constant explanatory variables to explain the variation in individual growth curves. Hwang and Takane (2005) also pointed out that the conventional GCM assumes that the covariance matrix of repeated measurements is unstructured. Typically, GCM in the social sciences is normally used for data with a relatively small and equal number of time points for each subject.

The linear mixed-effects model represents a broader framework of models than the GCM. Typically, time-varying explanatory variables may also be included to explain within individual-level variation and the number of time points can vary across each subject. It can also provide a more flexible structure to define the covariance matrix, such as various forms of serial correlation within subjects over time. Therefore, GCM can be viewed as one type of model within the class of linear mixed-effects models.

These two models therefore share a common approach to trajectory estimation – a mean trajectory for all cases is estimated through a polynomial function of time, and variability over cases is represented by random effects terms on the intercept, slope and higher order polynomial terms. Therefore, the LME model approach is adopted to the modelling of crime seriousness. This approach allows both within-individual and between-individual variation to be estimated, and for both time varying and time constant covariates to be included.

The linear mixed-effects model can be defined as follows:

$$Y_{it} = \boldsymbol{\beta} \boldsymbol{X}_{it} + \boldsymbol{u}_i \boldsymbol{Z}_{it} + W_{it} + \epsilon_{it}. \tag{5.1}$$

where $\boldsymbol{X}_{it}$ is a p-vector of fixed-effects covariates, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ is a p-vector of unknown regression coefficients for the fixed effects. The fixed-effects of $\boldsymbol{X}_{it}$ can include both time varying and time constant covariates. $\boldsymbol{Z}_{it}$ is a q-vector of random-effects covariates, with a q-vector of unknown subject-specific coefficients $\boldsymbol{u}_i = (u_{i1}, ..., u_{iq})$. It is common that $\boldsymbol{u}_i = (u_{i1}, u_{i2})$, where u_{i1} is the random intercept and u_{i2} is the random slope for time; with $\boldsymbol{u}_i \sim MVN(\boldsymbol{0}, \boldsymbol{V})$ and with $\boldsymbol{V}$ a two by two variance-covariance matrix of the $\boldsymbol{u}_i$, with diagonal terms $var(u_{i1}) = v_{11}$, and $var(u_{i2}) = v_{22}$, and an off-diagonal covariance $cov(u_{i1}, u_{i2}) = v_{12} = v_{21}$. Hence, the two random terms have a correlation of $\frac{v_{12}}{\sqrt{v_{11}v_{22}}}$. W_{it} is a serial correlation term which depends on the assumed within-offender correlation structure. Finally, ϵ_{it} is the residual error term with $\epsilon_{it} \sim N(0, \tau^2)$.

This model assumes that offenders' criminal histories are independent from each other. While this is mostly true, there will be some instances of co-offending, where offenders are working together and committing and being convicted of the same crimes. As this study analyses a birth cohort of offenders who were born in four selected weeks, and two co-offenders are unlikely to be in this sample together, it is likely that this assumption is valid.

One potential criticism of the linear mixed-effects model is the assumption of multivariate normality of the random effects ($\boldsymbol{u}_i \sim MVN(\boldsymbol{0}, \boldsymbol{V})$). Verbeke and Lesaffre (1996) state that violation of this assumption may seriously influence the parameter estimates, especially for the estimate of random effects.

5.4.2 Mixture modelling approach

The second common approach to trajectory estimation is through *group-based trajectory modelling* (Nagin, 1999, 2005). This approach assumes that the population is composed of a mixture of distinct groups defined by their developmental trajectories. Thus, instead of assuming a multivariate normal distribution of random effects in the linear mixed-effects model, this approach uses a finite number of groups to approximate a continuous distribution of random effects. The groups can be considered to be latent classes. Each individual will have a probability of belonging to a specific trajectory class – thus variability between individuals is represented through the varying individual probabilities of class trajectory membership. Therefore, there is no specific inclusion of any underlying random effects, and homogeneity is assumed within each identified trajectory class.

An alternative way of thinking about the GBTM approach is to conceptualise it as a linear mixed-effects model but with a finite number of discrete random effects or mass points (Laird, 1978). Therefore, the GBTM approach benefits from no assumption on the distribution of random effect. The unknown mass points interact with the growth factors of time, time-squared etc. to provide the equivalent of the random slopes in the mixed-effects model. This model is sometimes known as the non-parametric maximum likelihood (NPML) approach to the mixed-effects model (Aitkin, 1999).

The model can be generalised from the linear mixed-effects model (Equation 5.1) in the following way. Assuming the existence of K classes, and given the latent class k with $k = 1, .., K$, the model can be written as:

$$Y_{it|c_i=k} = \boldsymbol{\beta X}_{1it} + \boldsymbol{\alpha}_k \boldsymbol{X}_{2it} + \epsilon_{it}. \tag{5.2}$$

where $\boldsymbol{X}_{1it}$ is a p-vector of common effect covariates, and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ is a p-vector of unknown regression coefficients that have common effects across all classes. On the other hand, $\boldsymbol{X}_{2it}$ is a q-vector of class-specific covariates, and $\boldsymbol{\alpha}_k =$

$(\alpha_{k1}, ..., \alpha_{kq})$ is a q-vector of unknown regression coefficients with the coefficients varying across classes. ϵ_{it} is the residual error term for each individual i at time t, where $\epsilon_{it} \sim N(0, \tau^2)$. Therefore, the residual variance τ^2 is assumed to have a common variance across different classes. However, this assumption can be extended by allowing class-specific residual variances (ϵ_{itk}).

Thus, conceptually, there are now two types of covariates which can be included in the model. The first ($\boldsymbol{\beta}$) acts at the population level, and assumes the effects are common for all individuals. The second ($\boldsymbol{\alpha}$) acts at the class-level, and so the effects here will vary across classes. Thus if time and powers of time are treated as class-specific covariates, then the shape of the developmental trajectory among each latent class of individuals will vary.

In summary, the group-based trajectory approach is more flexible as it allows risk factors (both time-varying and time-constant variables) to vary across each latent class of individuals. Linear mixed-effects model can let a covariate have a random regression parameter, which allows the effect to vary across individuals, but in a more restrictive way.

Implementations of this model for balanced data with the same number of time points per case are available through the SAS procedure PROC TRAJ (Jones et al., 2001), and via the MPLUS package, where the method is referred to as latent class growth analysis (LCGA). For unbalanced data (unequal number of repeated measurements within each observation), the *lcmm* package in R (Proust-Lima and Liquet, 2011), the Latent Gold package (Vermunt and Magidson, 2005) (using the latent regression option), and MPLUS (by fitting a two-level model with a latent factor though the `TWOLEVEL MIXTURE`) command are suitable options. The difference in terms of modelling assumptions between the *lcmm* package with the other two software packages is that the *lcmm* package has an assumption of a class-independent residual variance (ϵ_{it}), but both the Latent Gold package and MPLUS allow a class-specific residual variance (ϵ_{itk}). All these packages allow covariates at both the class level and at the population level.

5.4.3 Mixtures of mixed models approach

While the group-based trajectory modelling provides a framework to identify latent subpopulations and to estimate their distinct trajectories, the model assumes that each class-specific trajectory is a good representation for all members of its class. In other words, variation around the expected trajectory within a class is assumed to be zero. Additional models which are termed as "mixtures of mixed models" have therefore been proposed to relax this assumption.

The simplest extension is the *heterogeneity model* (Verbeke and Lesaffre, 1996; Verbeke and Molenberghs, 2000) which is basically a form of finite mixture model (McLachlan and Peel, 2000; Titterington et al., 1985). This model assumes that the population distribution of trajectories is composed of a discrete number of latent subpopulations, each following a conventional linear mixed-effects model. To avoid numerical convergence issues, their method assumes a common variance covariance structure for the random effects in each class. In other words, the individuals' variation around the expected trajectories within each class is the same.

A more flexible extension – the *growth mixture model* (GMM) - was proposed by Muthén and Shedden (1999), and relaxes the assumption of a common covariance matrix. For each class, a unique covariance matrix of growth factors and intercept can be estimated. Proust and Jacqmin-Gadda (2005) have proposed an alternative name, the *latent class linear mixed model* (LCLMM), for either the growth mixture model or the heterogeneity model when modelling continuous response variables.

Both methods can be thought either as an extension of the linear mixed-effects model to handle heterogeneous populations (with the number of classes > 1), or as an extension of group-based trajectory modelling to account for correlation between repeated measures of the same subject and the variance within each sub-population.

The formal definition for GMM is as follows. Given the latent class k, the trajectory of the outcome is described using a linear mixed-effects model, and is given by:

$$Y_{it|c_i=k} = \boldsymbol{\beta} \boldsymbol{X}_{1it} + \boldsymbol{\alpha}_k \boldsymbol{X}_{2it} + \boldsymbol{u}_{ik} \boldsymbol{Z}_{it} + \epsilon_{it}. \tag{5.3}$$

where the vectors of $\boldsymbol{X}_{1it}$, $\boldsymbol{X}_{2it}$ are defined as in Equation 5.2. The third term $\boldsymbol{Z}_{it}$ is the vectors of class specific random effects (intercept and growth factors), where $\boldsymbol{u}_{ik}$ is assumed to follow a mixture of K multivariate Gaussians with probabilities π_k and with different means (c_k) and covariance matrices $\boldsymbol{V}_k$, with e.g. $\sum_{k=1}^{K} \pi_k c_k = 0$ for identifiability. When $k = 1$ this model becomes the linear mixed-effects model (Equation 5.1); alternatively, if the random effects are excluded($\boldsymbol{u}_{ik} = \boldsymbol{0}$), it becomes the group-based trajectory model (Equation 5.2).

A further extension to the GMM is to replace the assumption of multivariate normality of the class specific random effects above with a non-parametric alternative, estimating the random effects distributions within each class by a series of mass points with unknown masses and locations which are estimated from the data. This model is termed the non-parametric growth mixture model (NGMM) and has been considered by Kreuter and Muthén (2008) and Muthén and Asparouhov (2009).

Software implementations of the GMM model can be found in either MPLUS or in R. In MPLUS, both the `MIXTURE` and `TWOLEVEL MIXTURE` commands can be used. The `MIXTURE` command is for the analysis of balanced data. In contrast, the `TWOLEVEL MIXTURE` command can be used for unbalanced data with no time-dependent covariates. Therefore, analysis for GMM model by MPLUS is not flexible for unbalanced data and for models with large number of time-dependent covariates. Alternatively, the implementation of LCLMM (that is, GMM) which allows for unbalanced data and time-dependent covariates is provided by the R package *lcmm* (Proust-Lima and Liquet, 2011).

5.4.4 Comparison framework

To summarise, the three types of statistical approaches to longitudinal data above are distinguished from each other primarily by the assumptions regarding the underlying distribution of the individual trajectories in the population. There has been comparatively little work in comparing these approaches. There are two main research papers which have applied and compared such modelling approaches in the area of developmental trajectory studies.

Firstly, Kreuter and Muthén (2008) used four mixture modelling alternatives: the growth curve model, the group-based trajectory model (which they referred as latent class growth analysis), the growth mixture model (GMM) and the non-parametric GMM, to analyse conviction histories in two longitudinal criminological datasets (the Cambridge Study in Delinquent Development data and the Philadelphia cohort study data). They used both BIC and absolute standardised residuals for each response pattern as criteria for model selection. Their comparison methods focused on differences in overall fit, such as the average curve on convictions by age at offence, and significance of the age effects for each modelling approach. For the Cambridge data, they found that the four alternative models suggested no substantial differences in terms of number of classes, the characteristics of each class, the shape of curves over age and the proportion in each class. However, the four alternative approaches differed substantially for the Philadelphia cohort study. Their advice is essentially not to focus on one strategy, but to consider a variety of approaches before making inferences.

In contrast, the work of Bushway et al. (2009) focused on examining and comparing estimates of the *individual trajectories* from the *growth curve model* (GCM) and the *group-based trajectory models* (GBTM) based on offending prevalence data from a criminal career and life course study (CCLS) in the Netherlands. In terms of their comparison method, they first estimated separate trajectories for each individual offender by a method they called the *individual trajectory model* (ITM). ITM simply takes a sequence of observed offences from each offender as a subsample and estimates one trajectory for each person through a cubic regression function. They then computed Bayesian estimates of the individual trajectories from both the GCM and GBTM models. Finally they compared the Bayesian estimates to the estimates given by ITM using two statistical measures of bias: the *signed difference* (SDF) in the fitted probabilities of prevalence and the *absolute value of the signed difference* (ADF) of these probabilities, both of which were computed for each individual and at each age. Their comparison methods thus do not compare methods to the observed data, but rather assess bias towards ITM. They conclude that the average trajectories obtained from these three approaches are quite similar. On the other hand, for any given individual, these approaches tell very different stories, although GCM and GBTM are far more consistent relative to ITM.

Both of the above comparative studies also warn that care should be taken in assuming the existence of latent classes where none exist. Debates have been controversial (Nagin and Tremblay, 2005; Raudenbush, 2005; Sampson and Laub, 2005) and have been followed recently by a simulation study by Skarðhamar (2010) suggesting that evidence for groups is weak. However Bushway et al. (2009) also warn that GCM and GBTM may not detect classes with small numbers of cases which do not follow the general trend. Thus current practice suggests that mixture based models need to be used with care, but when well applied, can provide insight into underlying structure.

5.5 Summary

This chapter has defined and discussed methodological issues in the study of general escalation in crime seriousness. It firstly described the measurement of assessing crime seriousness. Then it defined the crime seriousness of a *conviction occasion* as the maximum seriousness score of the convicted offences at that court appearance. Secondly, the temporal scale of the conviction occasion is the preferred primary temporal scale, whereas the age and time spent in prison are considered as explanatory variables.

More importantly, this chapter reviewed three types of modern statistical approach which are designed to study repeated measurements over time. This is the first study that attempted to disentangle various statistical regression terminologies in the areas of longitudinal data analysis. Moreover, it describes the advantages and disadvantages of each approach, and summarised differences in major available statistical packages for each approach.

After talking through the related methodological issues in the study of general escalation, the empirical results of assessing general escalation will be illustrated in two stages. In the first part, 'Study 2' will analyse the Offenders Index data (see Chapter 3, section 3.3.2) by using the *linear mixed-effects* (LME) model to study the sequence of crime seriousness at both individual-level and population-level. The purpose of 'Study 2' is to initially explore the data in the respect of assessing the development of seriousness in offending. Two out of three temporal scales in crime escalation, namely age and conviction occasion, are considered in the initial analysis.

The second part of the result is shown in 'Study 3', which is a further development of 'Study 2' applying more sophisticated regression techniques. All three types of regression approach, which are the linear mixed-effects model, the group-based trajectory analysis, and the growth mixture model will be used and compared for assessing general escalation in crime seriousness. Moreover, all the regressions are examined with all three temporal scales and with more flexible age variables for improvement of accuracy in estimation of crime seriousness.

Chapter 6

Study 2: Escalation in General Crime Seriousness I – a linear mixed-effects modelling approach

6.1 Introduction

Following the previous Chapter on "Methodology for Assessing General Escalation", this chapter illustrates some initial results on the study of general escalation in crime seriousness. As previously introduced, there are various modern statistical methods for the study of repeated measurements over time. The purpose of this chapter is to explore the general developmental trajectory in crime seriousness at both individual-level and population-level. This study ('Study 2'), especially, focuses on using simple linear mixed-effects modelling approach to assess not only the marginal changes of seriousness but also variation within and between each individual offender.

The results of 'Study 2' has been recently published (Liu et al., 2011). The illustrated results in this chapter, however is slightly different than the published paper due to a slightly different dataset is used in this chapter. The dataset which is used in 'Study 2' includes more cases and has better matching (see Chapter 5, section 5.2) of the old offence codes with the current Home Office Standard List offences codings (2005). Therefore, analyses based on this dataset produce similar results to the published paper but with minor changes.

This chapter firstly describes variables which are used in this study (section 6.2). Then in section 6.3, it engages some basic descriptive statistics analysis, such as examination of observed overall and individual trend of seriousness by conviction occasions, assessing of serial correlation over convictions, and pre-checking multicollinearity between covariates. Section 6.4 is shown the results of the linear mixed-effects modelling in assessing the developmental trajectory of individual and overall crime seriousness. Finally, in section 6.5, findings from regressions are concluded, and contributions according to the analysis and potential development of this current approach will be discussed in the end.

6.2 Variables

This study analyses the 1953 birth cohort from the England and Wales Offenders Index (OI) data and followed through to 1999. The detailed description of this dataset is provided in Chapter 3 on 'Data Sources on Offending Behaviour' section 3.3.2. Variables of interest are extracted from this data and are described in the following.

A conviction occasion is referred as a distinct court appearance where an offender has been found guilty of one or more offences. Thus, an offender with two conviction occasions will have two separate court convictions at different dates. The term 'conviction' is then used as the shorthand for conviction occasion, and 'offences' is for convicted offences within a conviction occasion. The offender's age at any conviction is taken to be the age at court sentence.

As described earlier, the seriousness of a conviction is the maximum seriousness score for all offences at that conviction. The observed sequences of seriousness in crime from the first conviction are longitudinal sequences measured at each conviction occasion. Then the individual sequence of seriousness scores over convictions can be modelled through the LME model.

There are both time varying and time constant covariates in this analysis. Firstly, the time varying covariates are introduced in the following.

Order of conviction This is the number of current and prior conviction occasions. This provides a partial indication of the effect of criminal justice experience on escalation.

Age at conviction This is also a time varying covariate, and assesses the effect of maturation on escalation.

Age at conviction with one breakpoint This was used as an alternative to age at conviction. The concept here is to estimate separate maturation effects for ages less than 18, and for offenders 18 or more. Two dummies are constructed. The first takes the value of age of conviction up to age 17, and takes the constant value 17 for all ages above 17; the second takes the value zero if the age at conviction is less than 18, and $(age - 17)$ for ages 18 and above [1].

Number of offences This is the number of separate offences at the conviction occasion. The effect of this variable is expected to be positive, as the greater the number of offences brought before the court, the more likely the maximum seriousness at that conviction will be higher (Ewens and Grant, 2005). The variable was log-transformed as preliminary investigation showed that this gave a better fit.

There are also two time-constant covariates:

Gender Normally females offenders are expected to have an average lower seriousness score than male offenders.

[1]The breakpoint of 18 was chosen by fitting a sequence of mixed-effects models for a range of breakpoint values (age from 12 to 45), and choosing that value which minimised both the BIC and AIC criterion. See Stasinopoulos and Rigby (1992) for details.

Age at onset Offenders whom with an earlier age of onset are expected to have a higher average seriousness score.

The two age variables work together to form a breakpoint model at age 18 for age at conviction, with different slopes before and after age 18. The definitions ensure that the fitted seriousness scores are continuous at age 18.

As expected that the observed scores over convictions within each offender will be correlated, a traditional linear approach is not appropriate. The mixed-effects model approach extends the linear regression approach, allowing for consideration of variability both within- and between-individuals, and also accounting for any serial correlation which may exist. This approach allows the inclusion of a random intercept term which accounts for individual differences in the seriousness of the first conviction and also a random slope term for the order of the conviction, which allows individual differences in escalation over convictions.

In summary, the fundamental approach in this study is to look at crime seriousness by both order of conviction and age and to estimate the growth curve taking account of individual-level variation.

6.3 Exploratory analysis

Firstly this study explores graphically how seriousness changes over conviction number at both the population-level (the average of all offenders) and at the individual-level (each offender). This can be examined graphically through looking at the average sequence and individual sequence over convictions. Secondly, since the expected seriousness of an offender's convictions are likely to be serially correlated over his/her criminal career, the examination of the correlation within offenders over convictions is also needed. Finally, the correlation among the covariates of interest are briefly examined prior to any statistical modelling.

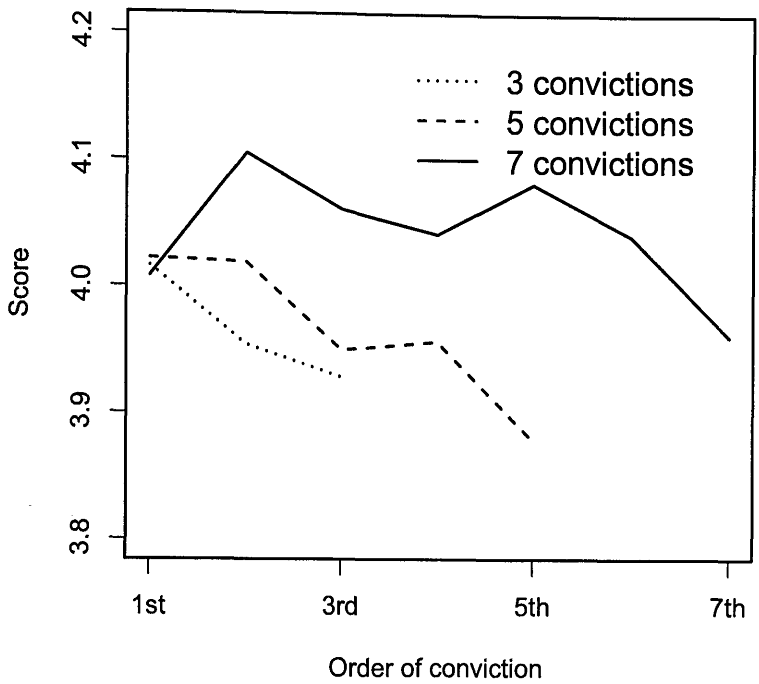

Figure 6.1: The mean seriousness score by conviction occasions for offenders with three, five and seven convictions.

6.3.1 Individual and average seriousness sequences

Figure 6.1 graphically examines the average seriousness by conviction occasions. The offenders who had three, five or seven convictions are shown in this figure for illustration purposes. This figure appears to suggest two things. Firstly, all subsets appear to have similar mean seriousness scores at the first conviction. Secondly, there appears to be evidence that those with a small number of convictions (three and five convictions) show evidence of de-escalation, whereas those with a more active criminal history (seven convictions) show evidence of escalation followed by later de-escalation. The range of the three sequences actually change very little, moving about ± 0.1 around a central value of 4.0.

Since these lines are marginal means, examination of the variability of each mean are needed. One way to examine this is to plot individual crime sequences

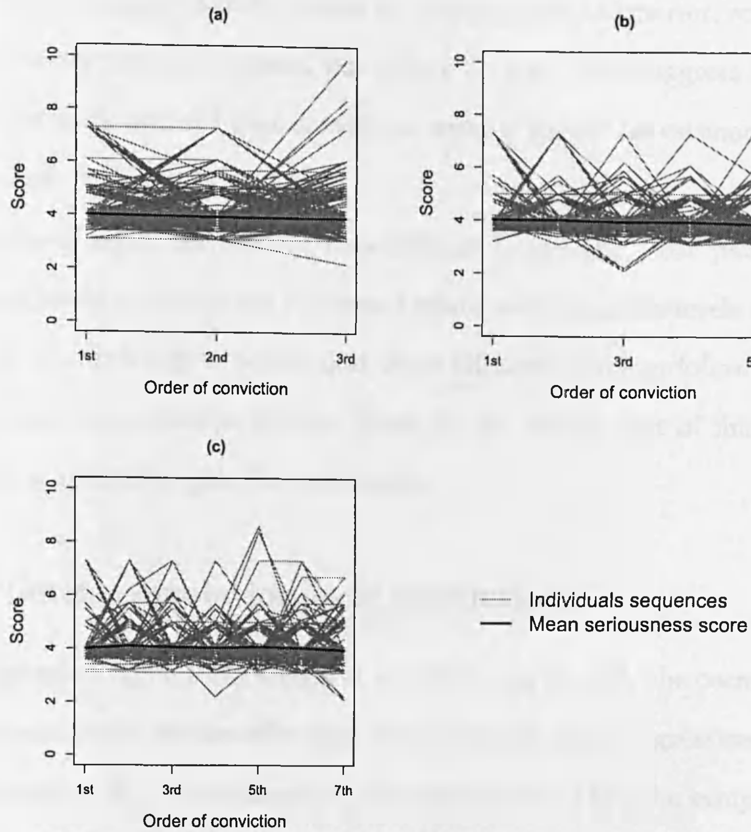

Figure 6.2: Individual seriousness score sequences for offenders with (a) three, (b) five, (c) seven convictions. The grey lines represent individual sequences and the black thick line represents the mean seriousness.

(Figure 6.2) for the same three subsets. The thick black lines repeat the mean lines shown in Figure 6.1, but the vertical scale has changed, allowing the variability of the individual sequences to be observed.

This plot can help to understand both variability within offenders and between offenders, also the serial correlation (over convictions). Firstly, considering variation *between offenders*, the seriousness scores of offenders' first conviction varied widely. This suggests that a different intercept for each offender should be estimated through a *random intercept* term by the mixed-effects model.

Secondly, considering variation *within offenders*, there appears to be two processes. On the one hand, a large number of offenders appear to have a low variability centered at a seriousness score of 4, indicated by the solid grey area. On the other hand, a smaller number of offenders change quite dramatically from less serious crimes to more serious crimes and back again. Moreover, some offenders tend to escalate over convictions, but others do not. This suggests that a different slope for each offender over conviction number should be estimated through a *random slope.*

In terms of serial correlation, it is difficult to see any clear pattern for how seriousness levels at conviction occasion t relate to seriousness levels at conviction occasion $t+1$, although it seems that some offenders tend to follow a pattern of escalation and de-escalation in turn. Next, in the second part of this section, the variogram is applied to gain further insight.

6.3.2 Serial correlation over convictions

The variogram (Diggle, 1990; Diggle et al., 2002) can identify the correlation structure over convictions within offenders. From a simple linear regression model of Y_{it} on the covariates $\boldsymbol{X}_{it}$, the residuals r_{it} are constructed. Then the sample variogram is calculated from a set of within-individual half squared differences

$$v_{ijk} = \frac{1}{2}(r_{ij} - r_{ik})^2, \tag{6.1}$$

where the r_{ij} is the residual for the offender i at the jth time of conviction, and the index of k stands for the kth time of conviction. For positive time differences of $u_{ijk} = (t_{ij} - t_{ik})$, is the gap difference of the conviction j and the conviction k. To construct the sample variogram $\hat{\gamma}(u)$, the values of the v_{ijk} are averaged for each value of u. The variogram graph then plots $\hat{\gamma}(u)$ against u. On the same plot, the process variance σ^2 can also be displayed; this is estimated to be the average of all half squared-differences $\frac{1}{2}(y_{ij} - y_{lk})^2$ with $i \neq l$.

In this plot, the ratio between the process variance and the sample variogram as the time difference u increases gives information on the autocorrelation present

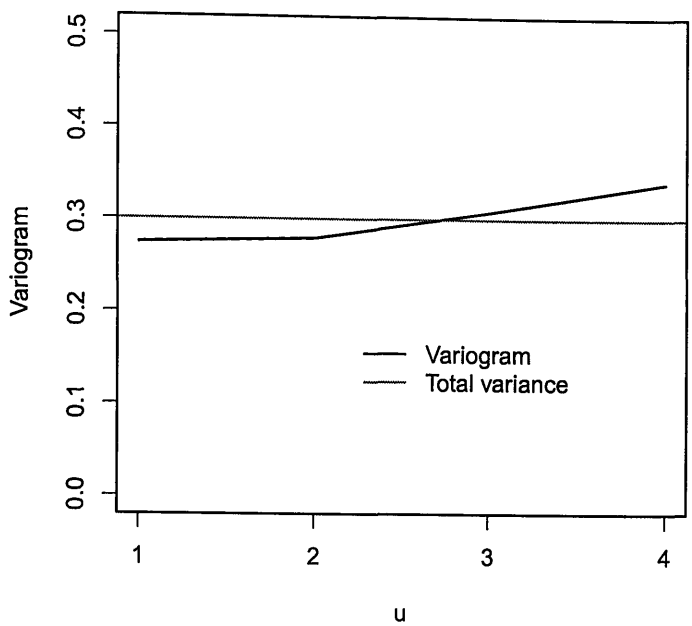

Figure 6.3: Empirical variogram for the five-conviction dataset. The grey line shows the total residual variance, and the solid black line the variogram.

in the data. Formally, the autocorrelation function is defined by

$$\hat{\rho}(u) = 1 - \frac{\hat{\gamma}(u)}{\hat{\sigma}^2} \tag{6.2}$$

For example, Figure 6.3 shows the sample variogram for offenders with five convictions. The lag quantities u take integer values between 1 to 4 as there are five time points. The value of the variogram is given by the solid black line; the grey horizontal line is the estimated process variance $\hat{\sigma}^2$.

The ratio of the variogram to the residual variance is very close to one for all four values of u, suggesting that the autocorrelation is close to zero for all lags. This in turn indicates that no complex form of decaying within-individual correlation is needed in this analysis.

6.3.3 Correlation between covariates

The primary concern here is to be able to detect multicollinearity between the covariates in this analysis. There have been two approaches taken to test this. The first was to look at correlations between the covariates. The Spearman rank correlation rather than the Pearson correlation were chosen as there was substantial skew in many of these covariates. All correlations were small, with the largest correlation being between age and order of conviction (0.489) and relatively modest in size. The second method was to calculate variance inflation factors (VIFs) for all covariates in the model (Marquardt, 1970). VIFs assess how much multicollinearity increases the variance of a parameter estimate; commonly, researchers have suggested that a value of 5 or more suggests significant multicollinearity (Stine, 1995). Fitting a linear model to this data produced VIFs which were all close to 1. Both methods, therefore, indicated no problem with multicollinearity.

6.4 Results of the linear mixed-effects modelling

A selection of mixed-effects models were fitted [2] to the complete data. According to the analysis from the previous section, all models included the random intercept and slope terms, and assumed that no serial correlation was present ($W_{it} = 0$). The information criteria AIC and BIC were used, and also performed a likelihood ratio (LR) test between nested models in order to determine the best fitting covariate model. The model selection process is shown in Table 6.1.

Model 1 estimates the effects for order of conviction, gender, and age at conviction. In adding the logarithm of number of offences at each conviction Model 2 was then obtained. The AIC and BIC from Model 2 are smaller than Model 1 and also the LR p-value is less than 0.05, indicating that Model 2 is preferred to Model 1.

Then Model 3 replaced the age at conviction variable with the age at conviction dummies which break the effect at age 18[3]. Again, the AIC and BIC both decrease and the LR-test again gives a highly significant p-value. Therefore Model 3 is chosen over Model 2. As a final stage, age at onset was added to give Model 4.

[2] Models were fitted using the R statistical program using function `lme` in the package `nlme`.

[3] Note that model 2 is nested in model 3 since constraining the two breakpoint dummies to have equal parameter estimates will produce model 2.

Table 6.1: Selecting the final mixed-effects mode : all offenders.

Models	1	2	3	4
		Fixed Effects		
Intercept	+	+	+	+
Order of conviction	+	+	+	+
Sex: female	+	+	+	+
Age at conviction	+	+	-	-
Age at conviction with breakpoint	-	-	+	+
log(No. of offences)	-	+	+	+
Age at onset	-	-	-	+
		Random Effects		
Intercept	+	+	+	+
Order of conviction	+	+	+	+
AIC	52254.34	51165.81	50994.76	50996.74
BIC	52319.75	51239.40	51076.53	51086.69
Likelihood-ratio test		M1 vs M2	M2 vs M3	M3 vs M4
P-value		$< .0001$	$< .0001$	0.8849

A + sign indicates that the relevant variable is included in the model.

Here the LR p-value is larger than 0.05 (0.8849) and the BIC increases, although the AIC decreases slightly. There is no evidence that age of onset increases average seriousness, and therefore Model 3 is selected as the final model.

Table 6.2 gives the estimated coefficients obtained by applying the final model (Model 3) to all valid offenders. In order to check the consistency of effects in criminal histories of different lengths, analyses of distinct subsets of offenders with two to three, four to six, and seven or more total convictions are also proceeded.

The results for all offenders are firstly described from the Table 6.2. The estimated baseline mean (the intercept) is 4.856; the random intercept standard deviation ($\sqrt{v_{11}}$) is 0.201, suggesting there is modest variation in mean seriousness over offenders. Unsurprisingly, the gender effect for females is significant and negative (-0.135), indicating that on average females are convicted of less serious crimes than males over their conviction history. The "number of offences" estimate is also positive (0.273) and significant, indicating as expected that the greater the number of offences per conviction, the higher the seriousness of the conviction. For example, in moving from one offence to two at the current conviction, the estimated seriousness score increases by $log(2) \times 0.273$ or 0.189.

Table 6.2: Final model: estimated coefficients of the mixed-effects model for offenders with two to three, four to six, and seven or more convictions, and all offenders respectively.

	2 to 3		4 to 6		7+		All	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Fixed Effects:								
Intercept	4.691	0.088*	5.022	0.106*	4.874	0.071*	4.856	0.0485*
Order of conviction	0.022	0.013*	0.017	0.007*	0.008	0.002*	0.009	0.0012*
Sex: female	-0.122	0.023*	-0.083	0.032*	-0.176	0.036*	-0.135	0.0172*
Age at conviction < 18	-0.047	0.006*	-0.066	0.007*	-0.054	0.005*	-0.055	0.0031*
Age at conviction 18+	-0.006	0.001*	-0.012	0.001*	-0.013	0.001*	-0.011	0.0007*
log(No. of offences)	0.297	0.019*	0.241	0.018*	0.272	0.010*	0.273	0.0080*
Random Effects :								
Intercept (v_{11})	0.1494		0.1427		0.0667		0.0404	
Slope (v_{22})	0.0374		0.0059		0.0004		0.0003	
Covariance (v_{12})	-0.0678		-0.0285		-0.0044		-0.0027	
Residual (τ^2)	0.2965		0.3429		0.4120		0.3822	

* indicates significance at the 5% level

The order of conviction estimate assesses escalation effects which may be considered to be due to experience of the criminal justice system. The coefficient is positive (0.009) and significant, indicating that, on average, offenders increase their seriousness score 0.009 with each conviction. The estimated standard deviation of the random slope for order of conviction ($\sqrt{v_{22}}$) is however large (0.018) when compared with the estimate of the mean slope (0.009), indicating that there is substantial variability in this slope across offenders.

Escalation effects due to maturity are given by the coefficients for age at conviction. In contrast to the positive sign for order of conviction, the age at conviction effects are both negative. Interestingly, the slope for offenders aged under 18 (-0.055) is more steeply negative than that for offenders 18 and over (-0.011). It suggests that for offenders younger than 18, seriousness decreases by 0.055 with each extra year of age. For those who are 18 and over, the seriousness score decreases by only 0.011 with every one year of maturation. This indicates that offenders de-escalate in offence seriousness as they get older, with the effect lessening for adult offenders.

Table 6.3: Various alternative mixed-effects models for all offenders.

Models	1		2		3		4	
	Coef.	S.E	Coef.	S.E	Coef.	S.E	Coef.	S.E
Fixed Effects:								
Intercept	4.293	0.0135*	4.239	0.0133*	4.856	0.0485*	4.855	0.0486*
Order of conviction	0.012	0.0012*	0.008	0.0012*	0.009	0.0012*	0.009	0.0014*
Sex: female	-0.134	0.0177*	-0.138	0.0173*	-0.135	0.0172*	-0.135	0.0175*
Age at conviction	-0.015	0.0006*	-0.015	0.0006*	NA	NA	NA	NA
Age at conviction < 18	NA	NA	NA	NA	-0.055	0.0031*	-0.055	0.0033*
Age at conviction 18+	NA	NA	NA	NA	-0.011	0.0007*	-0.011	0.0008*
log(No. of offences)	NA	NA	0.268	0.0090*	0.273	0.0080*	0.273	0.0080*
Age at onset	NA	NA	NA	NA	NA	NA	0.000	0.0013
Random Effects :								
Intercept (v_{11})	0.0448		0.0434		0.0404		0.0404	
Slope (v_{22})	0.0004		0.0003		0.0003		0.0003	
Covariance(v_{12})	-0.0029		-0.0028		-0.0027		-0.0027	
Residual (τ^2)	0.3997		0.3838		0.3822		0.3822	

* indicates significance at the 5% level; Models 1, 2, 3, and 4 refer to the models described in Table 6.1.

Finally, the random effects inform that there is greater variability within offenders (0.3822) than between offenders at the intercept (0.0404).

To enable to check the sensitivity of the final model (Model 3), offenders were divided into three subsets: those with two to three convictions, four to six convictions, and seven or more convictions (also in Table 6.2). In general, a similar story is produced from the analyses of subsets of offenders.

For these three subsets, the estimated 'order of conviction' effects are always positive (0.022, 0.017 and 0.008 respectively). In contrast, the estimated age effects for under 18s are all negative (-0.047, -0.066 and -0.054) and more steeply sloped than the estimated age effects for 18 and over offenders (-0.006, -0.012 and -0.013). The effects of gender and number of offences are also all similar. Although the average seriousness (the intercepts) are various across different subsets of offenders, the total number of convictions was not significant while tested through the LME model.

Although the final preferred model is Model 3, the likelihood ratio tests used

Table 6.4: Estimated marginal means from Model 3, for offenders with one offence per conviction date, and with selected number of convictions and age.

Order of	Juvenile			Young adult			Adult		
conviction	Age12	Age13	Age14	Age18	Age19	Age20	Age24	Age25	Age26
Male									
2	4.214	4.159	4.104	3.928	3.917	3.906	3.862	3.851	3.840
3	4.223	4.168	4.113	3.937	3.926	3.915	3.871	3.860	3.849
4	4.232	4.177	4.122	3.946	3.935	3.924	3.880	3.869	3.858
5	4.241	4.186	4.131	3.955	3.944	3.933	3.889	3.878	3.867
6	4.250	4.195	4.140	3.964	3.953	3.942	3.898	3.887	3.876
7	4.259	4.204	4.149	3.973	3.962	3.951	3.907	3.896	3.885
8	4.268	4.213	4.158	3.982	3.971	3.960	3.916	3.905	3.894
9	4.277	4.222	4.167	3.991	3.980	3.969	3.925	3.914	3.903
10	4.286	4.231	4.176	4.000	3.989	3.978	3.934	3.923	3.912
Female									
2	4.079	4.024	3.969	3.793	3.782	3.771	3.727	3.716	3.705
3	4.088	4.033	3.978	3.802	3.791	3.780	3.736	3.725	3.714
4	4.097	4.042	3.987	3.811	3.800	3.789	3.745	3.734	3.723
5	4.106	4.051	3.996	3.820	3.809	3.798	3.754	3.743	3.732
6	4.115	4.060	4.005	3.829	3.818	3.807	3.763	3.752	3.741
7	4.124	4.069	4.014	3.838	3.827	3.816	3.772	3.761	3.750
8	4.133	4.078	4.023	3.847	3.836	3.825	3.781	3.770	3.759
9	4.142	4.087	4.032	3.856	3.845	3.834	3.790	3.779	3.768
10	4.151	4.096	4.041	3.865	3.854	3.843	3.799	3.788	3.777

to select this model should be viewed as diagnostic tests for model selection, and it is important to view estimates for other fitted models to assess their stability. Additionally, therefore, the estimated coefficients from all four models (Model 1 to Model 4) are shown in Table 6.3. This shows that the estimated effects are remarkably stable across models. The estimate for gender, for example, varies from -0.138 to -0.134 over the four models.

Table 6.4 presents the marginal estimated means by order of convictions, age and gender, assuming one offence at each conviction date. Age is considered in three groups – juvenile (12-14), young adult (18-20) and adult (24-26). The table is useful for comparing average seriousness values across different offenders.

There are two effects which should be pointed in this table. Firstly, the effects of the parameter estimates of age can be seen directly. Thus, a male offender with two convictions, at age 13 has a seriousness mean of 4.159 whereas that for a 14 year old is 4.104 – a decrease of 0.055; mean seriousness from age 18 (3.928) to age 19 (3.917) decreases by 0.011.

Secondly, this table can compare the change in mean seriousness in adjacent age categories as the number of convictions changes. Offenders tend to escalate with increasing number of convictions. Male offenders younger than age 18 with more than six convictions a year or offenders aged 18 or over with two or more convictions a year are likely to show escalation. For instance, from Table 6.4, male offenders with two convictions at age 12, have approximately the same mean seriousness as male offenders with eight convictions at age 13 (4.213), and with escalation for those with *more* than six convictions (4.222). Male offenders with two convictions at age 18 have approximately the same mean seriousness as those aged 19 with three convictions (3.926). Similar statements can be made for females, where mean seriousness is lower (by 0.135).

The combination effect of age and experience on escalation/de-escalation can also be plotted through fitted seriousness trajectories under different offending scenarios. Figure 6.4 shows fitted trajectories for a male offender who started offending at age 10 and had only one offence per conviction. Three hypothetical scenarios are presented – the first where the offender has one court conviction a year, the second with three convictions a year, and the third with seven convictions a year. It indicates that for an offender with one conviction a year, the model suggests that there is no escalation, with mean seriousness declining from year to year. For three convictions a year, the fitted model shows declining seriousness up to age 18, followed by moderate escalation from age 18 onwards. An offender with seven convictions a year, in contrast, will show increasing escalation over the entire age ranges, which increases further at age 18. The figure highlights the point made earlier – that offenders with a small number of convictions a year on average will show de-escalation, whereas those with a large number of convictions a year will show escalation.

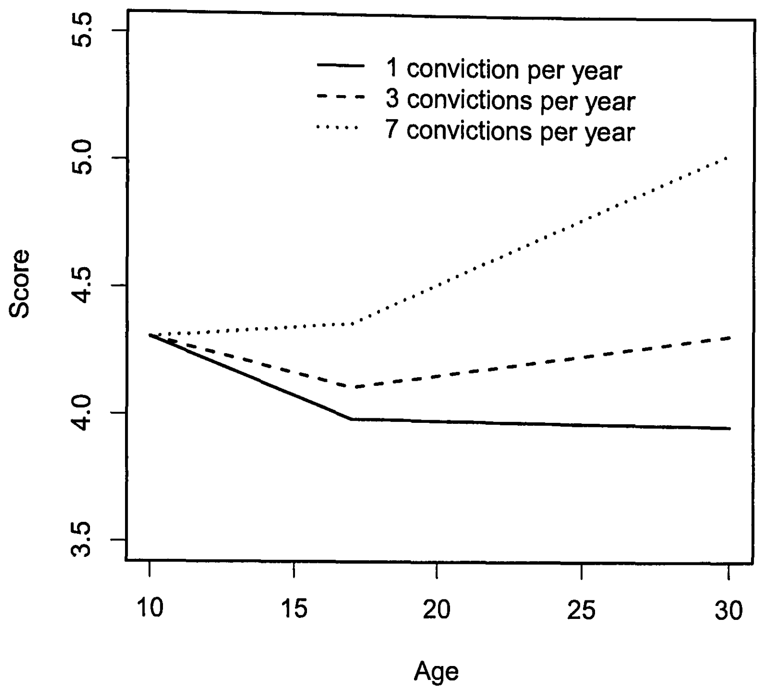

Figure 6.4: Fitted escalation trajectories for a male offender with one, three and seven convictions a year.

6.5 Conclusion and discussion

This analysis has enabled the detection and measurement of two distinct types of escalation – that due to increasing experience and exposure to the criminal justice system, and that due to increasing maturation. Importantly, these two measures are pulling in different directions: the first effect is positive, producing an escalation effect; and the second is negative, producing a de-escalation effect. This outcome is novel and interesting. The resulting conceptual framework helps to disentangle previously confusing results which sometimes seemed to suggest an escalation effect and sometimes a de-escalation effect.

In brief, if offenders have a large number of convictions over a short period of time, then experience will dominate maturation, and the overall effect will be one of escalation. Alternatively, if offenders have long periods without a conviction, then maturation wins out over experience, and the effect will be one of de-escalation. The parameter estimates identify that offenders with more than six convictions a year for age under 18, or two or more convictions a year for age 18 and over are likely to show escalation in their offending. If offenders have less than these amounts, then de-escalation is likely to take place. The model also allows that offenders can move from escalation to de-escalation through their criminal career as their frequency of convictions changes over the lifecourse.

The methodology used in this analysis is appropriate for assessing escalation effects from seriousness scores. First, the model allows for different sources of variation, both variation within offender, and variation between offenders, and can include between-offender random slope differences. Although there was no evidence of serial correlation in this data, the model also allows such correlation to be fitted if necessary. Finally, it allows to control for a wide range of time-varying and time constant covariates which affect seriousness.

There are some caveats to this present work. One potential disadvantage of measuring crime seriousness through court sentencing outcomes is that the severity of the sentence awarded for a specific offence may change because of plea bargaining. Thus an offender may agree to plead guilty in return for a reduction in sentence. Thomas (1978) notes that in the United States, plea bargaining is an open process, whereas in England and Wales, the plea bargaining process is more hidden, and is therefore under-researched. The likely impact of plea bargaining is therefore unknown, although it is unlikely to affect the seriousness scale strongly.

Another issue is that of time spent in prison. This has been shown to be an important potential confounder in trajectory research (Piquero et al., 2001) but one that is often neglected and difficult to collect (Piquero et al., 2003). In the context of this study, it might sensibly be argued that time spent in prison and not accumulating convictions is of a different nature to time spent in the outside world not accumulating convictions. An offender in prison might indeed be subject to influences from other offenders which may encourage escalation. Alternatively, training and education schemes in prison may well give the offender skills which will encourage de-escalation. This analysis has not controlled for street time as dates of release are not available, but in the following study this potential problem will be dealt through using an alternative measurement of total sentence length served in prison as an estimation of time spent in custody. The effect of imprisonment on risk of escalation in crime seriousness will be examined in the next chapter ('Study 3').

In the published paper of this study (Liu et al., 2011), one referee raised the issue as to whether the order of convictions variable really represents the effect of criminal justice experience on escalation. We agree that there may be other interpretations for this variable. For example, the number of times an offender is brought before a court could represent a lack of skill in avoiding arrest, or it may be related to low self-control which in turn would affect the frequency of offending. We do however maintain that the gaining of experience is an important component of this variable.

In the next chapter on 'Study 3', the focus of this research will be shifted more into methodological development based on this analysis. Such development includes: firstly, 'Study 3' will be assessing whether there are latent types of offenders who have different patterns on development of crime seriousness (heterogeneity in the underlying distribution of random effects). For example, the results in this study suggest that there is large variance in the random slope (conviction occasions). This may due to the existence of different types (unobserved) of developmental trajectories in crime seriousness among the population of the offenders. Secondly, 'Study 3' will be considering the impact of time spent in custody on the risk of escalation and will be testing two forms of non-linear age effects for better estimation of crime seriousness. More importantly, the successive study ('Study 3') will provide a comparison framework for comparing this current statistical approach with other alternatives.

Chapter 7

Study 3: Escalation in General Crime Seriousness II – mixture and mixed modelling approaches

7.1 Introduction

In the previous chapter, a *linear mixed-effects model* was used to model escalation in offence seriousness over the criminal lifespan. That work was innovative in taking a multi-level modelling approach to the sequence of seriousness scores from each offender. This statistical approach modelled sequences of seriousness scores with time-varying covariates and accounted for between individual and within individual variability. The study also identified that there are two temporal scales, age and conviction occasion, and so examined two types of escalation process – escalation associated with experience of the criminal justice process, and escalation associated with age and maturation. The resulting model suggested some interesting findings where ageing is associated with de-escalation whereas increasing conviction occasions (court appearances) are associated with escalation.

In this chapter, 'Study 3' is based on the initial study ('Study 2') but with further statistical developments in assessing general escalation. The current study develops further in various ways. The first direction of the development is related to statistical regression methodology. There was one potential criticism of 'Study 2' which the study did not consider: that there may be different subpopulations of offenders with different escalation processes. Statistically speaking, it did not allow heterogeneity in the assumption of the underlying distribution of random effects. In a previous chapter (Chapter 5), there are two types of alternative approaches in the study of developmental trajectory which can be used to address this problem.

Therefore, the main focus of the current study is to assess general escalation through all three types of introduced statistical approaches, which are linear mixed-effects model, group-based trajectory model, and growth mixture model. This study also provides a comparison framework which enables assessment of goodness-of-fit at both individual-level and marginal-level for each approach statistically. The same selection of explanatory variables are used for consistency of model comparison across different types of approaches. This work also attempts to examine the assumption of a multivariate normal distribution in the random effects based on linear mixed-effects model.

The second direction of the development is on the choice of covariates. In this chapter, some covariates differ from 'Study 2'. Firstly, this study considers all three temporal scales (see Chapter 2 section 2.4) together, which are age at conviction, conviction occasion, and time spent in prison. Criminologically speaking, it is important to understand how time spent in prison can affect offenders' seriousness in successive offending after they released from prison. This concept is also rarely examined through quantitative research due to difficulty of measuring the exact time spent in prison, or time at risk on conviction (street time). Although resources on information of the exact dates when in and released from prison are available from official authorities, there are still practical problems, such as public availability and data matching issues. Therefore, this study uses an accumulated total length of imprisonment prior to the current conviction as an alternative measurement of time spent in prison.

Secondly, as explained in 'Study 2', the main purpose of the previous study is an initial exploratory study on general escalation. Therefore the variables which were used there were more straightforward. However, this chapter allows various approaches of smoothing (non-linear) age effects on estimating crime seriousness. The well-known age-crime curve is therefore examined more thoroughly in this work through two different smoothing approaches. This is necessary to allow different types of offenders to develop their sequence of crimes with various age effects, rather than to assume everyone has the same trajectory over age at each conviction. Previously, the conviction occasion, which reflects some information on experience of going through criminal justice system, was assumed to have a linear effect on escalation in crime seriousness. In this chapter, this assumption is challenged by adding a quadratic term of conviction occasion. Moreover, two more new variables are constructed which are total number of conviction occasions and length of criminal careers. These two variables may give more explanatory power to the effect of escalation from the criminological point of view.

In brief, this current study is developed based on the previous study ('Study 2'). However, this work can be distinguished from the previous study by two main dimensions. Firstly, in terms of statistical methodology, it focuses on applying a mixture approach to assess latent types of offenders in the population who have different patterns of developmental trajectories in crime seriousness. Then it compares results across mixed-effects modelling approach and mixture modelling approaches on study of general escalation. Secondly, from the perspective of criminological theory in studying criminal careers, the potential impact of imprisonment on escalation is considered. Moreover, various approaches of non-linear age effects on escalation are also examined, in order to improve the accuracy of estimates of crime seriousness.

Therefore the structure of this chapter is: in section 7.2 it briefly summarises variables that were used in the previous study ('Study 2'), and introduces new variables of interest in the current study. In section 7.3, it shows the basic evidence of existence of heterogeneity in the underlying distribution of estimated of random effects, which suggests mixture modelling approaches are needed for assessing heterogeneity in the population of offenders. Then in section 7.4, all three competing methods: linear mixed-effects model, group-based trajectory model and growth mixture model are used to assess the existence of subpopulations. The following section (section 7.5) then compares the results from the three statistical approaches. Finally, in section 7.6, the modelling results will be summarised and some tentative substantive conclusions will be reached. In addition, methodologically, the advantage and disadvantage in using the mixed-effects approach and mixture regression approach for this particular study will be discussed.

7.2 Explanatory variables

Based on these current available methodologies, the research questions of interest are: (a) How to examine escalation trajectories at both population-level and individual-level? (b) How can the age at conviction (effect of maturation) affect escalation? (c) How can the order of convictions(effects of experience of going through the criminal justice system) affect escalation? (d) How can custodial sentence (time spent in prison) affect escalation? The time spent in prison is relatively rarely assessed in the criminology literature. In addition, variables of interest were extracted from the OI dataset (for detailed description about the dataset see Chapter 3 section 3.3.2) accordingly.

The present work is continuously examining general escalation in crime seriousness following Chapter 6. Explanatory variables thus include order of conviction, age at conviction, number of offences at each conviction occasion (logarithm transformed), gender and age at onset (description of each see Chapter 6 section 6.2) will be examined again through the alternative statistical approaches. Apart from those used variables, in this work, new variables including two forms of non-linear age effects, a quadratic term of order of conviction, cumulative custodial sentence length, total number of conviction occasions, and length of criminal careers are also examined to determine their effects on increasing/decreasing seriousness over conviction occasions. Descriptions of these variables are:

Custodial sentence This is cumulative custodial sentence length (in years) up to the previous conviction occasion for each offender. This provides information on cumulative time spent in prison prior to the current conviction occasion.

Age at conviction with two breakpoints In the previous study, one breakpoint at age 17 was selected according to the best goodness-of-fit (AIC and BIC) among a range of breakpoints from age 12 to 45. The idea was basically to allow non-linear age effects on the estimation of crime seriousness. Therefore, a similar strategy is adopted in the current work for searching over a two-dimensional grid. Then ages with two breakpoints at age 14 and 16 are chosen from various combinations of breakpoints (chosen from age between 12 to 45). The reason to use the two breakpoints in this study is that this may improve the accuracy of estimation in crime seriousness. Therefore three breakpoint dummy variables are constructed. The first takes the value of 0 to 4, and takes constant value 4 for all age above 14. The first dummy variable therefore indicates age at conviction between age 10 and 14, but subtracts 10 from it in order to ensure that the estimated intercept represents the mean seriousness at age 10. The second takes the value zero if the age at conviction either less than 15 or greater than 16, and $(age - 14)$ for age 15 and 16. Then the third takes the value zero if the age is less than 17, and $(age - 17)$ for age 17 and above. Therefore this set of three dummy variables fully describes the age information.

Non-parametric smoothing of age An alternative method for computing a non-linear age effect is to create a number of dummy variables through *natural cubic splines* (Hastie, 1993) based on the required degrees of freedom.

The concept of degrees of freedom (df) here is to use the value of *df* as suitably chosen quantiles of ages. For example, if $df = 4$, then the ages are divided by four equal quantiles (25%) with three breakpoints. The aim in using this alternative form of non-linear age effect is to seek an age effect which is smoother than the breakpoint approach and which may more closely represent the relationship between age at conviction and crime seriousness. Additionally, allowing non-linear age effects to vary across latent groups of offenders through a mixture modelling approach can also allow the developmental trajectories in crime seriousness over age to have different smooth paths within each latent type of offenders.

Quadratic order of conviction In the previous study, the relationship between order of conviction and crime seriousness is assumed simply linear. This chapter explores the potential existence of a non-linear relationship through a quadratic term of order of conviction.

Total number of conviction occasions The fact of whether offenders have large or small number of conviction occasions can affect the sensitivity of estimate of crime seriousness has been examined in 'Study 2'. In this chapter, the effect of total number of conviction occasions of each offender on escalation is examined statistically through regression models.

Length of criminal career The definition of length of criminal career is the total length (in year) between an offender's first conviction occasion to the last observed conviction occasion in this study. However, this length may not be the true length of some offenders' criminal careers for reasons, such as death of offenders, or unreported offending officially. The research question of interest here is whether an offender who has larger than average criminal career may have higher tendency to have high crime seriousness.

7.3 Assessing the nature of heterogeneity among offenders

Now in this section, the attention shifts back to the models which were outlined in Chapter 5 section 5.4, and discusses how the various alternatives might be chosen. The simplest approach – the linear mixed-effects model – is based on the assumption of multivariate normality of the random effects, but Verbeke and Lesaffre (1996) state that violation of this assumption may seriously influence the parameter estimates. Therefore, in this section, prior to any detailed modelling, this assumption will be assessed through graphical diagnostics of the fit of a basic linear mixed-effects model (including both random intercept and slope as well as controlling for other variables; see Model 1 in Table 7.2). In the following section, the other two alternative statistical models will be fitted, then results from each model can be assessed and compared.

In testing the multivariate normality of the estimated random effects ($\hat{u}_{i1}$ and $\hat{u}_{i2}$), a joint test proposed by Holgersson (2006) is applied in this study, which combines two graphical methods. The first graphical method is a *correlation scatterplot* of means against variances which are computed from the multivariate data, the second method is a *Q-Q plot* of Mahalanobis d^2 and chi-square distribution quantiles. The joint visual examination of the two graphs can provide a more robust test for detecting non-multivariate normality in situations when one graph fails to detect this but the other does. For example, the correlation scatterplot has the power to detect non-normality which the *Q-Q* plot cannot detect for simulated skewed normally distributed data. In contrast, for data which comes from a mixture of normals with the same mean but heterogenous variances, the *Q-Q* plot is likely to detect non-normality, whereas the correlation scatterplot supports normality. Therefore, the combination of these two tests are powerful graphical tools to detect non-normality.

The *correlation scatterplot* is defined in the following way. Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ be

n i.i.d. random variables, where $\boldsymbol{X}_j = (x_{j1}, ..., x_{jp})$ is a p-vector of realisations, with $j = 1, ..., n$. In this study, n is the total number of offenders, and $p = 2$ representing the estimated random intercept and estimated random slope for each offender. Let $\overline{\boldsymbol{X}} = (1/n)\sum_{j=1}^{n} \boldsymbol{X}_j$, where $\overline{\boldsymbol{X}}$ is a p-vector of means $\overline{\boldsymbol{X}} = (\overline{x}_1, ..., \overline{x}_p)$ and $\boldsymbol{S} = (1/n)\sum_{j=1}^{n}\sum(\boldsymbol{X}_j - \overline{\boldsymbol{X}})(\boldsymbol{X}_j - \overline{\boldsymbol{X}})'$, with $\boldsymbol{S} = (s_1, ..., s_p)$. If the n random variables are normally distributed then the value of $\boldsymbol{L}'\overline{\boldsymbol{X}}$ and $\boldsymbol{L}'\boldsymbol{S}\boldsymbol{L}$ are independent (Lukacs, 1942). Normally, either $\boldsymbol{L} = \boldsymbol{1}$ (i.e. a sum) or $\boldsymbol{L} = \boldsymbol{1}/\boldsymbol{p}$ (i.e. an average).

The $\boldsymbol{X}_j$ are multivariate normally distributed if and only if $\boldsymbol{L}'\overline{\boldsymbol{X}}$ and $\boldsymbol{L}'\boldsymbol{S}\boldsymbol{L}$ are independent. Therefore, M bootstrap samples of realisations from $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ can be computed and then the M paired values of $\boldsymbol{L}'\overline{\boldsymbol{X}}$ and $\boldsymbol{L}'\boldsymbol{S}\boldsymbol{L}$ can be calculated. If $\boldsymbol{X}_j$ is normally distributed then the scatterplot of $\boldsymbol{L}'\overline{\boldsymbol{X}}$ against $\boldsymbol{L}'\boldsymbol{S}\boldsymbol{L}$ should have no pattern of correlation.

The second graphical tool is the Q-Q plot of Mahalanobis distance d^2 (Mahalanobis, 1936) and is given by:

$$d_j^2 = (\boldsymbol{X}_j - \overline{\boldsymbol{X}})'\boldsymbol{S}^{-1}(\boldsymbol{X}_j - \overline{\boldsymbol{X}}) \tag{7.1}$$

Given that $\boldsymbol{X}_j$ is i.i.d. normally distributed, then the d^2 measures are chi-square distributed. Therefore, the basic idea of this Q-Q plot of the distance d^2 is to display the graph of the chi-square distribution quantiles $Q_p(\frac{j}{n+1})$ against d_j^2 which should display an approximately straight line on the diagonal if the data is multivariate normal.

These two graphical methods are applied to the data in this study, in order to test multivariate normality on the distribution of estimated random slope (the order of conviction) and random intercept from the linear mixed-effects model presented in the next section (see Model 1 in Table 7.2). Then 400 bootstrapped samples are taken from the estimated random effects ($\hat{u}_{i1}$ and $\hat{u}_{i2}$) which were obtained from this mixed-effects model, and the values of $\boldsymbol{L}'\overline{\boldsymbol{X}}$ and $\boldsymbol{L}'\boldsymbol{S}\boldsymbol{L}$ were computed, taking $L = \boldsymbol{1}$. The 400 paired statistics are graphed in Figure 7.1(a). It

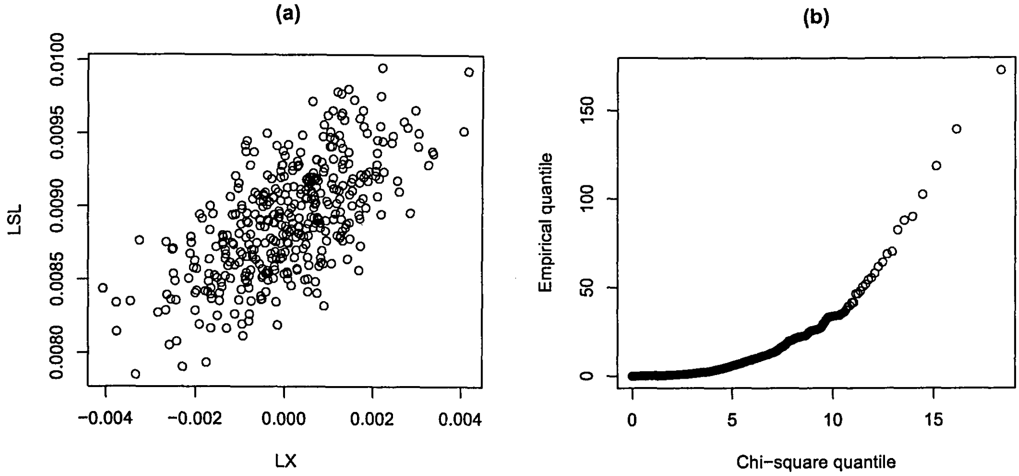

Figure 7.1: (a) The scatterplot for $\boldsymbol{L}'\overline{\boldsymbol{X}}$ vs. $\boldsymbol{L}'\boldsymbol{S}\boldsymbol{L}$ of 400 bootstrapping samples from estimated random effects. (b) Q-Q plot of Mahalanobis D^2 vs. quantiles of χ^2_2.

clearly suggests that there is a strong linear correlation between the means ($\boldsymbol{L}'\overline{\boldsymbol{X}}$) and variances ($\boldsymbol{L}'\boldsymbol{S}\boldsymbol{L}$) of the joint distribution of estimated random intercept and slope. As the variance is increasing with the mean, the plot rejects the assumption of multivariate normality. Figure 7.1(b) shows the Malahanobis Q-Q plot. It shows a curvilinear relationship rather than the expected straight line, which suggests that heterogeneity of the random effects is present with structure arising from a mixture of normals (Holgersson, 2006).

In summary, both the scatterplot and the Q-Q plot suggest that the joint distribution of estimated random intercept and slope does not follow a bivariate normal. Therefore a note of warning is needed. The work by Verbeke and Lesaffre (1996) states that the test of heterogeneity on random effects is fundamentally difficult as both the random intercept and slopes are already estimated under the multivariate normality assumption. Therefore, the estimates of the random effects may be biased if this assumption is wrong. In addition, Verbeke and Molenberghs (2000) suggest that Q-Q plots of the type suggested by Lange and Ryan (1989) cannot differentiate a wrong distributional assumption for the random effects or the error terms from a wrong choice of covariates. However Eberly and Thacheray (2005) suggest that in the presence of a correctly specified mean model, the normality test of Lange and Ryan (1989) detected non-normal random effect distributions with reasonable power that increased as the non-normality grew more pronounced. In the presence of a misspecified mean model, they go on to state that such plots are more useful as a general diagnostic procedure. In this work, it is concluded that there is sufficient evidence from these plots to justify the investigation of heterogeneity in more detail.

Therefore, in the next stage, it is necessary to apply both types of mixture modelling approaches to investigate the heterogeneity in the population of offenders and to identify potential latent types of offender in terms of their development of seriousness in crime.

7.4 Statistical modelling results

There are three stages in the following analyses. Firstly, it is necessary to identify the effect of covariates as either class-specific (with different parameter estimates in each class) or class-independent (with the same estimates in each class) effects. The primary interest in this analysis is in identifying any potential differences in the effects of age and criminal justice experience between classes. Therefore the age and the order of conviction are considered as class-specific covariates. The number of offences at each conviction occasion, gender, custodial sentence length, total number of conviction occasions, and length of career are tested both as class-specific covariates and also as class-independent covariates.

Secondly, as identified in section 7.2, two forms of non-linear age effects are of interest in this work. They both are considered as class-specific covariates. The two forms of non-linear age effects will be assessed and compared through AIC and BIC by applying to a set of linear mixed-effects models and a set of growth mixture models.

Thirdly, the three statistical models which are described in Chapter 5 section 5.4 will be applied, using the covariates described section 7.2, and trying two,

Table 7.1: AIC and BIC values for different forms of non-linear age effects for the linear mixed-effects model (LME model), the group-based trajectory model (GBTM) with two classes and the growth mixture model (GMM) with two classes.

Model: age with break-point(s)						
Break-points	LME model		GBTM (2 classes)		GMM (2 classes)	
	BIC	AIC	BIC	AIC	BIC	AIC
1	51079.28	50989.34	49251.88	49167.61	48515.76	48405.55
2	51009.84	50911.72	48756.28	48659.04	48027.06	47903.88
Model: age with natural splines						
df=2	51176.51	51086.56	49576.80	49492.52	48782.41	48672.20
df=3	51059.89	50961.77	49281.72	49184.47	48561.30	48438.13
df=4	51055.31	50949.01	49278.05	49167.84	48532.55	48396.41
df=5	51057.89	50943.42	49271.75	49148.58	48544.19	48395.09
df=6	51035.13	50912.48	49207.35	49071.21	48467.70	48305.63
df=7	51019.65	50888.82	48973.51	48824.41	48244.65	48069.62
df=8	51014.28	50875.27	48833.65	48671.58	48124.26	47936.26
df=9	51022.68	50875.50	48824.47	48649.44	48128.94	47927.97

three, and four class models for the mixture based approaches. Then in the next section, the three statistical models of their final models are compared in terms of their goodness-of-fit.

7.4.1 Choice of non-linear age effect

The two forms of non-linearity were compared by fitting the three different types of statistical model under consideration – namely, the linear mixed-effects model, the group-based trajectory model (fitted with two classes) and the growth mixture model (fitted with two classes). Then their AICs and BICs were compared. Table 7.1 shows the three models with one age breakpoint (see 'Study 2'), two age breakpoints, and a range of different degrees of smoothing age effects through the natural cubic spline function. The models also controlled for the class-independent covariates of gender, sentence length and the log of the number of offences at each conviction occasion, and the class-specific effect of order of conviction.

Table 7.1 gives the AIC and BIC values for various forms of age non-linearity under the LME model, the GBTM (two classes) and the GMM with two classes.

Firstly AICs and BICs are examined for the breakpoint results. As mentioned before with one breakpoint, the break was estimated at age 18 for both models, whereas with two breakpoints, the breaks for both models were estimated at age 14 and age 16. Moreover, all the AIC and BIC values suggest that for the three models the age effect with two breakpoints is better than with one breakpoint.

Examination of the natural cubic splines is carried out with a varying number of degrees of freedom from 2 to 9. Firstly, for the linear mixed-effects model, both AIC and BIC have a decreasing trend with increasing degrees of freedom but both reach a minimum at 8 degrees of freedom. The decrease is not strictly monotone, however, and the BIC results exhibit the multimodality behaviour described by Rosenberg et al. (2003). For the second model – GBTM (two classes), both AIC and BIC values have a general decreasing trend with increasing degrees of freedom and have not reached a minimum at 9 degrees of freedom. Finally, for the GMM model, the results are less clear, but in this case a minimum is reached for BIC at eight degrees of freedom, whereas the AIC continues to decrease and has not reached a minimum.

In comparing the fit of the best cubic spline model to the best breakpoint model, the BIC from the LME model (51009.84), and both BIC and AIC from the GMM model (48027.06 and 47903.88) all suggest that the two breakpoint model is preferred to the best spline (df=8) model. The BIC from the GBTM two breakpoint model (48756.28) is smaller than the cubic spline model with 9 degrees of freedom (48824.47), also suggesting that the two breakpoint model is preferred to the spline (df=9) model. Only the AICs for both the LME model and GBTM appear to suggest that the spline model is to be preferred. For the case of the GBTM, although a minimum of AIC/BIC may be reached at some large degrees of freedom ($df > 9$), still a model with fewer parameters but similar goodness-of-fit is preferred. Therefore, in the following statistical analyses, a non-linear age effect with two breakpoints is used consistently to model the seriousness of crime.

Table 7.2: Model 1: The linear mixed-effects modelling.

Fixed Effects:	Coef.	S.E
Intercept	4.137	0.026*
Order of conviction	0.006	0.001*
Sex(female)	-0.135	0.017*
Age at conviction:		
≤ 14	0.021	0.009*
$15 - 16$	-0.144	0.009*
17+	-0.011	0.001*
log(No. of offences)	0.273	0.008*
custodial sentence	0.008	0.003*
Random Effects (Var):		
Intercept (v_{11})	0.0404	
Order of conviction (v_{22})	0.0003	
Covariance (v_{12})	-0.0027	
Residual (τ^2):	0.3819	0.003
BIC (AIC):	51009.84	(50911.72)

* indicates significance at the 5% level.

7.4.2 Three statistical models for criminal career escalation

Linear mixed-effects model

The first model to be considered is the *linear mixed-effects model* and was fitted by using the *nlme* package in R. The parameter estimates are shown in Table 7.2. The results differ slightly from those shown in the previous study Table 6.2; there are two reasons for this. Firstly, an additional covariate of cumulative custodial sentence length has been included, and this model allows a greater degree of non-linearity for age with two breakpoints rather than one. Secondly, this model includes a new constructed variable – length of criminal careers.

Neither the quadratic term of order of conviction is significant (with a P-vale of 0.2036), nor the effect of total number of conviction occasions (with a P-vale of 0.6070). The effect of length of criminal careers has very little impact on escalation with coefficient of 0.001, although it is statistically significant (P-value=0.014). Therefore, the linear mixed-effects model which is presented in Table 7.2 focuses on effects of order of conviction, age, gender, number of offences at each conviction occasion (logarithm), and cumulative custodial sentence.

The model shows that offenders who are younger than 14 are more like to escalate while increasing in age. However, after 14, age plays a major effect for de-escalation, especially from age 15 to 16, when offenders decrease their expected crime seriousness by 0.144. The effect of cumulative years spent in prison has a positive value 0.008, indicating the effect of escalation in crime seriousness. However, it is not a practically significant effect. For instance, given a cumulative 100 year time in prison, the effect of escalation in seriousness is 0.8. Therefore, although the model suggests that custodial sentence is statistically significant, practically they are insignificant.

Apart from these new explanatory variables, the other effects are still similar with the model in the previous study. In brief, on average female offenders were convicted of less serious offences than males. The major age effect (after age 14) and number of convictions (experience) are pulling in different directions. De-escalation with increasing maturation, escalation with increasing experience. Offenders, who are convicted a large number of offences within a single conviction occasion, are more likely to have a high serious crime.

Group-based trajectory model

The second model is the *group-based trajectory model* (GBTM) which focuses on assessing the existence of latent types of developmental trajectories. As described in section 6.2, this model allows for heterogeneity in the underlying distribution of random effects which were estimated from the previous model (Model 1) by assuming a discrete mass-point distribution for the random effects rather than bivariate normality. Effectively, this means that this model allows for latent groups with different intercepts and different covariate effects for age and order. To fit the model,the `R` package *lcmm* is used which assumes a common residual variance across classes. The MPLUS `TWOLEVEL MIXTURE` command can be used to relax this

Table 7.3: Model 2-C2: Group-based trajectory model with two-class solution.

	Class 1 (90.6%) Coef.	S.E	Class 2 (9.4%) Coef.	S.E
Intercept	4.186	0.031*	5.292	0.103*
Order of conviction	0.0004	0.001	-0.007	0.003*
Age at conviction:				
≤ 14	-0.013	0.009	0.358	0.037*
$15-16$	-0.012	0.009	-1.315	0.039*
17+	-0.013	< 0.001*	0.021	0.003*
Common effect:	Coef.		S.E	
Sex(female)	-0.150		0.015*	
log(offences)	0.265		0.007*	
Custodial sentence	0.013		0.002*	
Residual Var (τ^2):	0.5928		0.003	
BIC (AIC):	48756.28 (48659.04)			

* indicates significance at the 5% level.

assumption, but for consistency of numerical method and likelihood calculation among the three types of statistical modelling approaches, the *lcmm* package is preferred for the GBTM analysis.

Again, the quadratic term of order of conviction and the effect of total number of conviction occasions are not significant. Similarly, the effect of length of criminal careers has very little effect (< 0.001) on escalation. The effect of gender, number of offences at each conviction occasion (logarithm), and custodial sentence are examined as class-specific variables, but with no statistical evidence of significance. Therefore, the GBTM with two, three, and four classes solutions are examined in the following. These are controlled with order of conviction and age as class-specific variables, and gender, number of offences, and custodial sentence as class-independent variables.

Firstly, the model (Model 2-C2) with two classes is fitted and is shown in Table 7.3. The effects of gender, number of offences (logarithmic), and custodial sentence were treated as common effects across both classes. In contrast, the intercept, order of conviction (slope), and age with two breakpoints were allowed to vary across the two classes. The results from looking at the common effects show that gender, number of offences, and custodial sentence have similar effects to Model 1 (Table 7.2).

The results suggests that the majority (90.6%) of offenders belong to class 1. Note that the proportion of class membership which presented in this table is the average of estimated posterior class probability [1] of each individual. The intercept, which represents the average seriousness at age 10, is estimated at 4.2. Neither the conviction occasion nor the age between 10 and 16 have effects on escalation/de-escalation (having non-significant p-values). The offenders show some evidence of de-escalation (by 0.013) by age.

On the other hand, class two consists of a relatively small proportion (9.4%) of offenders from the population, with a relative higher intercept at age 10 (about 5.3). This group of offenders has an unexpected negative value (-0.007) for order of conviction, indicating de-escalation with increasing experience which is in contrast to 'Study 2'. Although this is a significant effect, the absolute value is quite small compared to the effects of age. Moreover, these offenders have a strong positive age effect (0.358) showing escalation while they are aged between 10 to 14, and followed by a strong negative age effect (-1.315) showing de-escalation between ages 15 and 16. The effect of age from 17 onwards, however, is again positive (0.021) showing some tendency to escalate with increasing age.

Therefore, the two-class GBTM (Model 2-C2) showed that the majority of offenders neither escalate nor de-escalate over conviction occasions, although there is some tendency to de-escalate after age 16. In contrast, the interpretation for the 9.4% of offenders shows contradictory results especially for showing an escalation effect after age 16, and de-escalation effect with increasing criminal experience. Thus, the three-class GBTM is examined for more clarification.

The three-class GBTM solution is named as Model 2-C3 and is shown in Table 7.4. Firstly, in terms of the common effects, the estimates of gender, number of

[1] The posterior probability is the probability of each individual belongs to certain class k given data $\boldsymbol{X}$, $P(c_i = k \mid \boldsymbol{X}_{it})$.

Table 7.4: Model 2-C3: Group-based trajectory model with three-class solution.

	Class 1 (87.1%) Coef.	S.E	Class 2 (5.2%) Coef.	S.E	Class 3 (7.7%) Coef.	S.E
Intercept	4.099	0.032*	6.535	0.133*	4.340	0.093*
Order of conviction	0.001	0.001	-0.035	0.002*	0.013	0.003*
Age at conviction:						
≤ 14	-0.002	0.009	-0.371	0.040*	0.718	0.035*
$15 - 16$	0.002	0.009	-0.531	0.051*	-1.533	0.042*
17+	-0.013	< 0.001*	0.068	0.004*	-0.019	0.003*
Common effect:	Coef.			S.E		
Sex(female)	-0.139			0.014*		
log(offences)	0.267			0.007*		
Custodial sentence	0.010			0.002*		
Residual Var (τ^2):	0.5767			0.002		
BIC (AIC):	47941.28 (47805.14)					

* indicates significance at the 5% level.

offences, and custodial sentence are still relatively consistent with Model 1 (Table 7.2). Importantly, however, the three-class model has produced some interesting findings when the class-specific estimates are examined.

In the three-class solution, class 1 (87.1%) is identified as being similar to the old class 1 in the two-class model. The majority of offenders who belong to this class (class 1) have relatively constant seriousness over both conviction occasions and age when under 17, with all estimates non-significant. As before, the model shows a small and significant tendency to de-escalate with increasing age from age 17 onwards.

Class 2 (5.2%) is identified as a new class which is formed from a relatively small subset of offenders. This group of offenders have a very high intercept of 6.54 (representing the average seriousness at age 10), but with a tendency to de-escalate with experience (estimate of -0.035) and with strong de-escalation with age between ages 10 to 16. The effect of age after 16 is positive (0.068) which is around double the absolute value of the experience effect (0.035).

Although class 3 (7.7%) is identified as being similar to the old class 2 in the two-class model, the estimates of order (0.013) and age from 17 onwards (-0.019) are changed to have the opposite signs. The reason is that a small group of offenders have been extracted from the old group 2 and assigned to the new class (class 2). The new class 3 is telling a similar story to Model 1, which suggests that de-escalation occurs with increasing experience and increasing age from age 15 onwards, but also shows increasing seriousness between the ages of 10 to 14 (0.718).

A model with a four-class solution has also been attempted. Although both the AIC and BIC are smaller, suggesting a better goodness-of-fit, the interpretation of the class-specific parameter estimates are far less clear. The four-class solution is basically, splitting class 3 from the three-class model into two even smaller groups with very similar directional effects of the class-specific estimates in the two new groups, but with different magnitudes. Since the dataset used in this study consists of 4,831 offenders, then the AIC and BIC may not reach their minima until a large number of classes have been fitted. Recent guidance suggests that it is important to stop at a meaningful model with a smaller number of classes rather than searching for the best AIC/BIC with larger number of classes which is less interpretable (Nagin and Tremblay, 2005). Therefore, the GBTM with three-classes is preferred in this study.

Previously, the interpretations of both two-class and three-class GBTM solutions have been described. Following, some descriptive statistics are given in Table 7.5, which summarise some characteristics of offenders. This table presents the offenders as a whole and the various groups which were identified in both the GBTM two-class and the GBTM three-class solutions.

Firstly, from the aspect of overall offenders in this study, there are in total 4,831 offenders, which 89% of them are male offenders. Their average age at onset (the first conviction occasion) is at age 17 with standard deviance 5.21. Their average length of criminal career is about 12 years with substantial variation ($sd = 9.42$). Offenders on average have five conviction occasions, but the number of conviction

Table 7.5: Descriptive statistics are used, which include frequency, proportion, and mean (standard deviance) of lists of variables, to illustrate certain aspects of the characteristics in the identified various subgroups by group-based trajectory models.

	Two classes GBTM		Three classes GBTM			Total No. of offenders
	1	2	1	2	3	
No. of offenders:	4,573	258	4,520	166	145	4,831
% of males:	88%	93%	88%	93%	96%	89%
Average age at onset:	17(5.19)	15(4.96)	18(5.20)	16(5.44)	13(1.76)	17(5.21)
Average length of criminal careers:	12(9.37)	16(9.64)	12(9.34)	15(9.13)	16(10.54)	12(9.42)
Total No. of conviction occasions:	5(6.00)	8(6.39)	5(5.89)	8(8.20)	8(6.93)	5(6.05)

occasion varies substantially ($sd = 6.05$) from offender to offender as well.

Secondly, by looking at offenders at class-level, each offender is assigned to a certain group according to the estimated posterior probability through GBTM. Under the two-class solution, there are 4,573 (94.7%) offenders who are classified in class one. This proportion is larger than the size of class one in Table 7.3 (90.6%), as explained previously, this is due to the size of class membership (presented in Table 7.3 and Table 7.4) which is computed from the averaged estimate of individual posterior probability. The proportion of male offenders is higher in class two (93%) than in class one (88%). The average age at onset is younger in class two (15 years old) than in class one (17 years old), but the average length of criminal careers is longer in class two (16 years) than in class one (12 years). On average, there is also more total number of conviction occasions (8 conviction occasions) in class two than in class one. Additionally, there is substantial variance in both average length of criminal careers and the total number of conviction occasions, especially in the latter one.

Under the three-class solution, class two and class three have similar size, with 166 offenders in class two and 145 in class three. The total size of class one (4,520 offenders) is slightly smaller than class one in the two-class solution (4,573). It seems that there is higher proportion of male offenders in class two (93%) and class three (96%) than in class one (88%). Therefore, a further effect of sex examined at class-level through three-class GBTM, but the effects of gender on crime seriousness at each class are very similar. Therefore, it is reasonable to assume that the effect of gender is constant over classes.

In terms of the average age at onset, it seems that offenders have younger average age at onset in class three and class two than in class one. Especially in class three, not only the average age of onset is very young (13 years old) but also with relatively very small variance ($sd = 1.76$). It seems to suggest that offenders who are convicted of their first offence at similar young age are likely to escalate with increasing experience and de-escalate while aging. However, statistically there is no significant age onset effect on crime seriousness. For both the average length of criminal careers and the total number of conviction occasions, the pattern in class two and class three are quite similar and also consistent with the values from class two in the two-class GBTM solution.

Growth mixture model

The final approach is the *growth mixture model* (GMM). This GMM as described in section 2, allows estimation of class-specific random effects. The GMM extends the GBTM, with discrete random effects allowing members who belong to the same class to have individual intercepts and slopes. Thus, in contrast with the GBTM, the GMM estimates class-specific variances of each random slope and random intercept. The random effects distribution is therefore assumed as a mixture of multivariate normals.

The model is fitted through the *lcmm* package in R. The two-class solution (Model 3-C2) is shown in Table 7.6. There are two classes of offenders, consisting of a large first class with 92% of offenders (class 1) and a smaller second class with the remainder (class 2). For both classes the intercepts give the estimated average seriousness at age 10. This average seriousness is higher in the second class (5.036) than the first (4.154), although it needs to be remembered that most offenders do not start their offending careers until age 14 to 16. In class one the effects of age

Table 7.6: Model 3-C2: Growth mixture modelling with two-class solution.

	Class 1 (92.5%) Coef.	S.E	Class 2 (7.5%) Coef.	S.E
Intercept	4.154	0.029*	5.036	0.124*
Order of conviction	0.013	0.002*	0.003	0.027
Age at conviction:				
≤ 14	-0.018	0.008*	0.460	0.035*
$15 - 16$	-0.017	0.009*	-1.341	0.039*
17+	-0.013	0.001*	0.009	0.005
Random Effects (Var):				
Intercept (v_{11})	0.0467		2.5031	
Order of conviction (v_{22})	0.0117		0.6252	
Covariance (v_{12})	-0.0212		-1.1354	
Common effect:	Coef.		S.E	
Sex(female)	-0.128		0.015*	
log(offences)	0.253		0.007*	
Custodial sentence	-0.003		0.003	
Residual Var (τ^2):	0.5580		0.003	
BIC (AIC):	48027.05 (47903.88)			

* indicates significance at the 5% level.

over the two breakpoints are very similar (-0.018, -0.017 and -0.013) and show a near constant declining slope with age. In contrast, the offenders in this class increase crime seriousness with experience (0.013 for each extra conviction).

The statement of choice of variables in GMM here is in fact the same as in previous GBTM. There are differences in the interpretation of parameters. Class 2 consists of a small proportion (8%) of offenders who start with relative higher seriousness. Moreover, those offenders have an increasing age effect (0.460) showing escalation while they are aged between 10 to 14, and followed by a strong declining age effect (-1.341) showing de-escalation between ages 15 and 16. The effect of age from age 17 onwards then becomes small (0.009) with a non-significant p-value. Compared to class 1, class 2 also suggests substantial variation within offenders exists, with variances of 2.5031 and 0.6252 for the random intercept and random slope respectively. This compares with the far smaller variances of 0.0467 and 0.0117 for class one.

Table 7.7: Model 3-C3: Growth mixture modelling with three-class solution.

	Class 1 (88%)		Class 2 (6.4%)		Class 3 (5.6%)	
	Coef.	S.E	Coef.	S.E	Coef.	S.E
Intercept	4.109	0.030*	7.622	0.1117*	3.764	0.185*
Order of conviction	0.013	0.002*	0.006	0.004	0.104	0.075
Age at conviction:						
≤ 14	-0.021	0.009*	-0.560	0.037*	0.812	0.041*
$15 - 16$	0.004	0.009	-0.645	0.042*	-1.412	0.046*
17+	-0.013	0.001*	-0.016	0.003*	0.001	0.009
Random Effects (Var):						
Intercept (v_{11})	0.0566		0.0221		3.5397	
Order of conviction (v_{22})	0.0160		0.0062		0.9991	
Covariance (v_{12})	-0.0275		-0.0107		-0.9143	
Common effect:	Coef.			S.E		
Sex(female)	-0.123			0.014*		
log(offences)	0.248			0.007*		
Custodial sentence	0.0002			0.007		
Residual Var (τ^2):	0.5485			0.003		
BIC (AIC):	47371.76 (47203.20)					

* indicates significance at the 5% level.

The class-independent effects show that females have a significantly lower crime seriousness score compared to males (-0.128) and that the effect of time spent in custody is small and non-significant (-0.003). As with the linear mixed-effects model, the effect of the number of offences per conviction shows the expected positive estimate (0.253). Next a model with three-class solution explores whether additional class is needed.

The three-class GMM solution is named as Model 3-C3 and is shown in Table 7.7. In terms of the common effects, the estimates of gender, and number of offences are similar with the two-class GMM solution (Model 3-C2). The effect of custodial sentence is also small and non-significant. The residual variance is also about the same (0.5485). However, this three-class model has suggested some interesting insights when age and conviction escalation patterns within each class are examined.

In the three-class model, class 1 (88%) and class 3 (5.6%) are identified as being similar to the old class 1 and class 2 in the two-class model. Class 2 is identified as a new class (6.4%) which is formed from a small subset of offenders from the original class 1 and class 2 in the two-class model (Model 3-C2 in Table 7.6). This group of offenders have a very high estimate of seriousness at age 10 (7.622), but de-escalate over all three age periods. The de-escalation is strongest up to age 16, and then becomes smaller for those aged 17 and older (-0.016). In contrast with the strong de-escalation with age, no significant effect of conviction order is found in this class. This class also has the smallest variation among individual's intercepts (0.0221) and slopes (0.0062).

The model estimates for the other two classes also shows very interesting findings. Class 3 (which is similar to class 2 in the two-class model), contains 5.6% of offenders. The estimate for the intercept which gives the mean seriousness level at age 10 is about one unit lower (3.764). Escalation by age shows a similar pattern, with escalation between age 10 and 14, de-escalation from 15 to 16 and then little change thereafter. The coefficients however are larger with escalation increasing by 0.812 for each year of age between 10 and 14, and decreasing by 1.412 units for age 15 and 16. The effect of the order of convictions is large and positive (0.104) but not statistically significant, showing increasing escalation with the number of distinct convictions.

One potential problem that should be pointed out is that there is substantial variation within offenders in this group of offenders, with variances of 3.5397 and 0.991 for the random intercept and random slope respectively. Ideally, a standard error of each random effect should be provided as an indication of robustness of the estimates of random effects through the growth mixture models. Although there are not standard errors of random effects computed in the R package of either nlme and lcmm, still the computation of the standard error can be calculated through the *bootstrap* (Efron and Tibshirani, 1994). However, in practice, this computation is very hard due to the long modelling time required for each bootsrap sample. For instance, a 3-class model, say Model 3-C3 which is shown in Table 7.7,

Table 7.8: Descriptive statistics are used, which include frequency, proportion, and mean (standard deviance) of lists of variables, to illustrate certain aspects of the characteristics in the identified various subgroups by growth mixture models.

	Two classes GMM		Three classes GMM			Total No.
	1	2	1	2	3	of offenders
No. of offenders:	4,540	291	4,496	109	226	4,831
% of males:	89%	92%	88%	95%	92%	89%
Average age at onset:	17(5.19)	16(5.19)	18(5.19)	13(1.40)	16(5.50)	17(5.21)
Average length of criminal careers:	12(9.39)	14(9.72)	12(9.33)	18(10.71)	14(9.61)	12(9.42)
Total No. of conviction occasions:	6(6.15)	6(5.49)	5(5.96)	10(8.89)	6(5.18)	5(6.05)

needs approximate 25 hours in R using a computer with two dual-core Opteron 2Ghz processors and 16Gb of RAM. Therefore, if only 200 boostrap samples are taken from the total 4,831 offenders, then at least 200 days are needed for the computation of the standard errors for random effects. It is also arguable that whether the number of 200 boostrap samples are efficient enough for a dataset with more than four thousands of observations.

Finally class 1 in the three-class model consists of the majority of offenders (88%). The intercept of 4.109 lies between the other two intercepts. Members of this class are generally de-escalating with age and escalating with their experience, although the age effect between age 15 and 16 while positive (0.004) is not statistically significant. The variances of the random effects in this class are also small (0.0566 and 0.0160 for the intercept and slope respectively).

A model with four-class solution is also attempted in the analysis of GMM. Similarly, both the AIC and BIC were smaller from the four-class solution than the three-class solution, but again the interpretation of members of classes are less clear with practical means. The four-class solution is basically, splitting the class 3 in the three-class model into two even smaller groups with no clear interpretation of either groups. Therefore, the GMM with three-class model is preferred in this study.

Again, the same descriptive statistics are examined for both the two-class GMM and the three-class GMM solution. For the three-class GMM solution, there are more offenders classified in class three (226 offenders) than in class two (109 offenders), where in the previous GBTM three-class solution (Table 7.5), these two classes were quite evenly sized. Again, like the result in GBTM three-class model, there is a higher proportion of male offenders in both class two (95%) and class three (92%) than in class one (88%). However, the patterns of the other variables are telling a different story in using GMM than the previous GBTM.

The picture of the three other variables are more complicated in the three-class solution. Firstly, the average age at onset is younger in class two (13) and class three (16) than class one (18). However, the figures from class two and class three are telling a contradicting story than the previous result in Table 7.5. The offenders in class two on average have the youngest age at onset (13 years old) and with small variation ($sd = 1.40$), rather than those offenders classified in class three by GBTM three-class solution. To add, the interpretations of the class two and the class three are in fact similar from both GBTM and GMM three-class regressions. Therefore, this reflects that the classifications of some offenders are not consistent over the two types of mixture regression approaches. This will be further discussed in section 7.6.

Moreover, offenders who belong to class two have particularly long average length of criminal careers (18 years) and large averaged total number of conviction occasions (10 convictions). Again, there is substantial variance in criminal careers length ($sd = 10.71$) and number of conviction occasions ($sd = 8.89$). It therefore seems to suggest that those offenders who have average high seriousness at age 10 are more likely to have average longer criminal careers. Although, as mentioned earlier in this chapter, the length of criminal careers had very little impact (coefficient<0.001) on increasing the crime seriousness.

It is important to emphasise here that, the purpose of this work is to assess the heterogeneity in the process of offence escalation among offenders through two types of mixture regression approaches (GBTM and GMM). Therefore, the interpretation of each regression model focuses on the characteristics of each type of identified offender in terms of their effect on changing crime seriousness. Although such descriptive statistics are provided in both Table 7.5 and Table 7.8, that only to provide some extra information from different aspect of offenders. Therefore it can not reflect whether these variables are significant factors to the effect of escalation.

In the following, this study compares among the three approaches (the two mixture approaches and the mixed-effects approach) in their goodness-of-fit graphically and statistically, which has been rarely done in the literature for this type of work.

7.5 Comparison the goodness-of-fit of the three statistical models

In the previous section 7.4.2, various modelling results are explained and compared in terms of the estimates of coefficients and AICs and BICs, and the difference in class classification between the two mixture approaches. Therefore, in this section, further examination of differences among the three statistical methodologies in terms of their goodness-of-fit is needed. Firstly, the goodness-of-fit of the three models (the LME model, the GBTM three-class model, the GMM three-class model) will be assessed through graphical tools by comparing the differences between the observed scores and the estimated scores at both marginal-level and individual cases. Then diagnostic measures such as AIC/BIC, and the Euclidean distance are used to compare the three models.

7.5.1 Graphical goodness-of-fit at marginal-level

Firstly, the goodness-of-fit of the LME model, the three-class GBTM, and the three-class GMM are examined at the marginal level within each class. The class membership which has been estimated from the three-class GMM model (Model

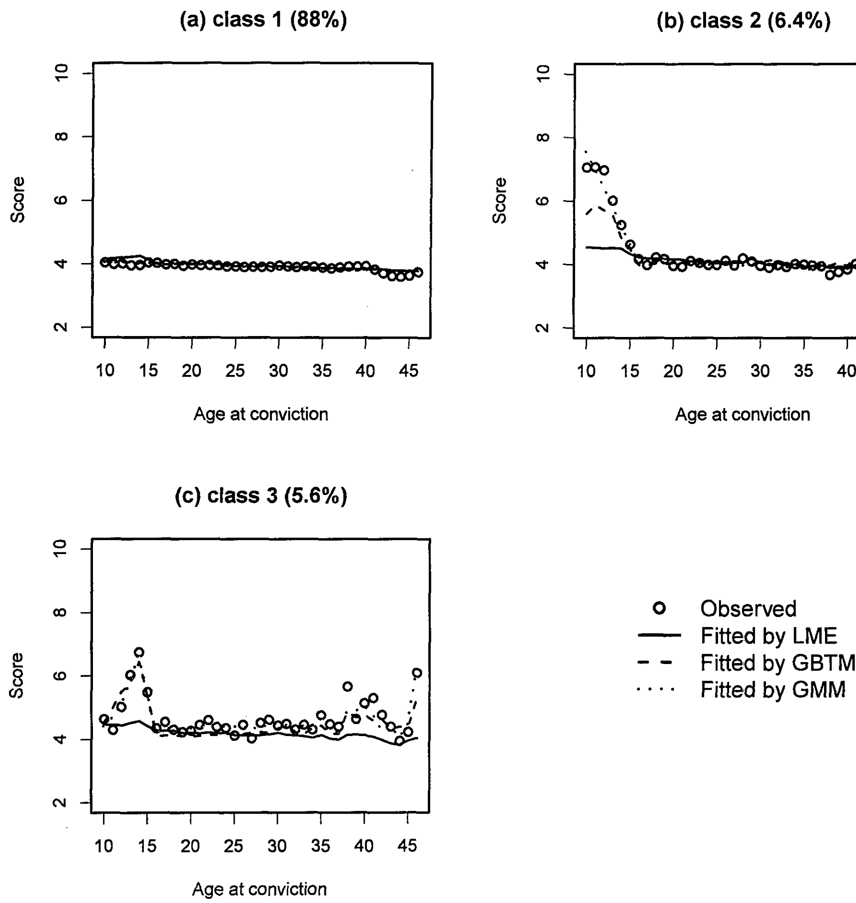

Figure 7.2: Comparison of the observed marginal seriousness scores and the estimated mean scores for the three models plotted against age at conviction. Offenders have been grouped into three classes by assigned class membership according to Model 3-C3. Plot (a): for offenders who are classified in class 1; Plot (b): for offenders who are classified in class 2; Plot (c): for offenders who are classified in class 3.

3-C3) is assigned to each individual offender. The marginal means of observed seriousness scores and predicted scores from the LME model, the GBTM approach, the GMM approach are computed for each age of conviction and for each of the three classes. The reason to look at the marginal seriousness scores by age at conviction is to be able to present graphs of the marginal crime seriousness effects within the three groups by age, as the age escalation effects differ strongly between the groups. The plots of the observed scores and the fitted scores against age at conviction are shown in Figure 7.2. It is important to clarify that in Figure 7.2 different offenders will contribute to each observed mean point, as each offender has a different set of conviction ages.

Firstly, the character of each class is examined by looking at the *observed* mean scores. It is clearly shown that for class 1 – Plot (a) – the majority (88%) of offenders stay relatively constant in their crime severity, but with a small tendency to de-escalate with increasing age. Class 2 consists of 6.4% of offenders who, if they offend in early adolescence, will start with expected high seriousness in offence, then de-escalate quickly between the ages of 14 to 16 followed by a gentle de-escalation at later ages. In contrast, class 3 shows remarkable diversity in crime seriousness especially between age 10 to 16 and from age 35 onwards. This group seems to consist of groups of offender either involved with serious crimes at earlier age (between age 10 to 15), or late onset offenders with quite serious offences, or even those offenders who were most delinquent with high serious crimes at both a early age and from late 30s onwards.

Secondly, the differences in *fitted* marginal means among the fitted three models are examined for each class. There is hardly any difference in class 1 between the three models. However, for the more complex offending patterns found in class 2 and class 3, the differences among the three methods are starting to show. On average, for both class 2 and class 3, estimates from the GMM appear to capture the more serious crimes more accurately and also can fit the observed mean more smoothly than the GBTM, and certainly better than the LME model, although the estimates from the GBTM also follow the mean observed trajectories well compared with the mean estimated scores from the LME model.

7.5.2 Graphical goodness-of-fit for individual cases

From looking at Figure 7.2, a clear story of the characteristics of each class has been observed, and some general marginal goodness-of-fit diagnostics have been presented. Therefore, the next step is to examine the *individual* offenders' trajectories in crime seriousness and their fitted values over conviction occasions. For graphical examination of individual goodness-of-fit, two scales along the x-axis

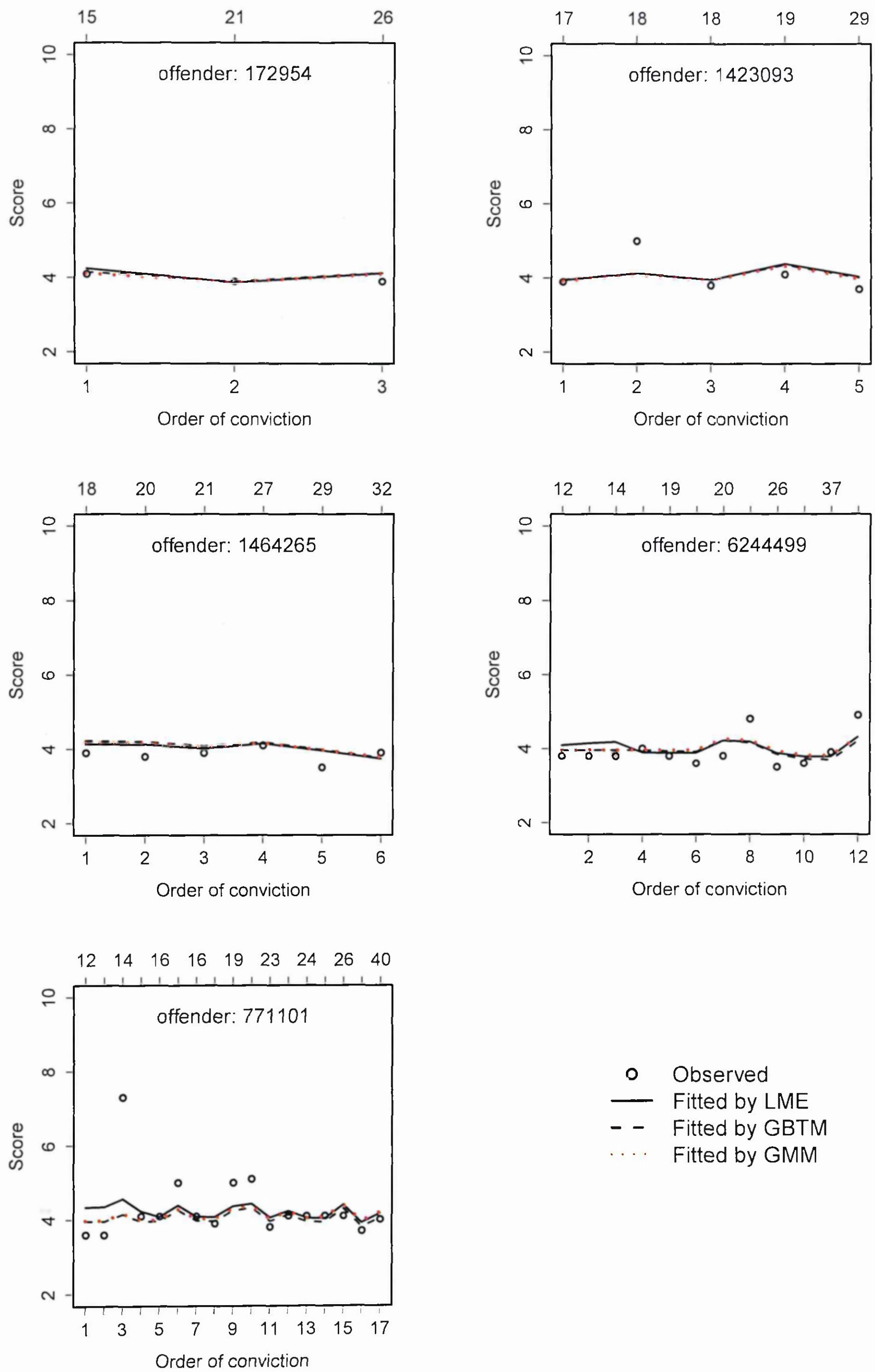

Figure 7.3: Comparison of the observed seriousness scores and estimated scores for the three models for five cases with varying number of convictions (labelled with offenders' id) in class 1 (Model 3-C3), plotted against order of conviction (bottom axis) and age (top axis).

are given. The order of conviction is shown on the bottom axis, and the age at each conviction is presented on the top x-axis. The plots are prepared as follows. First a random sample of around 100 individuals within each class is taken. Then, within each class, five offenders were selected who represent some common offending patterns from these samples, and also represent the range of total number of convictions. Graphical output from class 1 is shown in Figure 7.3, Figure 7.4 shows cases from class 2, and Figure 7.5 shows cases from class 3.

First, examination of the five individual offenders from class 1 is undertaken (shown in Figure 7.3). As described before, class 1 consists of the majority of offenders who are relatively stable in their seriousness in crime. In comparing the fitted models for the five offenders, similar findings are found to those given by the marginal plot diagnostics in Figure 7.2(a) - namely that the three models give very similar estimates. The bottom-left plot (offender `771101`) indicates that complex offending patterns will cause difficulty for any models. Basically, offender `771101` is active in offending from age 12 to 40, with the seriousness of most of his offending at about 4.0 but with a few irregular episodes of high seriousness offending in between. The sudden changes of severity in such a case can not be captured accurately by any of the three models. It is possible that this type of offending may needs its own small latent class which is not represented in the three group solution.

Five individual offenders from the second class (class 2) are now shown in Figure 7.4. As mentioned previously, class 2 consists of offenders with high seriousness at early ages but de-escalating with increasing age, and also escalating with increasing experience. For this class, estimates from both the GMM and the GBTM are a better fit than the LME model. In particular, the GMM captures the high seriousness at the beginning of each trajectory better than the other two models, and adjusts better for changing crime severity, such as in the middle right plot (offender `1139875`).

In Figure 7.4, a problematic example is also shown. Offender (`692880`) had

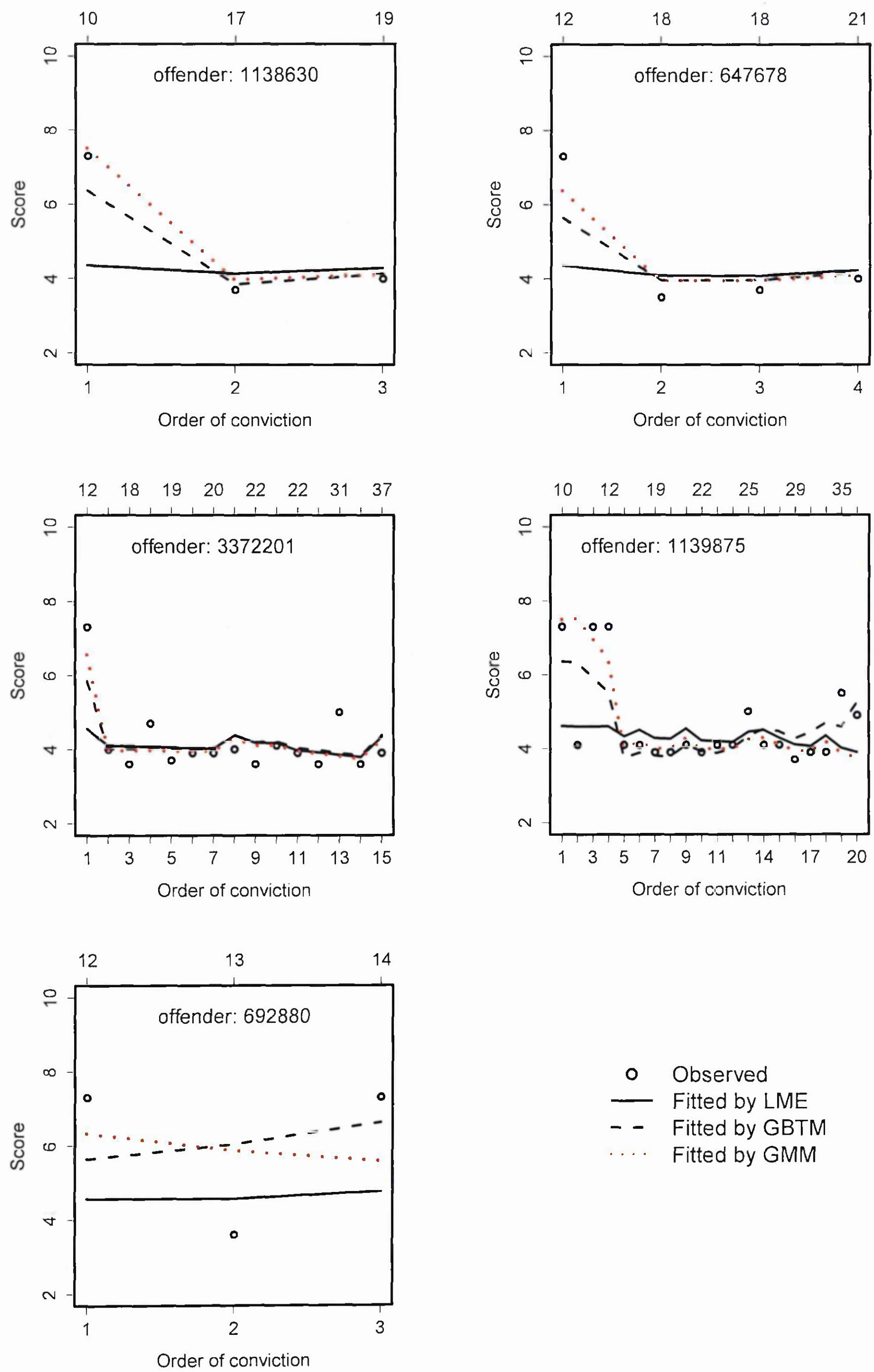

Figure 7.4: Comparison of the observed seriousness scores and estimated scores for the three models for five cases with varying number of convictions (labelled with offenders' id) in class 2 (Model 3-C3), plotted against order of conviction (bottom axis) and age (top axis).

one conviction occasion per year between age 12 to 14. This offender had a less serious offence in between two convictions of high seriousness crime (about 7.6) – aggravated burglary according to the original dataset. The estimates from the GMM captures the high starting point, then shows de-escalation thereafter. In contrast, the GBTM captures the increasing seriousness of the last offence, then with an estimate of positive slope before. In contrast, the LME model is taking an average of the seriousness of the three convictions and giving a nearly constant estimate of the seriousness over the convictions. The analysis of such a pattern is statistically difficult especially without controlling for any extra information of offending at each conviction occasion.

Finally, five offenders from class 3 are examined in Figure 7.5. Offenders in this class are generally more diverse in terms of their range of crime seriousness. The sequence of crime seriousness of both offender 3131743 and offender 672840 in particular are estimated quite accurately by the GMM method, and also the estimates for offender 1117309 well present the trend of the trajectory. Thus, the conclusion is the same as for class 2, with the GMM method performing more sensitively than the other two models. For this particular group of offenders, the common analytical issue is the sudden occurrence of the occasional high serious crime as part of the criminal history which occur more often in this class than for the offenders in the other two classes. This is represented in the model by the high estimates of v_{11} and v_{22}. Particularly, the offender 1049924 had a conviction of murder at age 42, although the previous two convictions seemed to suggest the start of a tendency to de-escalate. The offender 4245369, also had two occasions of high seriousness crimes in between relatively low serious crimes.

7.5.3 Comparison of goodness-of-fit by diagnostic measures

Finally, two forms of diagnostic measure are used to assess the goodness-of-fit for the three modelling approaches. The first method proceeds by examining AICs/BICs from various models which were fitted in section 6.5.2. The AIC/BIC

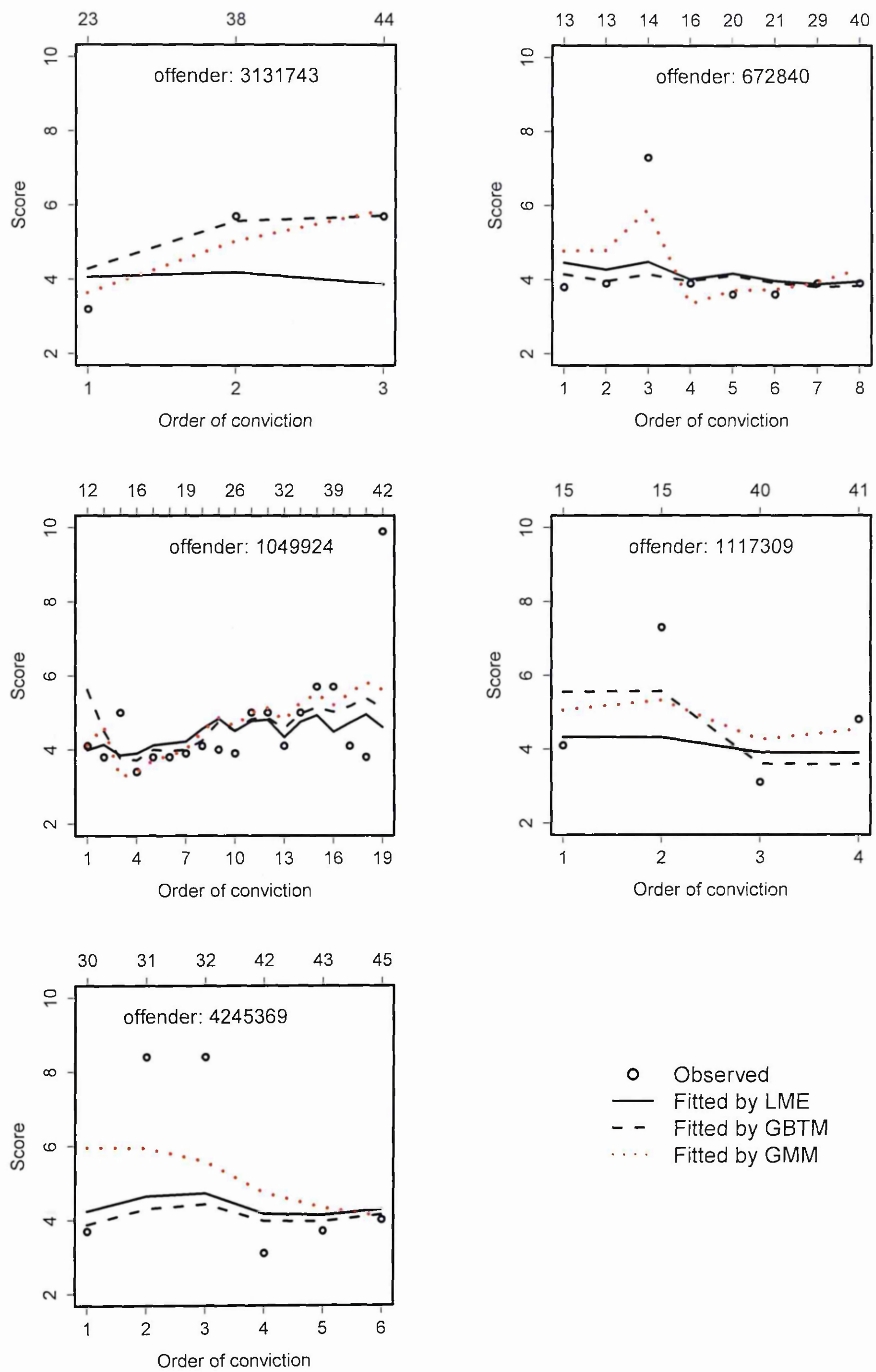

Figure 7.5: Comparison of the observed seriousness scores and estimated scores for the three models for five cases with varying number of convictions (labelled with offenders' id) in class 3 (Model 3-C3), plotted against order of conviction (bottom axis) and age (top axis).

Table 7.9: AIC and BIC values for various models by the LME model, the GBTM with two/three classes and the GMM with two/three classes.

	BIC	AIC
LME	51009.84	50911.72
GBTM-2CLASS	48756.28	48659.04
GBTM-3CLASS	47941.28	47805.14
GMM-2CLASS	48027.05	47903.88
GMM-3CLASS	47371.76	47203.20

Table 7.10: The Euclidean distance measures for the LME model, the GBTM with three classes and the GMM with three classes.

	Class 1	Class 2	Class 3
LME	0.8315	3.1530	3.3433
GBTM-3CLASS	0.8435	2.5619	2.1615
GMM-3CLASS	0.8036	2.3212	1.8410

values have been shown previously as part of the results for each model, but now they are grouped together and summarised in Table 7.9. It clearly shows that both the GBTM and the GMM with two or three classes have smaller BICs/AICs than the LME model, indicating better goodness-of-fit by using a mixture approach than the straightforward LME model. Moreover, in terms of the difference between the two mixture modelling approaches, the GMM two-class model has smaller BIC/AIC (48027.05/47903.88) than the GBTM two-class model (48756.28/48659.04), and similarly the GMM three-class model also has smaller BIC/AIC (47371.76/47203.20) than the GBTM three-class model (47941.28/47805.14). Therefore, the GMM approach is preferred as the best fitting model in this current dataset.

The second diagnostic measure which is used to examine the goodness-of-fit is the *Euclidean distance.* The Euclidean distance is a mathematical term which used to measure the "ordinary" distance between two points or sequences, and it is defined as follows:

$$D_{ik}(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i) = \sqrt{(y_{i1} - \hat{y}_{i1})^2 + ... + (y_{in_i} - \hat{y}_{in_i})^2}, \tag{7.2}$$

where $\boldsymbol{y}_i$ is a vector of an observed sequence of seriousness scores for offender i, with length n_i, and the vector of estimated scores is given by $\hat{\boldsymbol{y}}_i$. Within each offender i, the difference between each observed and estimated score at each conviction occasion is squared first and these squared differences are then summed together and the square root taken. Therefore, the Euclidean distance, which measures the distance between the observed sequence and the fitted sequence, can be computed. The average Euclidean distances by class (assigned membership according to Model 3-C3) for each fitted model are then shown in Table 7.10.

Firstly, class 1 shows that all three Euclidean distance measurements from the three models are nearly identical. In fact, the GBTM three-class model has a slightly greater average distance than the LME model, indicating the LME model fits the data slightly better the GBTM for class 1. In general, out of the three models, the GMM three-class model fits all three classes the best, with the smallest distance for class 1 (0.8036), class 2 (2.3212), and class 3 (1.8410). In addition, the two mixture modelling approaches (the GBTM three-class and the GMM three-class) have improved the goodness-of-fit substantially for class 2 and class 3.

7.6 Conclusion and discussion

This study has attempted to assess the existence of heterogeneity in the population of offenders in terms of their seriousness of crimes. Therefore two mixture modelling approaches are used, which are the group-based trajectory model, and the growth mixture model. Both mixture approaches identified three latent groups of offenders in the population. The study did not stop at the interpretation of each model. It also provided a comparison framework to compare models from both the linear mixed-effects model and the two mixture approaches. This comparison was carried out in three stages, firstly by looking at goodness-of-fit graphically at the marginal-level, then individual cases were selected for more detailed graphical examination. Finally two forms of diagnostic measure are used to compare the goodness-of-fit for the three modelling approaches as confirmation of the previous graphical methods. Therefore, based on all the previous analyses, the results can be summarised in the following points:

1. Three types of (latent) offenders are identified by both the GBTM and the GMM. Although increasing the number of classes will still reduce the values of AIC/BIC for either approach, the three-class solution is preferred in this study as it provides a clear practical meaning and distinct characteristics for each type of offender.

2. The comparison of goodness-of-fit by using both the graphical investigation and diagnostic measures suggests that the growth mixture model with three classes is preferred as the best model among all another fitted models. In more detail, the GMM three-class model captures the occurrence of high seriousness crimes more accurately and follows the sequence of crime seriousness scores more smoothly compared to both the GBTM and the LME model.

3. According to the GMM three-class model, the majority (88%) of offenders (class 1) are in fact quite stable in their seriousness in crimes. They are more likely to be specialised in terms of seriousness score, and therefore they may engage predominantly in only one offence or a group of closely related offence types, which is the classic meaning of 'specialists' by Blumstein's definition (Blumstein et al., 1986a). This group of offenders has some tendency to escalate with increasing experience and de-escalate with increasing age.

4. According to the GMM three-class model, about 6.4% of offenders belong to class 2. This group of offenders tends to have a high intercept (7.6) at age 10, and has a strong de-escalation effect with age, especially between the ages of 10 to 16.

5. Class 3 which is identified from the GMM three-class model consists of 5.6% of offenders. Offenders in this group show more diversity in crime seriousness,

and also are involved with more high seriousness crimes. These offenders on average tend to show a higher degree of escalation between ages 10 to 14 (by 0.812 per year) than offenders in the other two groups. They then de-escalate quickly at age 15 and 16 (by 1.412 by year). From age 17 onwards, they become more stable in their crime seriousness.

6. The three statistical modelling approaches all suggest that male offenders on average are more likely to be convicted of more serious offences than female offenders; in addition the larger the number of offences involved within a single conviction occasion, the higher the seriousness level in this conviction occasion will be. In contrast, the effect of custodial sentence varies from model to model. However, models with a statistically significant custodial sentence effect all show a small and positive effect, indicating that offenders escalate with increasing time spent in prison, but the effects are small, with changes of 0.01 of a seriousness score point or less per year. For the preferred GMM three-class model, in particular, the effect of length of custodial sentence is not significant.

This work also contributes to some important policy implications on how to identify and selectively target a small group of potentially dangerous offenders. In general, most offenders in this sample are more likely to be involved with similar types of crimes with similar crime seriousness as this study showed. Moreover, offenders who started with a relatively high seriousness crime at an early age have a tendency to de-escalate with age. For those offenders, policy implications from the previous study ('Study 2') can be applied to this study too, which is that it is important for criminal justice professionals to focus on persistent offenders – those with large numbers of convictions in a short period of time – as these individuals are most likely to escalate. This work importantly also identifies a group of offenders with high diversity and high seriousness in crime, and for this type offender, monitoring is needed as they are generalists in offending and more likely to be involved in high seriousness crimes in between offences than the other two types of offenders.

Some issues which have been raised in this particular study need discussing. One issue is to do with reliability (or robustness) in mixture regression approaches. In this work, two potential hazards which relate to the robustness have been noticed while using the group-based trajectory model and the growth mixture model.

The first hazard regards the classification disagreement in class two and class three of the three-class GBTM solution (Table 7.5) and the three-class GMM solution (Table 7.8). The number of offenders who belong to class two has shrunk in GMM (106) compared to GBTM (166). In contrast, the number of offenders who belong to class three has enlarged in GMM (226) compared to GBTM (145). There are in fact, 67 out of 166 offenders (40%) who had belonged to class two in GBTM but classified in class three in GMM. Only 48 out of 166 offenders (29%) stayed in class two in both GBTM three-class and GMM three-class. Section 7.5 has proven graphically and statistically that the GMM three-class solution is a better fit at both marginally and individually than the GBTM three-class solution. However, the results indicate that there is a potential misclassification problem in using a mixture type of approach.

The second hazard regards, the substantial variances of the random intercept and random slope in class three which are identified by the GMM three-class solution. The potential problem is that some members of group three could have intercepts which are higher than ten (the maximum value of the seriousness scale), or could have large coefficient for slopes which can shoot outside the range of seriousness score (0 and 10). This problem can be avoided in a further analysis through using censored normal distribution for the response variable (seriousness score). This way can at least limit the estimated values in the boundaries between zero and ten. However, this issue also relates to the question of robustness in mixture regression approach.

The large variation may imply that either the assumption of a normal distribution made for the random intercept and slope in this class is incorrect, or that a larger number of class is needed. Theoretically, applying different distribution assumption in random effects have been done, such as a generalised mixed-effects model (Diggle et al., 2002). However, in the literature no work has been done which can allow a mix of different distribution assumptions in the random effects across latent classes.

Alternatively, it is possible to search a larger number of class as for the best of goodness-of-fit statistically. For example there are two recent studies in the area of 'testing the number of classes in mixture modelling' (Nylund et al., 2007; Lo et al., 2001). In particular, the simulation study by Nylund et al. (2007) compared traditionally used Information Criteria (ICs) including AIC, BIC and two types of likelihood ratio test (LRT). The two types of likelihood ratio test are Lo-Mendell-Rubin (LMR) LRT (Lo et al., 2001) and boostrap LRT (McLachlan and Peel, 2000). They conclude that over three types of mixture regression approach, the boostrap LRT (BLRT) gave the best performance in terms of consistency, and BIC is the second best. Note that although BLRT outperformed the other CIs, the computational cost of this method is very heavy.

In this study, the potential problems of large variance of random effects in class three may be reduced through searching a larger number of classes. However, the study contains a large number of offenders (4,831) with at least two observed conviction occasions, while one offender had 137 conviction occasions. One issue is that the model cannot find the best BIC easily as a large number of classes may be needed. Additionally, with such crime data, as illustrated previously, there are occasional extremely serious crimes or a rare pattern of offending behaviours. Therefore a very small number of offenders, who can not be identified in any other groups, may form a separate group together, or even an individual can form a group with only one offender if the ideal goodness-of-fit is used.

So far, two particular issues from class three in GMM three-class solution have been discussed. A more general question is whether such mixture regression approaches are indeed needed, especially in comparison with the traditional mixed-effects model for dealing with repeated observations over time. The second issue which has been raised in this particular study is whether such mixture approaches have more explanatory power than the linear mixed-effects model.

In terms of this particular study, the answer to this question is that there is statistical evidence of heterogeneity in the estimated random effects (section 7.3). Moreover, all the diagnostics of goodness-of-fit in section 7.5 are supporting the growth mixture model with three-class solution as the best model across all offenders and within each identified group. The selection of the best model has followed practical guidelines (Nagin and Tremblay, 2005), which is to search a model with the best BIC/AIC and with most meaningful interpretation. To clarify, searching for a model with a practical interpretation does not imply that the mixture approaches have more explanatory power than the linear mixed-effects model. In my personal point of view, social scientists may be in favour of mixture approaches, as it can provide some practical meanings to associate with theories in their disciplines. However, from statistical point of view, my interest to apply mixture models is to explore the hidden heterogeneity in the distribution of random effects by the traditional mixed-effect approach, which has not been paid much attention in the research using mixed-effects models.

Therefore, my argument for using either the linear mixed-effects model or the mixture regression model for this particular study is that the mixture regression can explain the hidden heterogeneities in the mixed-effects regression approach further, but caution is required to avoid over-interpreting the mixture models. The more general way to answer the question of whether to use mixed-effects model or mixture regression models is through consideration of advantages and disadvantages of each types of approach (Chapter 5 section 5.4). It is essential to ask what can suit each research question the best.

Chapter 8

Conclusions and Discussion

The motivation for this thesis was to assess whether the seriousness of offending for a typical offender increases, remains stable or decreases over time. The thesis has hopefully helped to demonstrate the complexity of this area, namely the study of escalation in crime seriousness. The purpose of this chapter is to draw together the key findings and themes that have emerged from the various studies presented above. In particular, the policy implications of this thesis in relation to the various models utilised will be discussed. Finally, this chapter will discuss the value and contribution of the thesis as a whole, and will discuss the limitations and potential areas of future development.

8.1 Summary of earlier conclusions

As would be expected, criminal careers vary substantially between individual offenders. However, some general findings can be made on the basis of the analyses in this thesis.

In the first instance, results relating to gender effects on escalation and crime seriousness have been considered. All three studies have shown that, as was expected, male offenders on average are more likely to be convicted of more serious crimes than female offenders. For example, for offenders who had been convicted of arson, blackmail, kidnapping, or threats to kill, the risk of subsequent homicide conviction for those male offenders is doubled compared to female offenders. For general escalation, the average seriousness score of male offenders is higher than female offenders by between 0.070 to 0.135, with the exact value depending on the particular modelling strategy.

Secondly, results relating to age, order of conviction (conviction occasions), and the time spent in prison – the three temporal scales – on the risk of escalation are summarised as follows. According to the two studies on general escalation ('Study 2' and 'Study 3'), this thesis has found that the process of escalation/de-escalation in crime seriousness depends on a combination of the effect of maturity (age) and the experience of going through the criminal justice system (conviction occasions). Therefore, changing crime severity at the individual level is not driven by only one single temporal process. This thesis has proposed a third temporal scale, namely the time spent in prison. The results suggest that the greater the time spent in prison the higher the average crime seriousness for the next offence. Although this effect is statistically significant, it is not considered practically significant because of its very small coefficient.

According to the results from 'Study 2' and 'Study 3', the effect of age and experience are pulling in different directions. The first effect is negative and produces a de-escalation, whilst the second effect is positive and produces an escalation. If an offender has a large number of convictions over a short period of time, then experience dominates maturity, and the overall effect will be a tendency to escalation. Alternatively, if an offender has a long period without a conviction, then maturity wins out over experience, and the overall effect will be a tendency to de-escalation.

In addition, 'Study 3' suggests that there is a tendency to escalate with increasing age when the offender is between 10 to 14 years old. This may be a specific effect which relates to the 1953 cohort dataset from the Offenders Index (OI) used in this thesis. The reason might be that the children who were born in the early 1950s were starting to become involved in gang culture, specifically the mods and rockers phenomenon which started in the mid-1960s. As a result, the 1953 cohort children may have been more likely to have been convicted of a serious crime at an early age. This is a possible explanation for this specific effect of escalation while increasing age between 10 to 14. In addition, 'Study 3' suggested that the majority of offenders are relatively specialised in their seriousness, i.e. their seriousness level stays constant rather than escalating or de-escalating. Similarly, 'Study 1' suggested that offenders who had committed any of the four focus offences were more likely to commit the same focus offence again, rather than to commit one of the other focus offences, or to escalate to homicide – a different form of specialisation.

Thirdly, results which are related to prior criminal history on the risk of escalation and crime seriousness are summarised as follows:

1. In terms of the effect of crime type on escalation, 'Study 1' identifies that offenders who had been convicted of arson, blackmail, threats to kill, or kidnapping, their likelihood of being convicted of homicide are much higher (one in 100 for kidnapping offenders, one in 200 for the three other types) than the likelihood of a male member of the general population over a 20-year period (one in $3,000$). Moreover, kidnapping offenders double the risk of subsequent homicide conviction than offenders who had been convicted of blackmail, arson, or threats to kill.

2. In terms of crime mix, 'Study 1' found that offenders who were involved with two or more of the focus offences double the chance of escalating to homicide than those offenders who had been involved with one single focus offence.

3. In terms of frequency of criminal activities, 'Study 1' shows that the risk of a homicide conviction from the four focus offences increases with increasing numbers of previous convictions. 'Study 2' and 'Study 3' show that the greater the number of offences reported at each conviction occasion, the higher the average crime seriousness will be at this conviction occasion.

8.2 Policy implications

In this thesis, the study of escalation has important policy implications. It is important for criminal justice professionals to use their limited resources to target a specific small group of offenders who have the potential to commit more serious offences. Information which can help them to identify and selectively target a small group of potential dangerous offenders will be useful in making decisions in practice.

Therefore, this thesis has resonances with the recent body of work on redemption, which looks at the likelihood of conviction or arrest after significant crime free periods (Kurlychek et al., 2007; Blumstein and Nakamura, 2009; Soothill and Francis, 2009) and on intermittency, where criminal careers restart after a period of non-offending (Piquero, 2004). The research findings in this work have suggested that a long conviction-free period over a number of years will lower the seriousness of any future crime which may be committed, and so such individuals will demonstrate partial redemption. This implies in turn that sentencing guidelines may wish to exclude the effect of prior convictions if there is a sizeable gap in time between the prior offence and the current offence.

Moreover, Le Blanc and Loeber (1998) make the important point that de-escalation is one of the three elements of desistance (with the other elements being deceleration and increasing specialisation). This thesis identifies those with infrequent convictions as being more likely to experience de-escalation as they get older and, therefore, according to Le Blanc and Loeber (1998), on the path to desistance. Intervention may need to be focused on persistent offenders – those with large numbers of convictions in a short period of time – as these individuals are most likely to escalate. The attention of the criminal justice system should focus less on those convicted less often, as they are likely to de-escalate naturally.

This thesis also identified a small group of offenders with high diversity and high seriousness in crime, and suggests that for this type offender, they are more likely to be involved in high seriousness crimes than other offenders in the population.

Therefore, monitoring is needed for such offenders.

Finally, this thesis looked at those offenders who had committed some specific serious offence, such as arson, blackmail, kidnapping, or threats to kill, and suggested important policy implications for monitoring such groups of offenders. Offenders with such serious offences but also with multiple types of serious offences could be registered in a "serious offenders" database. Although such a serious offenders database now exists in the United Kingdom (known as the Violent and Sex Offender Register ViSOR), it is presently restricted to violent and sex offenders. Other types of offenders such as those committed of arson, blackmail, threats to kill, and kidnapping could potentially be included.

8.3 Discussion

8.3.1 Contribution of this thesis

As mentioned at the beginning of this thesis, quantitative research on the escalation of "criminal careers" has not been well developed. In particular, there has been relatively little longitudinal analysis into an individual offender's development of crime seriousness. Therefore, this thesis used the Offenders Index to explore escalation in crime seriousness at both population-level and individual-level. For offenders who started their first conviction in 1963, their criminal histories can be followed and examined over more than 30 years.

This thesis takes a modern statistical approach and has identified more appropriate modelling strategies than those used previously. Survival analysis has been adopted in the analysis of 'serious offender escalation'. Therefore, the time from a target serious offence to a homicide conviction can be examined and compared between subgroups of offenders. These subgroups include, for example, offenders with different first time convictions, age, gender and prior criminal history.

The common shortfalls of the previous work in the study of escalation are: they had only looked at escalation at the population-level, but not at the individual-level. Therefore, the tendency to escalation/de-escalation was observed marginally. In the study of general escalation, modern regression approaches such as linear mixed-effects models, the group-based trajectory models, and the growth mixture models are applied to examine development of crime seriousness both for individual offenders and among offenders. The linear mixed-effects model is designed to study repeated observations not only at population-levels but also to allow variation at the individual-level through random effects. The group-based trajectory model subsequently relaxes the multivariate normal distribution assumption on the random effects (in mixed-effects models) through a discrete number of latent subpopulations. Further more, the growth mixture model allows individual differences within each estimated latent subpopulation.

Each study in this thesis also examined a number of research areas that have been rarely studied in previous work. 'Study 1', for instance, examined escalation in crime seriousness for low-frequency but high-tariff offences, since there has been much less criminological focus on such crime types. 'Study 1' also attempted to look at the inter-relationships among those crimes, as normally only one type of crime is studied in this research domain (Soothill et al., 2002). 'Study 3' compared three different statistical approaches to the study of general escalation. This work provided a comparison framework to compare the differences between the three statistical modelling approaches through both graphical investigation and statistical diagnostic measures. Methodologically, there seems to have been an increasing interest in the use of growth mixture models, although such comparative research has been rarely carried out to date.

More importantly, in the study of general escalation, this thesis has proposed that three temporal scales are important in examining escalation in crime seriousness. Previous work in the study of escalation only looked at changing crime seriousness over only one of the temporal scales (i.e. either age or conviction occasions) but did not consider both together in their analyses. By contrast, this thesis has attempted to examine escalation/de-escalation in crime seriousness by looking at the combination effects of age, conviction occasions, and time spent in prison. The results show that the effect of age (maturation) and the effect of conviction occasions, which can partially reflect the experience of going through the criminal justice system, both act together on the process of escalation/de-escalation. This discovery can partially explain the reason why in the literature there is mixed evidence of escalation (mentioned in Chapter 2 section 2.1). For instance, some studies suggested evidence of escalation for juvenile offenders (Wolfgang et al., 1972) and other evidence of general escalation from age 10 to age 25 by Le Blanc and Frèchette (1989), but also evidence of de-escalation for adults (Moitra, 1981; Blumstein et al., 1985). On the other hand, other studies have found no evidence of escalation, such as studies from Datesman and Aickin (1984) and Sheldon et al. (1987).

In this thesis, the interpretation of conviction occasions refers to the experience of going through the criminal justice system. The conviction number can be thought as an experiential process, where offenders gain expertise and knowledge of criminality as the number of previous contacts with the criminal justice system increases. In relation to criminological theory, the study of the experience of going through the criminal justice system is essentially linked with 'labelling theory'. Empirical evidence for changes of subsequent offending after an interaction with criminal justice system is contentious (Soothill et al., 2009, page 104). Therefore, this thesis has attempted to partially examine the experience effect on escalation by using conviction occasions.

Another issue is the time spent in incarceration. This has been shown to be an important potential confounder in trajectory research (Piquero et al., 2001) but one that is often neglected and difficult to collect (Piquero et al., 2003). Von Hirsch (1986) also highlighted some work on selective incapacitation which aims to direct incapacitation to those most at risk of committing future dangerous crimes. Von Hirsch (1986) states that the claims of earlier work in this area have been exaggerated, and that there has been relatively little good research on the effect of prison to reduce the severity of future crime. This thesis has also examined the effect of time spent in prison, and has shown that it has very little impact on the risk of escalation.

8.3.2 Future development

This thesis has discussed the advantages and disadvantages of using a dataset of *official records.* In short, one needs to recognize the limitations of using the OI data, such as an underestimation of crime rates, potential misclassification errors from the initial reported offence into a different category of conviction type, lack of information on the offenders' social background information and so on.

Therefore, this work acknowledges the need to replicate the analysis using other criminal history datasets and in other jurisdictions. For example, by comparing more recent cohorts with the 1953 cohort, it may tease out the cohort effect on escalation. Similarly, using data from different countries may show different patterns which affect the crime seriousness and escalation/de-escalation.

There are also a number of potential research areas in terms of statistical methodology development. For example, in 'Study 2', the study of general escalation in crime seriousness was examined through the linear mixed-effects model. It was used to estimate both individual and population trends in seriousness over the order of conviction. However, as mentioned in the relevant chapter, there is a normality assumption on the underlying distribution of the offenders' intercepts (the first time crime seriousness) and their slopes (changing over convictions). Therefore, one potential development would be to apply different types of distributions which can better estimate the underlying distribution of random effects, such as the gamma distribution.

Furthermore, the measurement of the crime seriousness scale used in the study of general escalation is continuous from 0 to 10. Therefore, it would be sensible to use a truncated normal distribution which restricts the range of estimated seriousness to the range 0 to 10. Although the issue of over-large or over-small fitted values was not a problem in this thesis, it is a potential problem when increasing the number of latent groups (mixture approach), since this can produce a group of offenders with a negative intercept and a large positive slope, or with large positive intercept and a large negative slope. This may subsequently lead to fitted values outside the range of 0 to 10.

Additional research is still required on the study of general escalation using the mixture approach. For example, this thesis compared three competing modelling approaches statistically and identified three types of offender according to their offending patterns. Offenders belonging to each class may share some common crime patterns in terms of the specific types of offences involved. Although some simple descriptive statistics are provided for each subgroup of offenders, a future study can focus on an examination of each class of offender by considering various features in more detail, such as age at onset, type of first crime, sequence of crimes, length of criminal career, and diversity of offending. Another potential area of development relates to the need to develop better searching methods for a model with a larger number of classes with more practical meaning (perhaps allowing the detection of classes with a small number of cases) and the need to identify different subsets of offenders with more distinct characteristics along with their seriousness scores.

Finally, although certainly not the least challenge, this thesis recognizes that the majority of offenders are more likely to stay within a small range of crime seriousness, with a tendency to escalate or de-escalate slightly according to their age and experience. However, the analysis has identified a small group of offenders who have been involved with more serious offences at different points in time. The diversity of offending can be very different among such offenders and more difficult to capture statistically. Therefore, further analysis of such offenders is needed and alternative statistical modelling approaches may be required to handle the complexity of the underlying structure of offending behaviour. For example, the mixture modelling approach can be applied to this subgroup of offenders, and may help to identify different patterns of offending, or regression models which can address the estimation of current probability of escalation conditional on each offender's prior conviction. Therefore, to help capture the movement of high seriousness and high diversity in offending, one future research direction on the study of escalation would be to focus on the detection of such serious offenders, and to estimate more accurately their tendency for either escalation or de-escalation.

Finally, it is hoped this work will encourage more focus on studying escalation in general since, as should be clear, the present author believes that it is an important concept in policy terms. By being able to identify whether a person is on an escalating or a de-escalating trajectory, should help in deciding whether and what sort of intervention is appropriate.

Bibliography

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55(1):117–128.

Beck, V. S., Ramsey, R. J., Lipps, T. R., and Travis, L. F. (2006). Juvenile diversion: An outcome study of the Hamilton County, Ohio unofficial juvenile community courts. *Juvenile and Family Court Journal*, 57(2):1–10.

Becker, H. (1963). *Outsiders: Studies in the Sociology of Deviance.* Free Press, New York.

Bhati, A. S. and Piquero, A. R. (2008). Estimating the impact of incarceration on subsequent offending trajectories: Deterrent, criminogenic, or null effect? *Journal of Criminal Law and Criminology*, 98:207–254.

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, MA.

Blokland, A. A. J. and Nieuwbeerta, P. (2010). Life course criminology. In Shoham, S. G., Knepper, P., and Kett, M., editors, *International Handbook of Criminology*, pages 51–94. CRC Press-Taylor and Francis Group.

Blumstein, A., Cohen, J., Das, S., and Moitra, S. D. (1988). Specialization and seriousness during adult criminal careers. *Journal of Quantitative Criminology*, 4:303–345.

Blumstein, A., Cohen, J., Roth, J. A., and Visher, C. A., editors (1986a). *Criminal*

Careers and "Career Criminals", volume 1. National Academy Press, Washington, D.C.

Blumstein, A., Cohen, J., Roth, J. A., and Visher, C. A., editors (1986b). *Criminal Careers and "Career Criminals"*, volume 2. National Academy Press, Washington, D.C.

Blumstein, A., Farrington, D., and Moitra, S. (1985). Delinquency careers: Innocents, desisters, and persisters. In Tonry, M. and Morris, N., editors, *Crime and Justice: An Annual Review of Research*, volume 6, Chicago. University of Chicago Press.

Blumstein, A. and Graddy, E. (1982). Prevalence and recidivism in index arrests: A feedback model. *Law and Society Review*, 16:265–290.

Blumstein, A. and Nakamura, K. (2009). Redemption in the presence of widespread criminal background checks. *Criminology*, 47(2):327–359.

Brame, R. and Piquero, A. R. (2003). Selective attrition and the age crime relationship. *Journal of Quantitative Criminology*, 19(2):107–127.

Britt, C. L. (1996). The measurement of specialization and escalation in the criminal career: An alternative modeling strategy. *Journal of Quantitative Criminology*, 12:193–222.

Brocke, M., Göldenitz, C., Holling, H., and Bilsky, W. (2004). Attitudes towards severity of punishment: A conjoint analytic approach. *Psychology, Crime and Law*, 10(2):205–219.

Bruschke, J. and Loges, W. (2004). *Free Press vs. Fair Trials: Examining Publicity's Role in Trial Outcomes.* Routledge, New York.

Bushway, D. S., Sweeten, G., and Nieuwbeerta, P. (2009). Measuring long term individual trajectories of offending using multiple methods. *Journal of Quantitative Criminology*, 25:259–286.

Carrington, P. J., Matarazzo, A., and DeSouza, P. (2005). Court careers of a Canadian birth cohort. Technical Report 85-561-MIE-006, Canadian Centre for Justice Statistics.

Caussinus, H. (1966). Contribution à l'analyse statistique des tableaux de corrélation. *Annals of the Faculty of Science of the University of Toulouse*, 29:77–182.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman and Hall, London.

Datesman, S. K. and Aickin, M. (1984). Offense specialization and escalation among status offenders. *Journal of Criminal Law and Criminology*, 75(4):1246–1275.

DeLisi, M. and Piquero, R. A. (2011). New frontiers in criminal careers research, 2000-2011: A state-of-the-art review. *Journal of Criminal Justice*, 39:289–301.

Diggle, P. (1990). *Time Series: A Biostatistical Introduction.* Oxford University Press, Oxford.

Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002). *Analysis of Longitudinal Data.* Oxford University Press, Oxford, 2nd edition.

Eberly, E. L. and Thacheray, M. L. (2005). On Lange and Ryan's plotting technique for diagnosing non-normality of random effects. *Statistics and Probability Letters*, 75:77–85.

Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap.* Chapman and Hall/CRC, London.

Elliott, D. S. (1994). 1993 presidential address – serious violent offenders: Onset, developmental course, and termination. *Criminology*, 32:1–22.

Elliott, D. S., Ageton, S. S., Huizinga, D., Knowles, B., and Rachelle., C. (1983). *The Prevalence and Incidence of Delinquent Behavior: 1976-1980.* National Youth Survey, Report no. 26. Behavioral Research Institute, Boulder, Colorado.

Elliott, D. S., Huizinga, D., and Menard, S. (1989). *Multiple Problem Youth: Delinquency, Substance Use, and Mental Health Problems.* Springer, New York.

Ewens, W. and Grant, G. (2005). *Statistical Methods in Bioinformatics.* Springer, New York.

Fagan, A. A. and Western, J. (2005). Escalation and deceleration of offending behaviours from adolescence to early adulthood. *Australian and New Zealand Journal of Criminology*, 38(1):59–76.

Farrington, D. (1979). Longitudinal research on crime and delinquency. *Crime and Justice*, 1:289–348.

Farrington, D. (1986a). Age and crime. In Tonry, M. and Morris, N., editors, *Crime and Justice: An Annual Review of Research*, volume 7, Chicago. University of Chicago Press.

Farrington, D. (1986b). Stepping stones to adult criminal careers. In Olweus, D., Block, J., and Radke-Yarrow, M., editors, *Development of antisocial and prosocial behaviour: Research, theories and issues*, Orlando. Academic Press.

Farrington, D. P. (1989). The Origins of crime: The Cambridge Study in Delinquent Development. *Home Office Research Bulletin*, 27:29–32.

Farrington, D. P. (2002). Key results from the first forty years of the Cambridge Study in delinquent development. In Terence, P. T. and Krohn, D. M., editors, *Taking Stock of Delinquency: An Overview of Findings from Contemporary Longitudinal Studies.* Kluwer/Plenum, New York.

Fearn, T. (1977). A two-stage model for growth curves which leads to Rao's covariance-adjusted estimates. *Biometrika*, 64:141–143.

Figlio, R. M., Tracy, P. E., and Wolfgang, M. E. (1994). Delinquency in a Birth Cohort II: Philadelphia, 1958-1988. `http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/09293/detail`.

Firestone, P., Kingston, D. A., Wexler, A., and Bradford, J. M. (2006). Long-term follow-up of exhibitionists: Psychological, phallometric, and offense characteristics. *Journal of the American Academy of Psychiatry and the Law*, 34(3):349–359.

Florida Department of Corrections (2009). Florida criminal punishment code - scoresheet preparation manual. `http://www.dc.state.fl.us/pub/sen_cpcm/index.html`.

Francis, B. and Crosland, P. (2002). *The Police National Computer and the Offenders Index: Can they be combined for research purposes? Full report.* `http://library.npia.police.uk/docs/hordsolr/pncandoir170.pdf`. Home Office, London.

Francis, B., Crosland, P., and Harman, J. (2002). *The Police National Computer and the Offenders Index: Can they be combined for research purposes? Four page summary.* Home Office Research Findings 170. Home Office, London.

Francis, B. and Soothill, K. (2000). Does sex offending lead to homicide? *Journal of Forensic Psychiatry*, 11:49–61.

Francis, B., Soothill, K., and Dittrich, R. (2001). A new approach for ranking 'serious' offences: The use of paired-comparisons methodology. *British Journal of Criminology*, 41:726–737.

Francis, B., Soothill, K., Humphreys, L., and Bezzina, A. (2005). Developing measures of severity and frequency of reconviction. `http://eprints.lancs.ac.uk/50137/1/seriousnessreport.pdf`.

Francis, B., Soothill, K., and Piquero, A. R. (2007). Estimation issues and generational changes in modeling criminal career length. *Crime and Delinquency*, 53:84–105.

Glueck, S. and Glueck, E. T. (1930). *500 Criminal Careers.* Alfred A. Knopf, New York.

Glueck, S. and Glueck, E. T. (1950). *Unraveling Juvenile Delinquency.* Harvard University Press, Cambridge, Mass.

Gove, W. (1985). The effect of age and gender on deviant behavior: A biopsychosocial perspective. In Rossi, A. S., editor, *Gender and the Life Course.* Aldine, New York.

Hamparian, D. M., Schuster, R., Dinitz, S., and Conrad, J. (1978). *The Violent Few: A Study of Dangerous Juvenile Offenders.* Lexington Books, Lexington, Massachusetts.

Hanson, R. K., Scott, H., and Steffy, A. R. (1995). A comparison of child molesters and nonsexual criminals: Risk predictors and long-term recidivism. *Journal of Research in Crime and Delinquency*, 32:325–337.

Harrell, E. F. (2001). *Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression.* Springer-Verlag, New York.

Hastie, T. J. (1993). Generalized additive models. In Chambers, J. M. and Hastie, T. J., editors, *Statistical Models in S.* Chapman and Hall, London.

Hindelang, M. J., Hirschi, T., and Weis, J. G. (1981). *Measuring Delinquency.* Sage, Beverly Hills, California.

Holgersson, H. (2006). A graphical method for assessing multivariate normality. *Computational Statistics*, 21:141–149.

Home Office (1998). The Offenders Index: Codebook. `http://www.homeoffice.gov.uk/rds/pdfs/oicodes.pdf`.

Hwang, H. and Takane, Y. (2005). Estimation of growth curve models with structured error covariances by generalized estimating equations. *Behaviormetrika*, 32(2):155–163.

Jones, B., Nagin, D., and Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Social Methods and Research*, 29(3):374–393.

Kirk, D. S. (2006). Examining the divergence across self-report and official data sources on inferences about the adolescent life-course of crime. *Journal of Quantitative Criminology*, 22:107–129.

Kitsuse, J. I. and Cicourel, A. V. (1963). A note of the uses of official statistics. *Social Problems*, 11(2):131–139.

Kreuter, F. and Muthén, B. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology*, 24(1):1–31.

Kurlychek, M. C., Brame, R., and Bushway, S. D. (2007). Enduring risk? old criminal records and predictions of future criminal involvement. *Crime and Delinquency*, 53(1):64–83.

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

Laird, N. (1978). Non-parametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73:805–811.

Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, 38:963–974.

Lanctôt, N. and Le Blanc, M. (2002). Explaining deviance by adolescent females. In Tonry, M., editor, *Crime and Justice: An Annual Review of Research*, volume 29. University of Chicago Press, Chicago.

Langan, P. A. and Levin, D. J. (2002). *Recidivism of Prisoners Released in 1994*. Bureau of Justice Statistics. Department of Justice, Washington D.C.: U.S.

Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *Annals of Statistics*, 17(2):624–642.

Laub, J. H. and Sampson, R. J. (1994). Criminal careers and crime control in Massachusetts (The Glueck Study): A matched-sample longitudinal research design, phase I, 1939-1963. http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/9735/detail.

Lauritsen, J. (1998). The age-crime debate: Assessing the limits of longitudinal self-report data. *Social Forces*, 77:127–155.

Le Blanc, M. and Frèchette, M. (1989). *Male Criminal Activity from Childhood through Youth*. Springer-Verlag, New York.

Le Blanc, M. and Loeber, R. (1998). Developmental criminology updated. In Tonry, M., editor, *Crime and Justice: A Review of Research*, volume 23. The University of Chicago Press, Chicago.

Lemert, E. M. (1951). *Social Pathology: A Systematic Approach to the Theory of Sociopathic Behavior*. McGraw-Hill, New York.

Lindsey, J. K. and Jones, B. (1998). Choosing among generalized linear models applied to medical data. *Statistics in Medicine*, 17:59–68.

Liu, J., Francis, B., and Soothill, K. (2008). Kidnapping offenders: Their risk of escalation to repeat offending and other serious crime. *Journal of Forensic Psychiatry and Psychology*, 19:164–179.

Liu, J., Francis, B., and Soothill, K. (2011). A longitudinal study of escalation in crime seriousness. *Journal of Quantitative Criminology*, 27:175–196.

Lo, Y., Mendell, R. N., and Rubin, B. D. (2001). Testing the number of components in a normal mixture. *Biometrika Trust*, 388(3):767–778.

Loeber, R., Farrington, D., Stouthamer-Loeber, M., Moffit, T., and Caspi, A. (1998). The development of male offending: Key findings from the Pittsburgh Youth Study. *Studies in Crime and Crime Prevention*, 3:197–247.

Lukacs, E. (1942). A characterization of the normal distribution. *Annals of Mathematical Statistics*, 13(1):91–93.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings National Institute of Science, India*, 2(1):49–55.

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612.

McGee, T. R. and Farrington, P. D. (2010). Are there any true adult-onset offenders? *The British Journal of Criminology*, 50(3):530–549.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models.* Wiley-Interscience, New York.

Merton, R. K. (1957). *Social Theory and Social Structure.* Free Press, New York, US, revised edition.

Michigan State Police (1983a). *CCH - A closer look: Update.* Internal document. Michigan State Police, East Lansing: Michigan.

Michigan State Police (1983b). *Michigan's Computerized Criminal History System: A Closer Look.* Internal document. Michigan State Police, East Lansing: Michigan.

Miles, J. T. and Ludwig, J. (2007). The silence of the lambdas: Deterring incapacitation research. *Journal of Quantitative Criminology*, 23(4):287–301.

Moitra, S. D. (1981). *Analysis of Sentencing Policies Considering Crime Switching Patterns and Imprisonment Constraints.* PhD thesis, School of Urban and Public Affairs, Carnegie-Mellon University.

Muthén, B. and Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G., editors, *Longitudinal Data Analysis*, chapter 6, pages 143–162. Chapman and Hall, New York.

Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55:463–469.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4(2):139–157.

Nagin, D. S. (2005). *Group-Based Modeling of Development.* Harvard Univ. Press, Cambridge, MA.

Nagin, D. S., Cullen, F. T., and Jonson, C. L. (2009). Imprisonment and reoffending. In Tonry, M., editor, *Crime and Justice: A Review of Research*, volume 38, pages 115–200. University of Chicago Press, Chicago.

Nagin, D. S. and Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed poisson model. *Criminology*, 31:327–362.

Nagin, D. S. and Tremblay, R. E. (2005). Developmental trajectory groups: Fact or a useful statistical fiction? *Criminology*, 43(4):873–904.

Nylund, K. L., Asparouhov, T., and Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14:535–569.

Osgood, D. W. and Schreck, C. J. (2007). A new method for studying the extent, stability, and predictors of individual specialization in violence. *Criminology*, 45(2):273–312.

Ouston, J. (1984). Delinquency family background and educational attainment. *British Journal of Criminology*, 24:120–133.

Pinheiro, J. and Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York.

Piquero, A. R. (2000). Assessing the relationships between gender, chronicity, seriousness, and offense skewness in criminal offending. *Journal of Criminal Justice*, 28:103–116.

Piquero, A. R. (2004). Somewhere between persistence and desistence: The intermittency of criminal careers. In Maruna, S. and Immarigeon, R., editors, *After Crime and Punishment: Pathways to Offender Reintegration*, pages 102–128, Cullompton, England. Willan Publishing.

Piquero, A. R., Blumstein, A., Brame, R., Haapanen, R., Mulvey, E., and Nagin, D. (2001). Assessing the impact of exposure time and incapacitation on longitudinal trajectories on criminal offending. *Journal of Adolescent Research*, 16:54–74.

Piquero, A. R., Brame, R., Fagan, J., and Moffitt, E. T. (2006). Assessing the offending activity of criminal domestic violence suspects: Offense specialization, escalation, and de-escalation evidence from the spouse assault replication program. *Public Health Report*, 121(4):409–418.

Piquero, A. R. and Buka, S. L. (2002). Linking juvenile and adult patterns of criminal activity in the Providence Cohort of the National Collaborative Perinatal Project. *Journal of Criminal Justice*, 30:1–14.

Piquero, A. R., Farrington, D. P., and Blumstein, A. (2003). The criminal career paradigm. *Crime and Justice*, 30:359–506.

Piquero, A. R. and Weisburd, D., editors (2010). *The Handbook of Quantitative Criminology*. Springer, New York.

Plewis, I. (1996). Statistical methods for understanding cognitive growth: A review, a synthesis and an application. *British Journal of Mathematical and Statistical Psychology*, 49:25–42.

Prime, J. White, S., Liriano, S., and Patel, K. (2001). *Criminal Careers of Those Born between 1953 and 1978.* UK Home Office Statistical Bulletin 4/01. Home Office, London.

Prison Reform Trust (2008). 7 August 2008 – Two thirds of prisons officially overcrowded. `http://www.prisonreformtrust.org.uk/PressPolicy/News/vw/1/ItemID/38`.

Proust, C. and Jacqmin-Gadda, H. (2005). Estimation of linear mixed models with a mixture of distribution for the random-effects. *Computer Methods and Programs in Biomedicine*, 78:165–173.

Proust-Lima, C. and Liquet, B. (2011). The `lcmm` package version 1.4. `http://cran.r-project.org/web/packages/lcmm/lcmm.pdf`.

Ramchand, R., MacDonald, J. M., Haviland, A., and Morral, A. R. (2009). A developmental approach for measuring the severity of crimes. *Journal of Quantitative Criminology*, 25(2):129–153.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, 52:447–458.

Raudenbush, S. (2005). How do we study "what happens next"? *The Annals of the American Academy of Political and Social Science*, 602:131–144.

Reilly, B. and Witt, R. (1996). Crime, deterrence and unemployment in England and Wales: An empirical analysis. *Bulletin of Economic Research*, 48:137–159.

Rojek, D. G. and Erikson, M. L. (1982a). Delinquent careers: A test of the careers escalation model. *Criminology*, 20:5–28.

Rojek, D. G. and Erikson, M. L. (1982b). Reforming the juvenile justice system: The diversion of status offenders. *Law and Society Review*, 16(2):241–264.

Rosenberg, P. S., Katki, H., Swanson, C. A., Brown, L. M., Wacholder, S., and Hoover, R. N. (2003). Quantifying epidemiologic risk factors using non-parametric regression: Model selection remains the greatest challenge. *Statistics in Medicine*, 22:3369–3381.

Sample, L. L. and Bray, T. M. (2003). Are sex offenders dangerous? *Criminology and Public Policy*, 3(1):59–82.

Sampson, R. and Laub, J. (2005). Seductions of methods: Rejoinder to Nagin and Tremblay's "developmental trajectory groups: fact or fiction?". *Criminology*, 43:905–913.

Sapsford, R. J. (1998). Further research application of the "Parole Prediction Index". *International Journal of Criminology and Penology*, 6:247–254.

Sheldon, G. R., Horvath, A. J., and Tracy, S. (1987). Do status offenders get worse? some clarifications on the question of escalation. *Crime and Delinquency*, 35:202–216.

Sherman, L. W., Schmidt, J. D., Rogan, D. P., Gartin, P. R., Cohn, E. G., Collins, D. J., and Bacich, A. R. (1991). From initial deterrence to longterm escalation: Short-custody arrest for poverty ghetto domestic violence. *Criminology*, 29(4):821–850.

Sherman, L. W., Smith, D. A., Schmidt, J. D., and Rogan, D. P. (1992). Crime, punishment and stake in conformity: legal and informal control of domestic violence. *American Sociological Review*, 57:680–690.

Sipe, R., Jensen, L. E., and Everett, S. R. (1998). Adolescent sexual offenders grow up recidivism in young adulthood. *Criminal Justice and Behavior*, 25:109–124.

Skarðhamar, T. (2010). Distinguishing facts and artifacts in group-based modeling. *Criminology*, 48(1):295–320.

Sobel, M. E. (1988). Some models for the multiway contingency table with a one-to-one correspondence among categories. *Sociological Methodology*, 18:165–192.

Sobel, M. E., Hout, M., and Duncan, O. D. (1985). Exchange, structure, and symmetry in occupational mobility. *American Journal of Sociology*, 19(2):359–372.

Soothill, K., Fitzpatrick, C., and Francis, B. (2009). *Understanding Criminal Careers.* Willan Publishing, Cullompton.

Soothill, K. and Francis, B. (1997). Sexual reconvictions and the Sex Offenders Act 1997. *New Law Journal*, 147:1324–1325.

Soothill, K. and Francis, B. (2009). When do ex-offenders become like non-offenders? *The Howard Journal of Criminal Justice*, 48(4):373–387.

Soothill, K., Francis, B., Ackerley, E., and Fligelstone, R. (2002). *Murder and Serious Sexual Assault: What Criminal Histories can Reveal about Future Serious Offending.* Police Research Series 144. Home Office, London.

Soothill, K., Francis, B., and Liu, J. (2008). Does serious offending lead to homicide? Exploring the interrelationships and sequencing of serious crime. *British Journal of Criminology*, 48(4):522–537.

Spelman, W. (1994). *Criminal Incapacitation.* Plenum, New York.

Stasinopoulos, D. M. and Rigby, R. A. (1992). Detecting break points in generalised linear models. *Computational Statistics and Data Analysis*, 13(4):461–471.

Stattin, H., Magnusson, D., and Reichel, H. (1989). Criminal activity at different ages: A study based on a Swedish longitudinal research population. *British Journal of Criminology*, 29:368–385.

Stine, R. A. (1995). Graphical interpretation of variance inflation factors. *The American Statistician*, 49(1):53–56.

Stylianou, S. (2003). Measuring crime seriousness perceptions: What have we learned and what else do we want to know. *Journal of Criminal Justice*, 31:37–56.

Tappan, P. W. (1950). *The Habitual Sex Offender: Report and Recommendations of the Commission on Habitual Sex Offenders.* Commission on Habitual Sex Offenders, Trenton, NJ.

Taxman, F. S. and Piquero, A. (1998). On preventing drunk driving recidivism: an examination of rehabilitation and punishment approaches. *Journal of Criminal Justice*, 26(2):126–143.

Therneau, T. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model.* Springer-Verlag, New York.

Thistlethwaite, A., Wooldredge, J., and Gibbs, D. (1998). Severity of dispositions and domestic violence recidivism. *Crime and Delinquency*, 44(3):388–397.

Thomas, P. (1978). Plea bargaining in England . *Journal of Criminal Law and Criminology*, 69(2):170–178.

Thurstone, L. L. (1927). The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology*, 21(4):384–400.

Titterington, D., Smith, A., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Models.* Wiley, New York.

Ventura, L. A. and Davis, G. (2005). Domestic violence - court case conviction and recidivism. *Violence Against Women*, 11(2):255–277.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *American Statistical Association*, 91(433):217–221.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* Springer, New York.

Verbyla, A. (1986). Conditioning in the growth curve model. *Biometrika*, 73:475–483.

Verbyla, A. and Venables, W. (1988). An extension of the growth curve model. *Biometrika*, 75:129–138.

Vermunt, J. and Magidson, J. (2005). *Latent Gold 4.0 User Manual.* Statistical Innovations Inc., Belmont, Massachusetts.

Von Hirsch, A. (1986). *Past or Future Crimes: Deservedness and Dangerousness in the Sentencing of Criminals.* Manchester University Press, Manchester, UK.

Wagner, H. and Pease, K. (1978). On adding up scales of offence seriousness. *British Journal of Criminology*, 18:175–178.

Warburton, H., May, T., and Hough, M. (2005). Looking the other way: The impact of reclassifying cannabis on police warnings, arrests and informal action in England and Wales. *British Journal of Criminology*, 45(2):113–128.

Weiner, N. A. (1989). Violent criminal careers and 'violent career criminals'. In Wolfgang, E. M., editor, *Violent Crime, Violent Criminals.* Sage, Newbury Park, California.

Weis, J. G. (1986). Issues in the measurement of criminal careers. In Blumstein, A., Cohen, J., Roth, J. A., and Visher, C. A., editors, *Criminal Careers and "Career Criminals"*, volume 2, pages 1–51. National Academy Press, Washington, D.C.

Wolfgang, M. E., Figlio, R. M., and Sellin, T. (1972). *Delinquency in a Birth Cohort.* University of Chicago Press, Chicago.

Wolfgang, M. E., Figlio, R. M., and Sellin, T. (1994). Delinquency in a Birth Cohort in Philadelphia, Pennsylvania, 1945-1963. `http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/07729/detail`.

Wolfgang, M. E., Figlio, R. M., Tracy, P. E., and Singer, S. J. (1985). *The National Survey of Crime Severity (NCJ-96017)*. US Government Printing Office, Washington, DC.