Essays on the Economics of Crime

by

Katharine Inglis

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Department of Economics

Birmingham Business School

College of Social Sciences

November 2017

Abstract

This thesis examines topics on the Economics of crime, with a specific focus on the application of Econometrics in studying issues around crime, community safety and policy in England and Wales.

Chapters two and three highlight the gender gap in crime rates and sentencing outcomes and endeavours to identify possible causes. Utilising an ordered logistic regression model and a decomposition method, we find that differing risk preferences between men and women go some way to explaining the difference in offending rates. The analysis in chapter three uses a rich, individual-level dataset for sentencing in England and Wales and, controlling for confounding factors, we find that women are less likely than men to receive a custodial sentence when committing the same crime and receive a significantly shorter sentence when they do.

Chapters four and five analyse key risk factors for "Killed or Seriously Injured" (KSI) road traffic accidents in Norfolk and Suffolk. While chapter four employs an ordered logistic regression model to identify specific risk factors, such as not wearing a seatbelt and poor visibility, chapter five adopts a more novel approach by estimating a Classification and Regression Tree (CART) model to identify groups of significant characteristics.

Acknowledgements

Firstly, I would like to thank my lead supervisor, Dr. Siddhartha Bandyopadhyay, for his unwavering support throughout my research. His dedication and guidance through my undergraduate and postgraduate studies have been invaluable.

I also appreciate the thorough feedback and advice on Econometrics from my second supervisor, Professor Anindya Banerjee.

Thank you to Mum, Dad, Aimee and Graeme for their constant love and encouragement.

This work was supported by the Economic and Social Research Council [grant number 1365606].

I would like to thank my examiners Professor Eddie Kane and Professor Matthew Cole for their comments on the thesis and suggestions for future extensions.

I thank Norfolk and Suffolk Constabulary for facilitating the data in chapters four and five.

Contents

List of Figures

List of Tables

1 INTRODUCTION

In light of public funding cuts in recent years, there is an increasing demand for evidence-based policing and policy-making in England and Wales. The use of Econometric methods allows us to identify patterns and relationships in the data so that future policy can be based on sound evidence. The chapters in this thesis use a variety of econometric techniques to analyse two issues which face policy-makers today: the gender gap in crime and sentencing and serious road traffic accidents.

Chapter two tries to understand why women commit far fewer crimes than men. In particular, it looks at the role of differing risk preferences and provides evidence that female offenders are more risk averse than male offenders. A regression of offender characteristics on earnings risk produces a negative and significant estimated coefficient for the female dummy. This result indicates that female offenders choose crime types with lower risk in earnings and provides a way to determine which crime types women are most likely to choose. A Blinder-Oaxaca style decomposition indicates that differences in elasticities of offender behaviour with respect to changes in expected earnings and probability of apprehension account for the entire gender participation gap.

Chapter three tries to understand whether there is a gender gap in sentencing. Specifically, it investigates whether women receive lower sentences than men for the same crimes. Generalised ordered logistic regressions (GOLOGIT) are estimated for sentence type and length to find the difference in the probability of each outcome occurring for male and female offenders. The results from the empirical analysis show that the severity of sentencing outcomes for women are lower than for men even when aggravating and mitigating factors are controlled for. Given that an offender receives a custodial sentence, women receive significantly shorter sentences than men in most offence categories.

The purpose of the research in chapter four is to determine the factors and characteristics which affect the severity of road traffic accidents and to identify driver groups who are most at risk. Several variables are found to significantly effect the severity of an accident, including gender, wearing a seatbelt and visibility. Drivers found to be most at risk of being involved in an accident are those aged 17-39 and female drivers over the age of 70.

Chapter five classifies drivers according to the severity of their accident using Classification and Regression Tree Analysis (CART). Developing this type of predictive model allows policy makers to identify groups who are most at risk of being involved in KSI accidents and allows them to target severity-reducing measures accordingly.

2 RISK PREFERENCES OF FEMALE OFFENDERS AND THE GENDER PARTICIPATION GAP IN CRIME

2.1 Introduction

In 2015, women made up 25% of first time offenders and only 14% of repeat offenders [Ministry of Justice, 2016]. These percentages are even lower for serious offences and illustrate the magnitude of the gender gap in crime. Various explanations for this gap have been offered in the literature, some suggest that criminality is a male characteristic [Lombrosso and Ferraro, 1895, Akerlof and Kranton, 2000] and others discuss barriers to crime such as children and income [Freeman, 1999, Hart, 1985]. In this chapter we will explore the role of differing risk preferences as there is evidence that women are more risk averse than men [Fehr-Duda et al., 2006, Schubert et al., 1999].

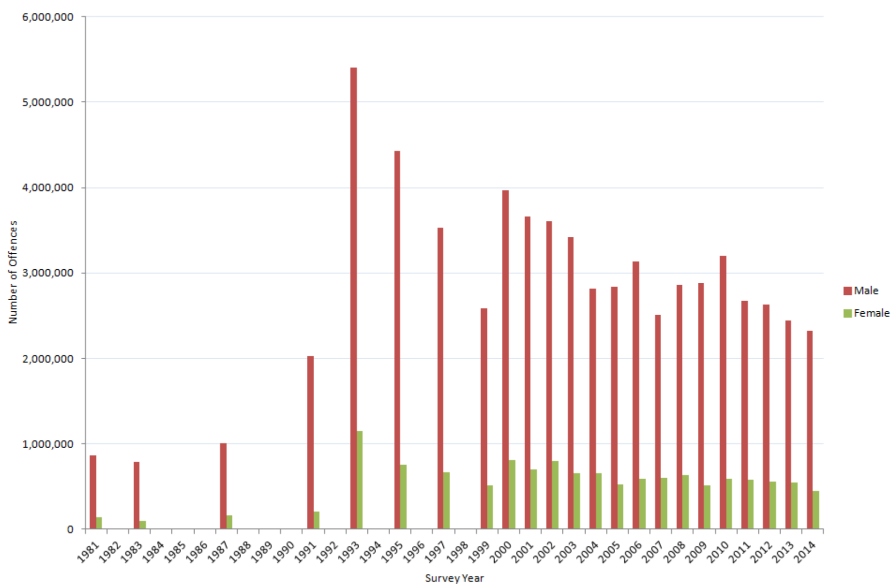
The analysis in this chapter finds that female offenders choose property crime types in which earnings risk is lower than in offence types chosen by men. This result complements the findings by Gavrilova and Campaniello [2015], who also find a negative and significant relationship between being female and the earnings risk of a crime. This result is found by running an OLS regression of earnings risk on gender, controlling for offender and environment characteristics, and estimating a statistically significant coefficient of -0.21. This indicates that women choose crime types in which the earnings risk is 21% lower than crime types chosen by men. Following this, a Blinder-Oaxaca decomposition technique [Blinder, 1973, Oaxaca, 1973, Jann, 2008] is used to show that gender differences in risk preference account for the entire gender gap in crime.

There are very few papers in the Economics literature which study the reasons for the vast gender gap in crime that we observe, fewer still explore the possibility that male and female offenders differ in behaviour due to their preferences. The analysis in this chapter is the first to study this relationship for a nationally representative dataset from England and Wales. For this reason, the outcomes of this chapter are interesting and add to the current literature, exploring the role of risk preferences within an inherently risky activity such as crime.

This chapter also contributes to the crime literature by adopting a novel measure of risk and applying it to an individual level dataset for England and Wales from 1981 to 2015. One reason for the relatively few studies on female crime in the Economics literature is that researchers often find it convenient to drop the small proportion of female offenders from the dataset before carrying out their analysis. The small numbers make it difficult to perform hypothesis tests and make robust conclusions about statistical significance. This is overcome by using a representative, repeated cross-sectional dataset which provides a large number of female observations to analyse.

Figure 1 shows the number of offences for single male and female offenders by survey year. Although the Crime Survey for England and Wales (CSEW) began in 1981, it was not carried out annually until 1999, so there are some missing years before this. The CSEW is weighted in such a way that it represents the whole population and provides a large sample of male and female offenders for analysis.

Figure 1: Number of Offences by Gender for Single Offenders



Using this representative dataset, empirical analysis is carried out to investigate male and female preferences for earnings risk and how responsiveness to incentives contributes to the gender gap. The earnings risk is calculated for each crime type by estimating a Mincer style wage equation, following the method outlined by Bonin et al.

[2007] and taking the standard deviation of the estimated error term for each crime category. The results from the analysis show a strong relationship between female offenders and earnings risk, implying that women choose crimes in which their pay-off is more certain. This observed behaviour reinforces the general idea that women are more risk averse than men, very few choose to commit crime and those who do tend to choose crime types with the lowest earnings risk. Following this, the participation gap is decomposed to find how much of the difference can be explained by the elasticity of offender behaviour with respect to two of the most important incentives: the expected pay-off from crime and the probability that they will not be apprehended. A synthetic panel dataset is constructed and used in a Blinder-Oaxaca decomposition and the results strikingly show that, if these elasticities were the same for women as they are for men, the gender gap would be reversed, with women committing more crimes than men.

2.2 Literature Review

Gender-specific risk preferences have been explored by several authors in a number of disciplines, but conclusive results are relatively sparse. These studies generally analyse self-reported preferences or calculate revealed preferences through observed behaviours and find that the behaviour of men and women in uncertain situations depends on the context.

Schubert et al. [1999] conduct an experiment where male and female subjects are offered lotteries which are framed either as insurance, investments or gambling opportunities. Their results indicate that risk preferences only differ between genders when the lottery is framed as an abstract gamble, with men seeking more risk in monetary gains and women seeking more risk in monetary losses. When framed as insurance or an investment, they find no significant difference between the certainty equivalents of each gender, implying that probability weighting does depend on the framing of a question. In a similar experiment, Fehr-Duda et al. [2006] use monetary incentives to test whether female subjects are more risk averse than males. They hypothesise that there is a gender difference in probability weighting and their results show that women do underestimate gains with a larger probability relative to men.

When analysing self-reported data, the validity of a characteristic such as risk aversion can be difficult to accept at face value. To deal with this problem, Dohmen et al. [2005] use survey data in which 22,000 individuals are asked to rate their willingness to take risks by choosing a number from 1 to 11. They test these reported preferences in an experimental setting by asking participants to choose one of two lotteries, one more risky than the other. These revealed preference results indicate that the self-reported risk preferences in the dataset are accurate and, on the whole, women are more risk averse than men.

The analysis in this chapter takes inspiration from a paper by Gavrilova and Campaniello [2015] in which they use a measure of earnings risk developed by Bonin et al. [2007] to test whether female offenders choose offence types with lower earnings risk. The premise of Bonin et al.'s paper is that jobs in the legitimate sector vary in several different types of risk, such as health risks, but one which has not been studied is risk in earnings when wages vary in different occupation groups. The aim of their paper is to find whether people sort into occupation types according to their risk preferences. They are able to use a unique dataset in order to analyse this relationship. The German Socio-Economic Panel (SOEP) dataset contains information about self-reported risk preferences for every individual in the survey, which measures their personal willingness to take risks by asking them to choose a number within a linear range. The authors recognise that this self-reported level of risk preference is very subjective so they conducted a separate field experiment [Dohmen et al., 2005] to show that these risk preferences and actual risky behaviour are very closely correlated. They found that those who describe themselves as risk-loving are more likely to take part in activities such as playing the lottery or smoking.

In order to categorise occupation types according to their earnings risk, they develop a measure of risk using the variation in monthly wages which cannot be explained by independent variables in a standard Mincer wage equation [Mincer, 1958, 1974]. They hypothesise that if variables related to human capital cannot explain the variation in wages, this variation is a risk associated with the job type.

In order to measure the earnings risk for a particular occupation, the authors estimate a Mincer wage equation using the log of the monthly earnings as the dependent variable and several characteristics of the job and human capital of the individual as independent variables. They take the estimated residuals for each wage equation and assign the standard deviation of these residuals to each occupation type, calling this the measure of earnings risk. The standard deviation of the residual will depend on which independent variables are included in the Mincer equation, some of the variance in the error may just be due to omitted relevant variables. The validity of the earnings risk measure relies on the assumption that all relevant, human capital variables have been included in the equation and any unexplained variation is considered to be earnings risk. The estimated Mincer [1958, 1974] wage regression is given by Equation

1.

$$
\begin{aligned}
logEarnings = \alpha &+ \beta_1 Experience_i + \beta_2 Experience_i^2 \\
&+ \beta_3 Experience_i^3 + \beta_4 Tenure_i + \beta_5 Tenure_i^2 \\
&+ \beta_6 Education_i + \beta_7 EastGermany_i \\
&+ \beta_8 PublicSector_i + \beta_9 RiskAttitude_i + \epsilon_i
\end{aligned} \tag{1}
$$

The authors seek to find whether those individuals who claim to be willing to take risks sort themselves into occupation categories with a high measure of earnings risk, according to the standard deviation of the residuals from the Mincer equation. Throughout the analysis, it is assumed that individuals take the earnings risk of occupation types as given and are fully aware of it. The authors initially carry out the analysis for male, full time workers aged 25-55 and calculate the earnings risk for each occupation type, which ranges from 0.2 to 0.8. By regressing earnings risk on risk attitude, amongst other control variables, they show that there is a positive and significant relationship, indicating that those individuals who are more willing to take risk do indeed sort themselves into occupations with higher earnings risk.

The analysis is then repeated for female workers and the results are very similar to the male results. The slightly weaker results are shown by a lower R^2 statistic for the Mincer wage equation is lower, which implies that the variation in wages for females is explained less by independent human capital variables than it is for men. They conclude that these results offer a potential explanation for the gender wage gap observed in most countries. Previous literature has shown that women are more risk averse than men [Dohmen et al., 2005] and that more risky occupations tend to have higher wages [McGoldrick, 1995], therefore we would expect to see a gender wage gap given these results.

It is of course possible that women are sorted into occupation types with lower wages through discrimination by the labour market rather than by choice. This would lead to lower earnings risk but this is not discussed by Dohmen et al. [2005]. These findings about risk preferences and earnings risk enable Gavrilova and Campaniello [2015] to analyse a similar relationship in the crime market.

Gavrilova and Campaniello [2015] outline a theoretical model which adapts Becker's theory [Becker, 1968] that potential offenders compare costs and benefits when deciding whether to commit a crime. They show that individuals have a threshold legitimate wage, $W*$, below which they will choose to commit crime over legitimate work.

$$(1-p)\,E\left[U\left(Earnings\right)\right] + pU\left(Punishment\right) > U\left(Wage\right) \quad (2)$$

$$\begin{aligned} W* = {} & U^{-1}\left\{(1-p)\,E\left[U\left(Earnings\right)\right] + pU\left(Punishment\right)\right\} \\ & > Wage \end{aligned} \quad (3)$$

Equation 3 indicates that an individual's threshold wage is determined by probability of apprehension, expected earnings from crime and expected punishment. The authors regress criminal earnings on gender and other characteristics to find that women earn 30% less than men in crime. However, when they control for the type of offence committed, this gender gap in criminal earnings disappears, indicating that the earnings gap is due to men and women choosing different types of crime. This sorting behaviour, which leads to differences in expected criminal earnings for men and women, would lead to a difference in threshold wages in Equation 3 and could, at least partly, explain why fewer women than men choose crime over legitimate work.

Gavrilova and Campaniello [2015] use the National Incident Based Reporting System dataset, a US dataset which contains details of criminal incidents for a sample of reporting agencies. Included in their analysis are 7, 812,439 observations for offenders aged 15 to 65 who commit property crimes during the years 1995 to 2010. To test whether female offenders sort into crime types with lower variance in monetary earnings, the authors calculate earnings risk for each crime type by estimating a Mincer style wage equation, shown by Equation 4. This parallels the method outlined by Bonin et al. [2007] and is followed by taking the standard deviation of the estimated error term for each crime category.

$$\begin{aligned} logEarnings = \alpha + \beta_1 Female_i + \beta_2 Age_i + \beta_3 Weapon_i \\ + \beta_4 Female_i * Weapon_i + \tau_{year*agency} + \epsilon_i \end{aligned} \tag{4}$$

After running this wage equation for each offence type and assign-ing the correct earnings risk to each individual crime, the authors run the second regression using the earnings risk as the dependent variable, to find whether the results are similar to those found by Bonin et al. [2007] in the legitimate job market. The results show an estimated coefficient for female offenders which is negative and statistically significant, implying that female offenders sort into crime types with lower earnings risk.

After showing that female offenders are more risk averse than males, the authors propose a theoretical model which shows how different risk preferences can lead to a gender participation gap in crime. Equation 5 gives the offender's expected utility from a crime in terms of the probability of apprehension (p), earnings from crime (E), coefficient of risk aversion (r) and dis-utility from time in jail (D). By summing over individuals, the authors are able to define the total number of crimes, C, in terms of these variables in Equation 6. Equation 7 takes logs on both sides and assigns the value a to the size of the potential population of offenders. This crime equation is used to specify an OLS equation with the log of the total number of crimes as the dependent variable in Equation 8, where g represents gender, j is the location, t is the time period, R is a dummy for race, A is age category and Y is a year dummy.

$$\begin{aligned} E\left(U\right) &= \left(1-p\right)\left(\frac{E^{1-r}}{1-r}\right) - pD \\ &> \varepsilon \end{aligned} \tag{5}$$

$$C \propto (1-p)\left(\frac{E^{1-r}}{1-r}\right) \tag{6}$$

$$log\,(C) = a + (1-r)\,logE + log\,(1-p) \tag{7}$$

$$\begin{aligned} logC_{gjt} = \alpha_j + \beta_{1g} logE_{jt-1} + \beta_{2g} log\,(1-p_{jt-1}) \\ + \beta_{3g} R_{jt} + \beta_{4g} Y_{jt} + \beta_{5g} A_{jt} + \varepsilon_{gjt} \end{aligned} \tag{8}$$

The results of this estimation show that male offenders react positively to both the expected earnings and the expected probability of getting away with the crime. The elasticities for female offenders are much lower and are only positive for expected earnings, the estimated elasticity for probability of not being apprehended is insignificant. This result shows that women are less responsive to changes in incentives than men and suggests that, since crime is a risky activity, female risk aversion reduces the probability that women will choose crime when incentives increase.

Finally, to measure the effect of these differences in incentive elasticity, they used a Blinder-Oaxaca decomposition to construct a counterfactual equation for female offenders using the estimated incentive coefficients for male offenders. The results from this estimated equation show that women would increase their number of offences by 40% of the participation gap if they responded to incentives in the same way as male offenders. Differences in risk preferences cannot explain the other 60% of the gender gap in crimes. A limitation of this paper is that the data is not representative of the US since only certain reporting agencies collect the data.

The gender gap in crime is large and persistent, with only 16% of crimes committed in England and Wales by female offenders, according to the CSEW. The bulk of research on this topic has been carried out in disciplines such as Psychology, Criminology and Sociology and several reasons for this difference have been cited from the "Biological Essentialism" argument, the idea that women are genetically programmed to care and nurture and are not predisposed to crime or violence, to psychological explanations. In 1915, Lombrosso and Ferraro suggested that women who commit crime lack female characteristics which leads them to behave like men. To support this claim, they produce several measurements of physical and mental characteristics of female criminals and show these to be similar to average male measurements at the time.

Due to the small number of female offenders, researchers often find it convenient to drop them from datasets used in their analysis and focus on male offenders. For this reason, there are relatively few papers on women in crime in the Economics literature. Economists have not yet studied the role of incentives in female crime, there are many different incentives at play such as legitimate wage differences, working hours and child-rearing [Freeman, 1999].

Several other theories have been published as to why women are less likely to commit crime, with many citing a woman's role in society as a barrier to criminal behaviour. Akerlof and Kranton [2000] discuss the idea that individuals have a particular identity within society and they lose utility when they behave in a way that does not fit in with the norm. In this model, there are different categories within society and each category has a certain type of ideal behaviour attached to it. Individuals identify with a particular category and their utility is dependent on their self image, their behaviour and the behaviour of others. People have a choice about their identity to a certain extent, but when identities are limited or prescribed in some way, their behaviour and utility functions are predetermined. This theory could be extended to explain offender behaviour if gender is assumed to be a prescribed category and crime is seen as a typically male activity. Women would therefore lose utility if they commit crime because they are demonstrating behaviour that is not compatible with their social identity. However, Hart [1985] argues that traditional theories about female crime are no longer relevant since the role of women in Britain has changed so dramatically. Arguments such as that women have fewer opportunities to commit crime and are only concerned with marriage and family no longer represent modern women, who have the same opportunities as men [Hart, 1985]. Although, despite the changing roles of women in society, there is evidence that women still face many barriers to entry when it comes to crime. Steffensmeier and

Terry [1986] interview several male offenders about their attitudes towards women and find that they, almost universally, do not think that women have the necessary skills to be successful criminals and therefore choose to commit crime with men.

2.3 Theory

Crime is an activity that was originally thought of as irrational and therefore not suitable for Economic analysis. Early theories discuss psychological reasons for committing crime and offenders were viewed as inherently bad. Becker (1968) was the first to suggest that committing crime was, in fact, an economic choice like any other. There are costs and benefits, each with particular probabilities attached, and individuals are likely to choose crime or legitimate work depending on which will maximise their expected utility. With this idea in mind, he developed a model outlining the decision process followed by potential offenders and claimed that they will choose to commit a crime if the expected utility from doing so outweighs the expected utility from legitimate work. He outlined the expected utility from committing crime which is illustrated by Equation 9, where j denotes the offence, p_j is the offender's probability of conviction for committing offence j, Y_j is the expected benefit from committing offence j, f_j is the expected punishment for committing offence j and W is the expected wage from a legitimate job.

$$\begin{aligned} E(U_j) &= p_j U_j(Y_j - f_j) + (1 - p_j) U_j(Y_j) \\ &> W \end{aligned} \tag{9}$$

This model has been adopted by scores of subsequent work and forms the basis of most economic papers in the crime literature. We can use this framework to consider the differences between male and female offenders. Let all potential offenders belong to one of two groups, males are denoted "M" and females are denoted "F". The expected utility function above can then be written separately for males and females. For a man to commit a crime, it must be true that

$$\begin{aligned} E(U_j^M) &= p_j^M U_j^M (Y_j^M - f_j^M) + (1 - p_j^M) U_j^M (Y_j^M) \\ &> W^M \end{aligned} \tag{10}$$

For a woman to commit a crime, it must be true that

$$\begin{aligned} E(U_j^F) &= p_j^F U_j^F (Y_j^F - f_j^F) + (1 - p_j^F) U_j^F (Y_j^F) \\ &> W^F \end{aligned} \tag{11}$$

It can be seen in Equations 10 and 11 that each of the variables has its own superscript, which implies that probability of apprehension, expected benefit, expected punishment and expected legitimate wage can differ for men and women. This is explored in the literature and there are several studies which claim that female offenders are less likely to be apprehended and are likely to face a lower punishment if they are [Gavrilova and Campaniello, 2015, Visher, 1983]. It is also shown in chapter three that sentencing for women is more lenient than for men, even when in the same offence category. These observed differences would, however, make crime more inviting to women and cannot be the explanation for the large gender gap in crime. The important difference to consider is, of course, the utility functions or more specifically the differences in risk preferences. For fewer women to choose crime than men, it could be true that their utility functions take different forms.

When deciding between crime and legitimate work, individuals from both groups face a choice between two lotteries. If they choose crime, the pay-offs are $Y_j - f_j$ with probability p_j and Y_j with probability $1 - p_j$. Whereas, if they choose legitimate work, the pay-off is W with certainty. These two lotteries are defined in Equations 12 and 13 and the expected values are calculated in Equations 14 and 15 respectively.

$$L_1 = (Y_j - f_j, Y_j; p_j, 1 - p_j) \tag{12}$$

$$L_2 = (W; 1) \tag{13}$$

$$E(L_1) = p_j(Y_j - f_j) + (1 - p_j)(Y_j) \tag{14}$$

$$E(L_2) = W \tag{15}$$

To begin with, let us assume that crime pays for women, for reasons discussed above. It is therefore true that the expected value of L_1 is higher than the expected value of L_2.

$$E^F(L_1) > E^F(L_2) \tag{16}$$

The expected pay-off from crime is higher than the expected pay-off from legitimate work for women. When all women are risk neutral, it would be reasonable to assume that they would all choose to commit crime over legitimate work since their expected utility is proportional to the expected value of a lottery. If their utility function were, for example, $u(x) = x$, then $E(u(L_1)) > E(u(L_2))$.

However, if we assume that women are risk averse, their utility function is now concave. The expected utility is no longer proportional to the expected value and the utility function needs to be taken into account when choosing between L_1 and L_2.

The aim here is to find the conditions under which women will choose legitimate work over crime, even when the expected pay off from crime is higher. When the expected value of crime is higher than the expected value from legitimate work, but the expected utility from crime is lower than the expected utility from legitimate work, the following conditions must hold.

$$W < p_j(Y_j - f_j) + (1 - p_j)Y_j \tag{17}$$

$$EU\left[W\right] > EU\left[p_j(Y_j - f_j) + (1 - p_j)Y_j\right] \tag{18}$$

The concavity of the utility function means that there is a certainty equivalent below the expected value of the lottery. At this point, the individual is indifferent between the gamble and the certain amount. To find the certainty equivalent (C.E):

$$U(C.E) = p_j\,[U(Y_j - f_j)] + (1 - p_j)\,[U(Y_j)]$$

Assuming a log form, we get:

$$\ln(C.E) = p_j \ln(Y_j - f_j) + (1 - p_j)\ln(Y_j)$$

$$e^{\ln(C.E)} = e^{p_j \ln(Y_j - f_j) + (1 - p_j)\ln(Y_j)}$$

$$C.E = e^{p_j \ln(Y_j - f_j)} . e^{(1 - p_j)\ln(Y_j)} \tag{19}$$

$$C.E = (e^{\ln(Y_J - f_j)})^{p_j} . (e^{\ln(Y_J)})^{1 - p_j}$$

$$C.E = (Y_j - f_j)^{p_j} . (Y_j)^{1 - p_j}$$

For argument's sake, if the individual is indifferent between the certainty equivalent and the gamble, let them choose legitimate work in order to avoid the social stigma attached to committing crime, or something other reason. Considering the condition given in Equation 17, there must be a range of values for W where the expected value of crime is higher but the expected utility is lower for women.

$$\begin{aligned}(Y_j - f_j)^{p_j} . (Y_j)^{1 - p_j} &\leq W \\ &< p_j(Y_j - f_j) + (1 - p_j)Y_j\end{aligned} \tag{20}$$

When Equation 20 is satisfied, risk averse individuals will choose to work in the legitimate sector over crime, despite the expected value from crime being higher. If it is the case that women are generally more risk averse than men, this theory could go some way to explaining the gender gap in crime.

2.4 Data

The dataset used in this chapter is the Crime Survey for England and Wales (CSEW), formerly known as the British Crime Survey. The CSEW was first carried out in 1981 and aims to survey around 50,000 households, biennially until 2000 and then annually. There are several benefits of using this dataset over Police Reported Crime (PRC), mainly the fact that it includes those crimes which have not been reported to the police. The survey data also give a better idea of long term trends in crime because they are not subject to trends in reporting.

The CSEW is designed in such a way that the core set of questions have been kept constant since the beginning, so it is possible to compare answers to identical questions over time. The data are also weighted in such a way that it can be used as representative of the whole population of England and Wales.

There are drawbacks to the dataset, namely that it excludes what can be thought of as “victimless” crimes such as fraud and drug possession. It also excludes homicide. These limitations are not too important for the analysis in this chapter as the main focus is on property crimes. This restriction is necessary for this analysis since the calculation of an offence’s earnings risk relies on observing the offender’s earnings from the crime, namely the value of the stolen property. The value of stolen property is recorded in the CSEW dataset as the answer to the question, “what was the total replacement value of what was stolen?” and this value is treated as the offender’s earnings in this analysis.

Variable descriptions for this dataset are given in Appendix A.1 and it can be seen that there are sub-categories within each offence category. The main offence categories are robbery, snatch theft, theft from the person, domestic burglary, theft from a dwelling, domestic burglary in an outhouse, vehicle-related thefts, theft from outside dwelling, other personal theft and criminal damage. Table 15 shows the questions that the victim is asked during the survey, the answers to which are used as offence characteristics in the analysis that follows.

2.5 Estimation Methodology

Female Offender Risk Preferences

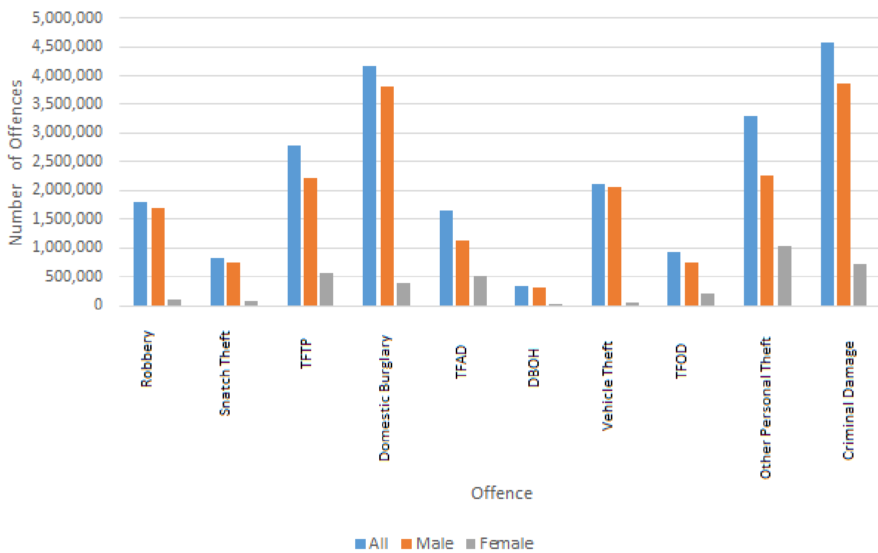
One of the principle concepts in economics is the idea that individuals are rational and behave according to their own personal preferences. The above theory section uses this concept to show that, given the other known constraints for men and women such as legitimate wages, probability of apprehension and so on, it may be the case that men and women have different preferences for there to be such a vast gap in the crime rates. More specifically, it may be true that women are more risk averse and, despite the possible high profits from crime, would almost always prefer the certainty of a legitimate wage.

The aim of this analysis is to explore the risk attitudes of female offenders compared with male offenders using individual-level data from England and Wales for the years 1981 to 2015. Following on from the work by Gavrilova and Campaniello [2015], it uses Bonin et al.'s measure of earnings risk to find whether offenders sort into different crime types according to the their risk preferences.

In order to use the total earnings of crime, it is necessary to restrict the dataset to property crimes, where the total value of stolen goods is recorded for each offence. This earnings value is used to run Mincer wage equations [Mincer, 1958, 1974] for each crime type, and the standard deviation of the residuals is recorded as the earnings risk. Each offence is then assigned the appropriate earnings risk value and an OLS regression is run using the earnings risk as the dependent variable. The idea behind Gavrilova and Campaniello [2015]'s paper is that a significant and negative estimated coefficient for the female dummy variable in this regression indicates that female offenders sort into crime types with lower levels of earnings risk. Gavrilova and Campaniello [2015] do find a negative and significant estimated coefficient in their analysis and interpret this as evidence that there is sorting. This leads to their conclusion that female offenders are more risk averse than their male peers.

It has been well documented that female offenders choose different crime types to male offenders. Figure 2 shows that, even within property crimes, there is low participation of female offenders in certain offence categories. The three most popular for women are theft from the person, personal theft and criminal damage (including pick-pocketing). The offence categories in Figure 2 are robbery, snatch theft, theft from the person (TFTP), domestic burglary, theft from a dwelling (TFAD), domestic burglary in an outhouse (DBOH), vehicle-related thefts, theft from outside a dwelling (TFOD), other personal theft and criminal damage.

Figure 2: Crimes by Type and Gender



To put the earnings risk measure into context, Bonin's measure of earnings risk can be used with the Labour Force Survey (LFS) dataset to calculate the earnings risk associated with different legitimate industries in England and Wales, the results are shown in Table 1. Each individual in the dataset belongs to one of nine industry categories and the earnings risk is calculated for each. There is very little variation in the measures which range from 0.84 to 0.89. These are very close to those found by Bonin et al. Reasons for the small differences could be that it's a different country and different time period.

Table 1: Earnings Risk by Occupation Type

Industry	Earnings Risk
Agriculture, Forestry and Fishing	0.861
Energy and Water	0.886
Manufacturing	0.844
Construction	0.850
Distribution, Hotels and Restaurants	0.846
Transport and Communication	0.850
Banking and Finance	0.847
Public admin, Education and Health	0.842
Other Services	0.850

When compared with work in the legitimate sector, crime is clearly a much riskier choice, as shown in Table 2. In addition to the potential problems faced in the legitimate sector, those choosing to commit crime also face the risk of apprehension and punishment, not to mention the fact that the illegal job market is not regulated like the legitimate market. These risks, however, are likely to be specific to individual crimes and will depend on a multitude of factors. In order to compare risks across offences, it must be something that is common across individuals and can be quantified in the same way. Gavrilova and Campaniello [2015] restrict their dataset to property crimes in which the total value of stolen goods can be measured and compared across the board, this can be regarded as the "earnings" from a particular offence.

The US dataset used by Gavrilova and Campaniello [2015] covers only a few crime reporting agencies and is therefore not representative of the country as a whole. This is not a problem for the CSEW dataset since the frequency weights allow it to be used as representative for the whole population.

Only crimes with a single offender are considered in this analysis since individual characteristics such as age and ethnicity are not given when there are multiple offenders. This means that 48% of the observations are excluded because either there are multiple offenders or the victim did not know how many there were. Although this is a reasonably large proportion to exclude, the characteristics of individual offenders are vital in order to calculate the wage equations shown in Equation 21. There are also many more independent variables that can be included in the Mincer wage equation than in Gavrilova and Campaniello [2015], for each separate offence type the following model is estimated where i denotes the individual offence.:

$$\begin{aligned}logEarnings_i = {} & Female_i + Age_i + Drug_i \\ & + Drink_i + Relationship_i + Race_i \\ & + Weapon_i + Contact_i + Force_i + Threaten_i \\ & + Sexual_i + Knew_i + StreetGang_i + \\ & Time_i + Weekend_i + Year_i + \varepsilon_i \end{aligned} \tag{21}$$

The standard deviation of the residuals is then recorded and each individual offence is assigned a measure of earnings risk according to the offence category that it comes under, which are given in Table 2.

Table 2: Offence Earnings Risks

Offence	Obs	Residual Mean	Residual S.D =Earnings risk
Robbery	5,576,170	2.224	2.624
Snatch Theft	1,567,959	-1.636	6.030
Theft from the Person	10,537,378	-1.522	3.158
Domestic Burglary	21,621,478	-0.698	2.391
Theft from a Dwelling	2,613,583	-0.613	1.858
Domestic Burglary Outhouse	6,302,198	0.089	2.766
Vehicle Theft	54,786,220	-0.843	1.777
Theft from Outside Dwelling	19,717,500	0.178	2.081
Other Personal Theft	24,991,293	-0.005	2.199
Criminal Damage	55,542,421	-5.173	6.327

The mean earnings and earnings risk for each offence category are illustrated in Figures 3 and 4. In the context of this analysis, earnings are given by the total value of stolen goods. It may be surprising that criminal damage and arson have such high mean earnings given that they don't necessarily involve stealing any property. However, the recorded offence is the main offence committed and not necessarily the only one, therefore offenders who commit criminal damage and arson may have stolen some property during the offence.

The mean earnings in Figure 3 vary across different types of offence and between men and women. Aside from arson and criminal damage, burglary in a dwelling has the highest average earnings among offence types, while theft from outside a dwelling has the lowest. Figure 4 shows that the crime type with the highest level of earnings risk is arson and criminal damage while the lowest is theft of/from a vehicle. As explained earlier, the earnings risk is the variability in earnings which cannot be explained by an offender's characteristics included in Equation 21. The level of earnings and earnings risk are not correlated with each other, as shown in Section 2.7, so the effect of earnings risk on female offender decision making is unlikely to be driven by the level of earnings.

Figure 3: Mean Earnings

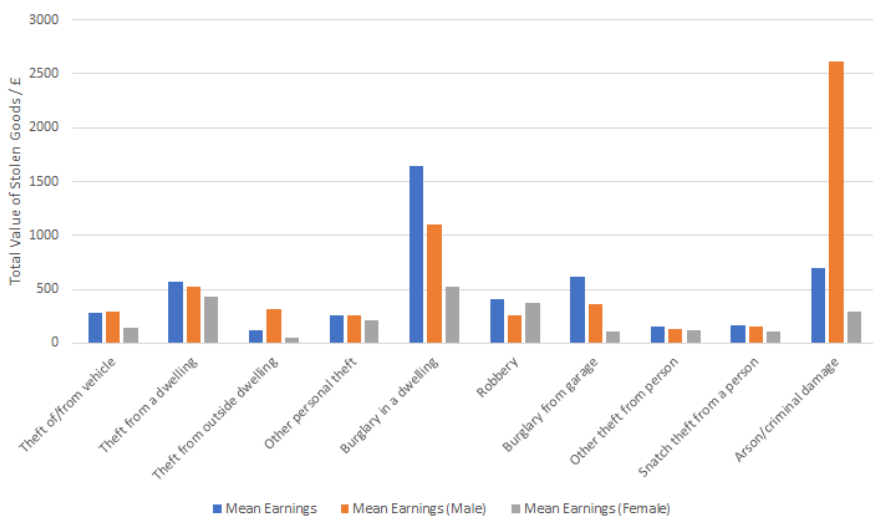
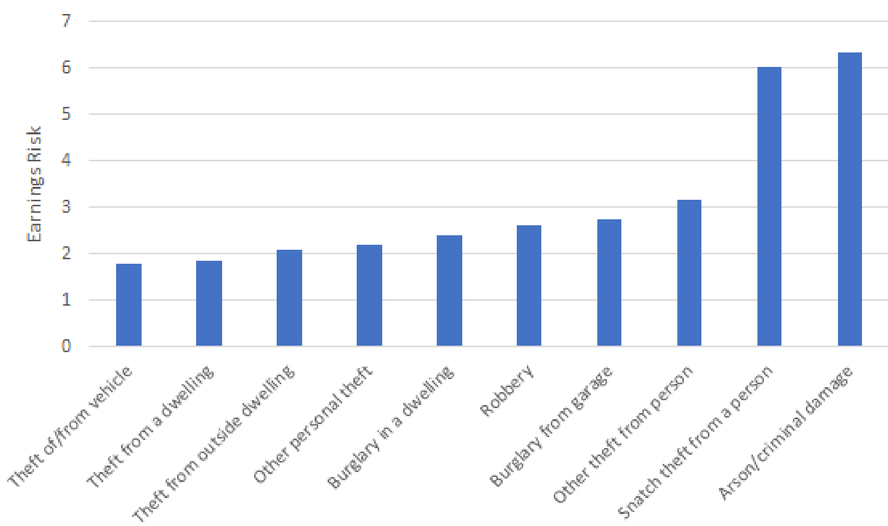


Figure 4: Earnings Risk



2.6 Why is there a Gender Gap?

To analyse the participation gap in crime, it is necessary to model the to number of crimes in terms of incentives faced by potential offenders. Equation 22 is the derived crime equation from Gavrilova and Campaniello [2015], adapted for this analysis, which gives the total number of crimes in terms of expected log earnings and the log probability of not being apprehended, with $g\epsilon[M, F]$ denoting gender.

$$\begin{aligned} logCrime_{gt} = \beta_{0g} + \beta_{1g} logEarnings_{t-1} + \beta_{2g} log\left(Probability_{t-1}\right) \\ + \beta_{3g} Race_t + \beta_{4g} Age_t + \beta_{5g} Year_t + \varepsilon_t \end{aligned} \tag{22}$$

$$\widehat{logCrime}_{Ft}^{Counterfactual} = \widehat{logCrime}_{Ft} + \left(\hat{\beta}_{1M} - \hat{\beta}_{1F}\right) logEarnings_{t-1} + \left(\hat{\beta}_{2M} - \hat{\beta}_{2F}\right) log\left(Probability_{t-1}\right) \tag{23}$$

The objective of this part of the analysis is to find whether men and women respond differently to expected earnings and probability of arrest when deciding whether to commit a crime and to find how much of the participation gap can be explained by these different responses. The Blinder-Oaxaca decomposition method [Blinder, 1973, Oaxaca, 1973, Jann, 2008] is used to estimate a counterfactual crime equation for female offenders in the case where they have the same estimated coefficients as male offenders, shown in Equation 23. The results give a decomposition of the participation gap in crime, one component is due to differences in endowments and the other is due to differences in elasticity of offender behaviour with respect to changes in the expected earnings or probability.

To carry out this analysis, the repeated cross-sectional dataset from the CSEW has been transformed into a synthetic panel which groups individuals into cohorts sharing similar characteristics. Specifically, they are grouped according to race, gender and age, creating 30 cohorts per year who are tracked over time. Cross sectional characteristics in this type of synthetic panel are consistent over time, since these are the characteristics on which they are constructed, so individual fixed effects are redundant as they would have perfect collinearity with the dummy variables for age and race.

Table 3: Summary Statistics for Synthetic Panel

Variable	Observations	Mean	S.D	Min	Max
Year	420	2007.5	4.036	2001	2014
Log Crime (F)	169	9.382	1.856	5.497	12.620
Log Crime (M)	187	10.821	1.936	6.596	14.090
Log Earnings	336	5.072	1.130	1.609	8.956
Log Probability	338	-0.334	0.229	-1.731	-0.027

2.7 Results

Risk Preferences

Table 4 shows the results for an OLS regression of the log of criminal earnings on offender characteristics. The unconditional log earnings for female offenders are 35% lower than for males, but when other attributes and the offence types are controlled for, this earnings gap falls to 4%. This implies that a large part of the gender gap in earnings is due to gender-specific sorting in crime and the results are very similar to those found by Gavrilova and Campaniello [2015].

Table 4: Gender Earnings Gap

	Log(Earn)	Log(Earn)	Log(Earn)	Log(Earn)	Log(Earn)
	(1)	(2)	(3)	(4)	(5)
Female	-0.353***	-0.234***	0.070***	-0.073***	-0.041***
Age=16-24		0.622***	0.316***	0.087***	0.259***
Age=25-39		1.349***	0.877***	0.799***	0.947***
Age=40+		1.516***	0.447***	0.423***	0.581***
Drug Influence			0.292***	-0.023***	-0.072***
Drink Influence			-0.018***	0.195***	0.131***
Rel=Family			-0.693***	-0.667***	-0.769***
Rel=Friend			-1.292***	-1.019***	-0.918***
Rel=Acquain.			-1.583***	-1.171***	-0.772***
Rel=F. Spouse			-0.798***	-0.255***	-0.325***
Race=Black			0.151***	-0.111***	-0.091***
Race=Asian			1.023***	0.121***	0.124***
Weapon				0.380***	0.398***
Contact				-0.065***	-0.099***
Force				-0.161***	-0.738***
Threaten				-0.217***	0.108***
Sexual				0.692***	1.285***
Knew Offender				0.374***	0.646***
Street Gang				0.044***	0.037***
Time=Night				0.006***	0.076***
Weekend				-0.030***	-0.116***
Constant	-4.419***	3.298***	5.335***	4.701***	4.572***
R Squared	0.027	0.0368	0.1549	0.1161	0.2039
Observations	10,132,691	6,951,226	2,174,878	908,503	845,274
Year dum.	Yes	Yes	Yes	Yes	Yes
Offence dum.	No	No	No	No	Yes

The same OLS regression is run using earnings risk as the dependent variable to find whether female offenders tend to choose offences with lower earnings risk. Table 5 shows the estimated results when using Gavrilova and Campaniello's specification, given by Equation 24. The estimated coefficient for the female dummy is positive for the US dataset and negative for the CSEW dataset, but both are relatively small in magnitude.

$$EarningsRisk_i = \alpha_i + \beta_1 Female_i + \beta_2 Age_i + \beta_3 Weapon_i + \beta_4 Female_i * Weapon_i + \varepsilon_i \tag{24}$$

Table 5: Specification from Gavrilova and Campaniello [2015]

Earnings Risk	Gavrilova and Campaniello [2015] Coefficient	CSEW Dataset Coefficient
Female	0.004***	-0.091***
(Age) Age=10-15	0.000***	-0.813***
Age=16-24	-	-1.225***
Age=25-39	-	-1.123***
Age=40+	-	-0.685***
Weapon	-0.001	0.434***
Female*Weapon	-0.015***	0.915***
Alone	0.014***	-
Female*Alone	-0.008***	-
Gang	0.018	-
Female*Gang	-0.014	-
Constant	1.772***	4.887***
Observations	9,205,070	16,029,381
R squared	0.521	0.0251
Year*Agency FE	Yes	-

The CSEW has the advantage that many more observable characteristics of the offenders are recorded in the survey and these are controlled for in the regression estimates outlined in Table 16.

The estimated coefficient for female is -0.213 and significant which indicates that female offenders generally choose offences with lower earnings risk and the magnitude of this effect is larger than in the previous specification. When considering the possible directions of causality, it seems unlikely that a crime type would have lower earnings risk because women are more likely to choose it. It is necessary to consider the possibility that an external factor causes some offence types to have both lower earnings risk and a higher rate of female offenders, for example the expected earnings. It would be reasonable to hypothesise that crime types with lower levels of expected earnings also display lower earnings risk, and women choose crimes with lower earnings potential than men. A simple estimate of Pearson's correlation coefficient for level of earnings and earnings risk, however, gives a value of 0.02, implying that there is no correlation between them. Of course this doesn't mean that the level of earnings isn't correlated with other types of risk, for example risk of apprehension. Assuming that the direction of causality is correct, the results support the idea that female offenders are more risk averse than males and sort themselves into crime types with lower earnings risk.

Gender Participation Gap

The results from the decomposition analysis are given in Table 6, with the endowment and coefficient effects in the second table. These estimates indicate that the average total crime for men is 78,042 and the average total crime for women is 15,214. The figures reported are the log total crimes, therefore total crime is calculated by taking the exponential of 11.265 and 9.630 for men and women respectively. If female behaviour had the same elasticity with respect to changes in expected earnings and probability of apprehension, there would be an average of 88,876 crimes committed by women each year, which is larger than the average number of male crimes. The calculation for this is shown in Equation 25.

$$
\begin{aligned}
& log(FemaleCrime) + CoefficientEffect \\
& \quad = 9.630 + 1.764 \\
& \quad = 11.395
\end{aligned}
$$

$$e^{11.395} = 88,876 \tag{25}$$

This result is very different in magnitude from that found by Gavrilova and Campaniello [2015] who suggest that differences in elasticities account for 40% of the gender participation gap. Here we are saying that, if elasticities were the same, the gender gap would in fact be reversed since they account for 108% of the gap.

The validity of this result could possibly be improved by considering limitations in the data and analysis methods. Individual offenders from the repeated cross section are split into cohorts according to gender, age and race and are treated as groups that can be tracked over time in a synthetic panel. In theory, the idea is that these groups will respond to incentives in the same way because of their shared characteristics. A more rigorous method may be to allocate offenders to cohorts by several different characteristics and to test which groupings most accurately split offenders into groups who respond to incentives in the same way. Characteristics used to split offenders into cohorts should be those which affect an individual's propensity to commit crime. In addition to those used in this analysis, these may include variables such as education, income, unemployment and number of children, which have been shown in previous literature to have a significant effect on crime [Gould et al., 2002, Machin and Meghir, 2004, Witt et al., 1998, Bartel, 1979]

Unfortunately, this level of information is unavailable in the CSEW, due to the nature of victim-based surveys. The number of descriptive characteristics about the offender is limited to the victim's observations, namely easily observable characteristics like age and race. Another key point to remember is that this analysis is only for property crimes with single offenders, it doesn't necessarily explain the participation gap for all crime types.

Table 6: Blinder-Oaxaca Decomposition

	Log Crime(Female)		Log Crime(Male)	
	Coefficient	S.E	Coefficient	S.E
Lag Log Earnings	-0.068	0.051	0.114***	0.041
Lag Log Probability	0.241	0.250	0.532**	0.204
Race=Black	-2.890***	0.108	-2.627***	0.088
Race=Asian	-3.587***	0.162	-3.125***	0.091
Age = 10 - 15	3.389***	0.309	1.887***	0.284
Age = 16 - 24	4.625***	0.314	4.139***	0.263
Age = 25 - 39	5.142***	0.318	4.396***	0.273
Age = 40 +	4.453***	0.325	3.528***	0.280
Constant	7.587***	0.324	9.398***	0.324
Observations	148		156	
R squared	0.8848		0.9390	
Year dummies	Yes		Yes	

Log Crime	Coefficient	S.E	P-Value
Differential			
Prediction (Male)	11.265***	0.139	0.000
Prediction (Female)	9.630***	0.148	0.000
Difference	1.635***	0.203	0.000
Decomposition			
Endowments	-0.096	0.193	0.621
Coefficients	1.765***	0.079	0.000
Interaction	-0.034	0.067	0.608

2.8 Conclusions

The large gender participation gap in crime has puzzled academics for many years, but very few Economists have researched this topic. The current chapter offers an analysis of the difference in risk preferences of male and female offenders using a novel measure of earnings risk also adopted by Gavrilova and Campaniello [2015]. Following this, a Blinder-Oaxaca style decomposition is used to examine the gap in participation and to find how much of this gap can be explained by preferences and behaviour. Propensity to commit crime is likely dependent on a multitude of factors for both men and women, many of which are discussed in previous literature. The analysis in this chapter seeks to examine the role of risk preferences and reactions to incentives, which may play a role in addition to other variables.

The results from the earnings risk analysis show that female offenders sort themselves into crime types where the earnings risk is 21% lower than the types that male offenders choose. Controlling for additional characteristics leads to an estimated coefficient which is larger in magnitude than in the specification used by Gavrilova and Campaniello [2015]. When calculating the earnings risk, it is important to keep in mind that the standard deviation of the residual will depend on which independent variables are included in the Mincer equation, some of the variance in the error may be due to omitted relevant variables. The validity of the earnings risk measure relies on the assumption that all relevant, human capital variables have been included in the equation and any unexplained variation is considered to be earnings risk. These results show that female offenders demonstrate risk averse behaviour and may explain why they tend to choose different types of crime.

The results from the second part of the analysis suggest that the entire participation gap can be explained by differences in the elasticity of male and female behaviour with respect to incentives and the gap would be reversed if women behaved in the same way as men. Compared to the results found by Gavrilova and Campaniello [2015], who find that this accounts for 40% of the gap, this finding is very striking indeed. This analysis could be extended by employing a more comprehensive data set with a larger number of offender characteristics available, such as income, education and number of children.

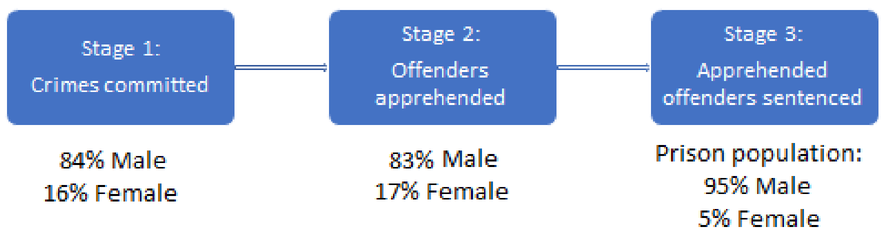
The main drawback of the data used in this chapter is the fact that all observations are from the victim's point of view. Their knowledge of events after the crime may be limited and as such, arrest and sentencing information about the offender are rarely recorded. The exclusion of all non-property crimes also means that the results found cannot necessarily be generalised to other crime types where the main pay-off to the offender is non-monetary.

3 THE GENDER GAP IN SENTENCING

3.1 Introduction

The previous chapter addressed the difference in offending rates between men and women and found that risk preferences may play a role in the large gender gap. When thinking about the gender gap in crime, however, there are three stages to consider in determining why the gap may be wide. The first stage is the subject of the previous chapter and that is the difference in the number of men and women who commit crime in the first place. This was analysed by focusing on the reasons why individuals commit crime or crimes and how different risk preferences might contribute to this gap. The second stage is the difference in the number of male and female offenders who are apprehended for crimes and reach the sentencing stage. During the year ending March 2017, 459,222 male offenders were cautioned or arrested in England and Wales compared with 91,603 female offenders. Although analysing this gender gap would be an interesting and logical next step in the overall analysis, data for this are insufficient since victims surveyed in the Crime Survey for England and Wales often do not know whether the offender was later apprehended for crimes committed against them. The third stage is the sentencing process and the difference in types and lengths of sentences given to male and female offenders. There is a possibility that these three stages are correlated, the gap in one stage may lead to the gap in the following stage, and these could cause individual studies of each stage to be misleading. However, the female percentage of offenders who were cautioned or arrested in 2017 is 16.63%, which is not significantly different to the percentage of total offenders who are female according to the Crime Survey for England and Wales.

Figure 5: Stages of the Gender Gap in Crime



This chapter, therefore, jumps to the third stage, which is the gender gap in sentencing, and seeks to find whether men and women are treated differently in the courts. While 16% of offenders are female and 17% of apprehended offenders are female, women make up only 5% of the prison population and tend to have shorter sentences when given a custodial sentence. The gender gap in prison population in England and Wales has persisted and risen over time, in 2010 there were 19 male prisoners for every female and this increased to 21 by 2016.

The analysis in this chapter aims to answer two key questions:

whether men and women receive different sentence types for the same offences and whether they receive different sentence lengths conditional on going to prison. Previous work in this area has focused on US data and does not control for many independent factors, such as offender characteristics, aggravating factors and mitigating factors. The analysis in this chapter is the first UK study and utilises a rich, individual-level dataset which includes several confounding factors which are recorded by the Judge after each Crown Court case.

Female sentencing has been discussed in the news in recent years, an example of which is the recent closure of Cornton Vale Womens' Prison in Scotland. This prison has been closed in an effort to redirect female offenders to rehabilitation facilities. There are, however, some papers that have tried to find whether women already receive more lenient treatment in the criminal justice system [Butcher and Park, 2017, Starr, 2015, Rodriguez et al., 2006, Sorensen et al., 2014].

The tables in Appendix B.1 show sentence outcomes and lengths by gender and offence type, as recorded in the Crown Court Sentencing Survey for the years 2011 to 2014. The most striking aspect of these tables at first glance is how few women are in each table compared to men. These statistics, of course, do not control for the seriousness of offences or any other relevant factors but do give an initial idea of the vast difference in the numbers.

Record-level data from the Crown Court Sentencing Survey is used to estimate the effect of gender on sentence type and length within offence categories using Ordered Logistic Regression (OLOGIT) analysis. The richness of the dataset allows one to control for several factors including the number of previous convictions and any aggravating or mitigating circumstances which affect the Judge's decision. The results show that, for all offence types bar sexual offences, women are significantly less likely to receive a custodial sentence than men. Similarly, for the length of the sentence, the probability of receiving a longer sentence is higher for men in most offence categories than for women. For a burglary offence, for example, women have a 37% chance of receiving a sentence of 12 months of less, this falls to only 28% for men[1].

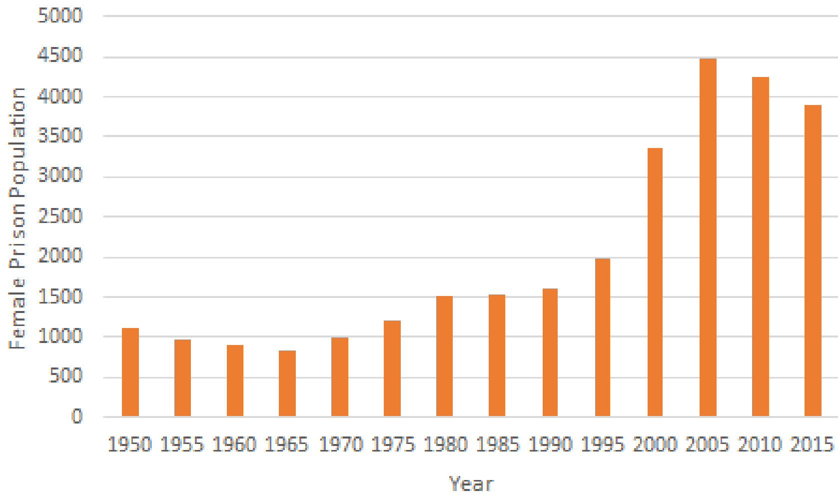
These results are consistent with findings from previous papers, which generally find that women are less likely to receive custodial sentences and receive shorter ones when they do. A problem that many papers face is that such a small proportion of the prison population is female, it can therefore be difficult to make robust conclusions about the differences between outcomes for men and women. This problem is overcome by using a large, record-level dataset which provides many details about both male and female offenders.

[1]Pearson's Chi-Squared test is used to test the statistical significance of these differences and the results are given in Appendix B.3

Some argue that women receive lesser sentences for reasons such as less serious crimes, responsibility for children or non-primary roles in offences. The analysis in this chapter shows that, even when these factors, and several others, are taken into account, women still receive more lenient treatment than men. Another possible explanation is judicial bias, which has been studied by Park [2014] in relation to offender race. He uses rank order tests to find whether judges choose sentences based on taste based or statistical discrimination and the possibility for this type of analysis with data for England and Wales is discussed in this chapter.

Figure 6 shows the female prison population for England and Wales for the years 1950 to 2015. The overall upward trend since the 1950s has sparked interest among academics in female sentencing, and crime in general, and many Economists have considered this gender gap in recent years [Butcher and Park, 2017, Starr, 2015, Park, 2014].

Figure 6: Female Prison Population



3.2 Literature Review

There has been an upward trend for both the male and female prison population over the past fifty years and academics have become more interested in the possibility of gender bias in sentencing. In order to test whether men and women receive different sentences for the same crimes, it is necessary to control for other determining factors. Previous papers have focused on datasets from the US and some have examined the effect of ethnicity and age on sentencing in addition to gender.

The first question to ask is whether , in fact, there is a gender gap in sentencing. Some would argue that sentence type and length should be determined by factors such as the type of crime and the number of previous convictions, however some papers have shown that this is not always the case. Butcher and Park [2017] use a decomposition method to find that 30% of the gender gap in incarceration rates cannot be explained by observable characteristics, which is consistent with the findings in this chapter. The analysis uses data from Kansas in the years 1998 to 2011 and focuses on two outcome variables: a dummy for incarceration and, conditional on incarceration, the sentence length, although the gap in sentence lengths is largely explained by observable factors. By presenting the sentencing guidelines for Kansas, the authors show that there is scope for judicial discretion since the guidelines offer a range rather than a specific length for each combination of previous convictions and offence severity and they are also able to adjust sentences according to aggravating and mitigating factors.

The authors use regression analysis, which controls for case facts and criminal history, to find that female offenders still receive more lenient sentencing than men. Equation 26 shows the logistic regression used in their analysis, where Y_i is equal to 1 for a custodial sentence, F_i is a gender dummy and X_i is a vector of control variables.

$$Y_i = \alpha + \beta F_i + X_i \gamma + \epsilon_i \tag{26}$$

These regressions are, however, only run for "drug offences" and "non-drug offences" within which there may be large variation in crime type. They also do not control for child care as a mitigation factor which is likely to be significant in the case of female offenders but they acknowledge that the unexplained gap found later in the paper may be partly due to the majority of parents in prison being female.

Following this, they use a decomposition method to create a counterfactual distribution of prison terms for women to find whether the gender gap in sentencing can be explained. They find, using this method, that 30% of the gap in incarceration rates is unexplained by observable factors but there is only a very small unexplained gap in sentence length conditional on incarceration. It is shown that there is heterogeneity across judges when it comes to female sentencing and the authors speculate that judges differ in their punishment philosophies. They find no evidence of chivalry, concluding that those judges who are more lenient towards women are also more lenient towards men.

This decomposition method has also been used by Starr [2015] who finds that women are favoured in sentencing and a large gap in sentence lengths is unexplained when controlling for other factors, men receive sentences which are 63% longer on average.

Evidence of gender bias in sentencing has been found in several other papers, all finding that women are either less likely to receive a custodial sentence, receive shorter custodial sentences or both.

Park [2014] uses the rank-order to find whether judges in Kansas use statistical discrimination or taste-based discrimination. Taste-based discrimination occurs when a decision is made purely on the individual's taste for a particular observable factor, whether that's gender, race, age or any other characteristic. Statistical discrimination, however, occurs when the individual cannot observe a particular variable so instead they base their decision on something they can observe which they know is correlated with the unobserved variable. If, for example, a Judge cannot observe the future criminality of an offender, but they do know that women are less likely to commit crime than men, they may base their decision on gender. Park finds that judges do not have consistent rankings for each race and concludes that this is evidence of taste-based discrimination. Both Sorensen et al. [2014] and Rodriguez et al. [2006] find that women receive more lenient sentences in US even when controlling for other factors. Rodriguez et al. [2006] analyses sentencing data from Texas to find that the prevalence of gender bias in sentencing varies across different types of crime. Custodial sentence is less likely for women who commit drug or property offences. For violent crimes, they are not less likely to receive a custodial sentence, but they are likely to be shorter when they do.

Other papers have studied the possibility of sentencing bias for groups other than women and find that they are possibly not the only group to receive lighter sentences. Mustard [2001] finds that black and male offenders are more likely to receive longer prison sentences than white and female offenders. In addition to race and gender, having children may also be a characteristic that reduces the severity of an offenders sentence. Pierce and Freiburger [2011] find that having children reduces the probability of receiving a custodial sentence if they are charged with child neglect. Tillyer et al. [2015] finds evidence for both the "chivalry" and "evil woman"[2] hypotheses, women with low criminal history receive more lenient sentencing whereas women with higher criminal records receive more severe sentences.

Despite evidence that women are treated more leniently by the criminal justice system, some suggest that there should be fewer female prison sentences as women commit less serious crimes and often struggle to keep custody of their children when in prison. Corton Vale, the only womens' prison in Scotland, is closing with a view to reduce the number of female prisoners and offer alternative sentencing options [BBC, 2016]. Since women generally commit less severe crimes [Butcher and Park, 2017], there is an argument that female offenders should receive non-custodial punishments. This view is

[2]The "chivalry" hypothesis and the "evil woman" hypothesis are terms used in criminology. The "chivalry" hypothesis was presented by Crew [1991] and suggests that the male-dominated justice system sees women as victims and therefore treats them with more leniency. Erez [1992] presents the "evil woman" hypothesis that women are, in fact, treated more harshly in the criminal justice system because men believe that they have not only broken the law but gone against expected female behaviour.

supported by Bagaric and Bagaric [2016] and Gelsthorpe and Morris [2002], who argue that women commit less serious crimes than men and most women in prison are not a risk to others, so the number of women being imprisoned should be reduced.

Another argument against female imprisonment is the strong correlation between being female and being a single mother. At the end of 2016, there were 2.9 million single parent families in the UK, 86% of which were headed by mothers. The Crown Court Sentencing Survey records whether an offender is the main or sole carer for dependent relatives, including children. Table 29 gives the number of male and female offenders in each offence category who have responsibility for dependent relatives as well as the percentage of the total and a Pearson χ^2 significance test of the difference. It is clear from this table that for all offence types, excluding offences causing death, the percentage of female offenders who have responsibility for dependent relatives is significantly higher than the percentage of male offenders.

It is therefore logical when thinking about female offenders to consider the effects of imprisonment on their children. Cho [2009] uses propensity score matching to examine the impact of mothers in prison on educational achievements of children. They find, surprisingly, that childrens' school work does not suffer when their mothers go to prison, in fact they find a very slight positive effect. However, they hypothesise that this may be the effect of sympathetic teachers encouraging children with imprisoned mothers and not necessarily the true effect on children's education.

In addition to the effects on children, studies have explored the effects of prison on female employment and welfare. Lalonde and Cho [2008] study a dataset from Illinois to find whether employment prospects are worsened for women after prison. They find that, in the short term, having been in prison has no adverse effect on employment prospects, employment rates simply return to their pre-prison level. Since pre-prison employment for female offenders tends to be lower than the national average, they conclude that going to prison is not the cause of lower employment, rather this group of women had lower employment levels regardless. Butcher and Lalonde [2006] analyse the effect of imprisonment on womens' subsequent social welfare receipts. Contrary to popular opinion, the authors find that womens' welfare receipts do not increase after prison. Rather women who are imprisoned tend to claim higher levels of benefits anyway and there is also a tendency for welfare receipts to drop just prior to imprisonment, so it may appear that they increase after but in fact they return to their previous levels.

3.3 Data

The Sentencing Council are responsible for providing and monitoring sentencing guidelines in England and Wales and are an independent body within the Ministry of Justice[3]. The sentencing guidelines do not specifically mention gender and therefore should not directly contribute to the gender gap in sentencing. In an effort to assess the effectiveness of current guidelines in 2010, the Council began to request that Judges complete a survey after every case in Crown Courts across the country. The survey ran until mid-2015 and has provided a rich dataset at the individual offender-level. The survey covers several aspects of the case, including offender characteristics and previous convictions, and details of the factors which influenced the Judge's decision during sentencing.

The tables in Appendix B.1 presents some descriptive statistics for the variables used in the analysis. The offender characteristics used are age and gender, both given as factor variables. Age falls into one of four categories: "18 to 24", "25 to 34", "35 to 44" and "45 to 54". The category "54 and over" has been eliminated from the analysis since such a small percentage of offenders fell into it, this lack of variability in the X variables can prevent the model from

[3]The sentencing guidelines for England and Wales can be found on the Sentencing Council's website: https://www.sentencingcouncil.org.uk/about-sentencing/about-guidelines/.

running due to their low predictive power. Gender is recorded as a binary variable with 0 for men and 1 for women.

The other independent variables common to all offence categories are "number of previous convictions", either "none", "1 to 3" or "4-9", and "guilty plea discount", either "none", "1% to 10%", "11% to 20%", "21% to 32%" or "33% or more". In each dataset, there are also offence-specific variables, theses include the seriousness of the crime, the subcategory offence and any aggravating or mitigating circumstances. As noted, where a binary variable has a value of 1 for a very small number of cases[4], this variable must unfortunately be excluded from the regression since a certain level of predictive power is required from explanatory variables in order to estimate coefficients.

3.4 Estimation Methodology

The nature of the data collected by the survey means that both the outcome and length of an offender's sentence is categorical rather than continuous. Judges are asked to choose one of the options shown in Table 7 and this choice is coded accordingly in the dataset. Details such as gender, age, criminal history and crime characteristics are recorded and coded in a similar way, meaning that

[4]This analysis uses 10% as a rule of thumb, since this is the percentage below which the model failed to run.

the majority of variables used in this analysis are categorical, also known as factor variables. Furthermore, the differences between the values assigned to each option are not representative of the differences between the options themselves, where “Immediate Custody” is assigned a value of 4 and “Community Order” is assigned a value of 2, this does not mean that a custodial sentence is twice as bad as community service. Therefore, these variables can be described further as ordinal categorical variables, for which there are specific modelling techniques.

Table 7: Values for Categorical Dependent Variables

Value	Sentence Outcome	Sentence Length
1	Other	Up to 1 year
2	Community Order	1 to 3 years
3	Suspended Sentence	3 to 5 years
4	Immediate Custody	5 years or more

Before considering regression models that are appropriate for ordinal dependent variables, it is important to understand why Ordinary Least Squares (OLS), and linear models in general, are inappropriate. Linear regression models require normality of errors with constant variance, which is violated by ordinal categorical variables due to the fact that they are non-continuous, bounded and cannot be measured on an interval or ratio scale. Therefore, estimating an OLS model for an ordinal dependent variable will lead to biased estimates, see the analysis carried out by Winship, C. & Mare [1984] and McKelvey and Zavoina [1975].

A superior method of estimation in this case is the Ordered Logit Model (OLOGIT), also known as the Proportional Odds Model. Estimating this type of model for a dependent variable with M levels is analogous to estimating a series of $M-1$ binary logistic regressions with grouped values of the dependent variables. This idea is illustrated more clearly in Table 8, where an example is given for a dependent variable with four levels, that is $M=4$. In this example, the OLOGIT model estimates three logistic regressions as shown in the first column and assigns a binary value to the dependent variable according to the values given in the second and third columns. In the first regression, the binary dependent variable is 0 for a value of 1 and 1 for values 2, 3 and 4. Effectively, the estimates for the coefficients in this first regression represent the effect of an increase in the independent variables on the odds of the dependent variable having a value of two, three or four rather than one. It follows that the estimated coefficients for the second and third regressions give the effects on the odds of the dependent variable having values 3 or 4 and 4 respectively.

Table 8: Ordered Logistic Regression Example

Logistic regression	Dep. variable = 0 if...	Dep. variable = 1 if...
1	$Y=1$	$Y=2,3,4$
2	$Y=1,2$	$Y=3,4$
3	$Y=1,2,3$	$Y=4$

Equation 27 gives the OLOGIT model as described by Williams [2016], where $j = 1, ..., M$ is again the number of levels for the dependent variable and X_i is a vector of observed explanatory variables. As described in Table 8, a set of logistic regressions is estimated to find the effect on the odds that the dependent variable has a value higher or lower than a particular cut off. To estimate an OLOGIT model, the proportional odds assumption must be met, which requires that the estimated log odds are the same regardless of the cut off. As can be seen in Equation 27, α and β do not have a j subscript, since only one set of coefficients is estimated.

$$P(Y_i > j) = \frac{exp\,(\alpha + X_i\beta)}{1 + [exp\,(\alpha + X_i\beta)]},\ j = 1, 2, ..., M - 1 \tag{27}$$

In practice, this assumption is very strong and violation can be tested for in two ways. The Brant test uses the idea that, if the proportional odds assumption holds, the estimated coefficients should not be significantly different from one logistic regression to the next. It calculates a χ^2 statistic for each explanatory variable and one for all, showing which, if any, violate the assumption. If the overall test is insignificant, OLOGIT is the appropriate model. If, however, the overall test is significant, the individual tests can be used to identify which explanatory variables do not meet the assumption, that is their estimated coefficients differ across regressions. The second way to test for violation is the Likelihood Ratio Test, which is similar to the Brant test but tests the assumption by estimating the model given by Equation 27 and the generalised model given by Equation 28 and comparing the goodness of fit for each. Again, an insignificant test statistic implies that the proportional odds assumption is not violated.

Equation 28 gives the model for the Generalised Ordered Logit Regression (GOLOGIT), of which OLOGIT is a special case. As can be seen in the model, the proportional odds assumption is relaxed and a separate set of coefficients is estimated for each logistic regression. This model is suitable when the above tests indicate that the assumption for OLOGIT is not met.

$$P(Y_i > j) = \frac{exp\,(\alpha_j + X_i\beta_j)}{1 + [exp\,(\alpha_j + X_i\beta_j)]}, j = 1, 2, ..., M - 1 \tag{28}$$

The disadvantage of this, however, is that there are now $M - 1$ sets of coefficients to interpret and work with, rather than just one. The most commonly used model is a mixture of the two discussed called the Partially Constrained Generalised Ordered Logit Regression (PC-GOLOGIT) [Williams, 2006, 2016, Long, 1997]. Equation 29 gives an example of the model given by Williams [2016], where the estimated coefficients β_1 and β_2 are constrained and β_3 is unconstrained. This model is often the most appropriate in practice as it can accommodate both OLOGIT and GOLOGIT models where needed.

$$P(Y_i > j) = \frac{exp\,(\alpha_j + X1_i\beta1 + X2_i\beta2 + X3_i\beta3_j)}{1 + [exp\,(\alpha_j + X1_i\beta1 + X2_i\beta2 + X3_i\beta3_j)]}, j = 1, 2, ..., M - 1 \tag{29}$$

The regression models described above are designed for categorical dependent variables and therefore lend themselves well to the analysis in this chapter. The aim of this analysis is to find whether men and women receive different sentence types and lengths for the same crime types when controlling for other variables. The contribution of this paper over previous studies is that it controls for a variety of specific aggravating and mitigating factors, outlined in Appendix B.1, which influence the Judge's sentencing decision. The inclusion of these additional confounding factors helps to achieve a comprehensive assessment of the research question.

First, to find whether sentence types differ between men and women, Equation 30 is estimated separately for every crime type as a PC-GOLOGIT model.

$$\begin{aligned} Outcome_i = {} & \alpha_i + Gender_i + Age_i + Offence_i \\ & + PreviousConvictions_i \\ & + Seriousness_i + AggravatingFactors_{ij} \\ & + MitigatingFactors_{ik} + GPDiscount_i + \epsilon_i \end{aligned} \tag{30}$$

The variables in Equation 30 are described in detail in the tables in Appendix B.1. The dependent variable is the sentencing outcome which is categorical. Gender, age, offence, previous convictions, seriousness and guilty plea discount are also categorical variables which are recorded for each individual offender. Aggravating and mitigating factors are vectors of factors which positively or negatively affect the judge's sentencing decision and these are also described in detail for each offence type in Appendix B.1.

If it is the case that, all other things being equal, men and women receive different sentence types for the same crimes, we would expect to see a negative and statistically significant estimated coefficient for the female dummy in Equation 30. If these estimated coefficients are statistically insignificant, we can conclude that it is unlikely that differences in sentence types are due to gender. Similarly, to find whether custodial sentence lengths depend on gender, Equation 31 is estimated separately for each crime type as a PC-GOLOGIT model.

$$\begin{aligned} Length_i = \alpha_i + Gender_i + Age_i + Offence_i + PreviousConvictions \\ + Seriousness_i + AggravatingFactors_{ij} \\ + MitigatingFactors_{ik} + GPDiscount_i + \epsilon_i \end{aligned} \tag{31}$$

The sentence length is, of course, only relevant for those offenders sentenced to immediate custody and is therefore only estimated for a subset of the individuals. The estimated coefficients should be interpreted as the effects on sentence length conditional on receiving a custodial sentence.

3.5 Results

As explained in the previous section, the OLOGIT models relies on the assumption of proportional odds and only estimates one set of coefficients. Since this assumption is violated for every offence type in the dataset[5], the appropriate estimation model is the PC-GOLOGIT regression, where only those variable which satisfy the assumption are limited to one estimated coefficient. Using data from the Crown Court Sentencing Survey, it is possible to control for independent factors which include the offender's age and gender, the number of previous convictions, any mitigating or aggravating circumstances and the percentage discounted in length when a guilty plea is entered. By using these controls, any significant differences found between male and female offenders are less likely to be due to differences in characteristics or circumstances, particularly in the case of offenders who are the sole carer for children. Unfortunately, as shown in Table 29, the percentage of offenders who are sole carers is very low for almost all offence categories. This means that the variable has very low predictive power and the model failed to run in the software when it was included. Therefore, this variable is only included for the offence type "theft and fraud" where 5.6% of male offenders and 18.8% of female offenders are sole carers.

The detailed output tables for these estimated regressions are presen-

[5]Both the Brant test and the Likelihood ratio test are carried out for every offence type and produce insignificant test statistics.

ted in Appendix B.2, where models for sentence outcome and length are estimated for each crime category. For arson and criminal damage offences, the offence type regression output is shown in Table 30. Starting with the variable of interest, the gender dummy, the estimated coefficient is -0.555 for every level of the dependent variable. It indicates that this particular variable does not violate the proportional odds assumption and it is therefore appropriate to estimate one coefficient rather than three. The estimate is also statistically significant which tells us that, taking all other independent variables into account, being female reduces the odds that an offender will receive a higher sentence type for arson and criminal damage offences. Similar results can be found for assault offences in Table 32, burglary offences in Table 34, death offences in Table 36, driving offences in Table 38, other offences in Table 42 and robbery offences in Table 44.

For drug offences, shown in Table 40, three different coefficients are estimated for the gender dummy with only two negative and significant. The coefficient in the first column is 0.088 and insignificant, which implies that being female does not affect the odds that the sentence type will be higher than level one (community order, suspended sentence or immediate custody), all other things being equal. Similarly for theft and fraud offences, the estimated coefficient in the first column is insignificant but the others are negative and significant. In Table 46, a positive and insignificant coefficient is estimated for the gender dummy at every level. A coefficient of 0.415 with a p-value of 0.164 indicates that gender has no effect on the odds of a particular type of sentence occurring. However, only 1.35% of sexual offenders in the dataset are women[6] which may mean that the effect of gender on sentence outcome is not fully realised in the model for this particular offence type.

In addition to the type of sentencing outcome, PC-GOLOGIT models are also used to analyse the effect of gender on the length of immediate custodial sentences. The female dummy has a significant estimated coefficient for four of the ten offence types and these are all negative. These offence types are assault in Table 33, burglary in Table 35, drugs in Table 41, robbery in Table 45 and theft and fraud in Table 49. These estimates imply that, when an immediate custodial sentence is given for these offences, female offenders are less likely to receive longer sentences than men, all other things being equal. For the remaining offence types, the estimated coefficients are insignificant which implies that gender does not influence sentence lengths for these crimes, although these do seem to be the crime types for which the number of women receiving custodial sentences is very low, sometimes in single digits, so it may be the case the the effect of being female on sentence length is not fully realised.

Although these estimated coefficients are useful in suggesting the sign and significance of the effects of gender, the magnitude of the

[6]See Table 27

log odds are themselves difficult to interpret. A more intuitive way of presenting these findings is to calculate the marginal effects, which give the probability of a certain outcome, holding all other variables constant. This provides a simpler way to compare effects across crime types and allows us to interpret how likely each outcome is. Detailed output tables showing the marginal effects for every crime type are presented in Appendix B.3 in Tables 50 and 51. It is clear from Table 50 that, all other variables being equal, female offenders have a lower probability of receiving a custodial sentence than men for almost all offence types. They only positive difference shown is for Sexual offences, which may again be due to the small number of female offenders in this offence category. The results show consistently that women are more likely to receive lower sentence types and less likely to receive custodial sentences. The "Difference" column gives the χ^2 statistic for the statistical significance of the difference between the probabilities for men and women and the corresponding p-value.

Table 51 presents the marginal effects for the estimated regression models for sentence length, conditional on a custodial sentence. All other variables being equal, the probability that a male or female offender receives a custodial sentence of a particular length is given for each offence type. For arson and criminal damage offences, for example, the probability that a male offender will receive a sentence up to 1 year long is 25%, whereas for a female the probability is only 12 %. From the results in this table, it is clear that for the majority of offence categories women are less likely than men to receive longer sentences and more likely to receive shorter sentences. These probabilities are calculated while controlling for other factors so the differences are unlikely to be due to different characteristics or circumstances.

The probabilities presented in Table 50 are illustrated in Figure 7 to demonstrate the persistence of the gender effect. Excluding sexual offences, the probability that the sentence type is "otherwise dealt with" is always higher for women and for men and vice versa for immediate custodial sentences. This pattern also holds for sentence lengths in the majority of offence types shown in Figure 8.

A unique feature of this analysis is the inclusion of several aggravating and mitigating factors which influence the Judges' decisions. If the offender's gender influences this decision, it's important to find whether there is an interaction effect between gender and these other influencing factors. However, when the female dummy is interacted with each of the factors for each crime type, none of the estimated coefficients are statistically significant[7], insinuating that it is gender alone and not the combination of gender and another factor, which is driving the gender gap in sentencing. These insignificant results are interesting in themself because it suggests that, if women are

[7] An example of these estimated coefficients for interaction terms are illustrated for Arson and Criminal Damage in Table 52.

receiving more lenient sentences than men, it is because they are female and not because they are female with a particular mitigating circumstance.

Figure 7: Marginal Effects for Sentencing Outcomes

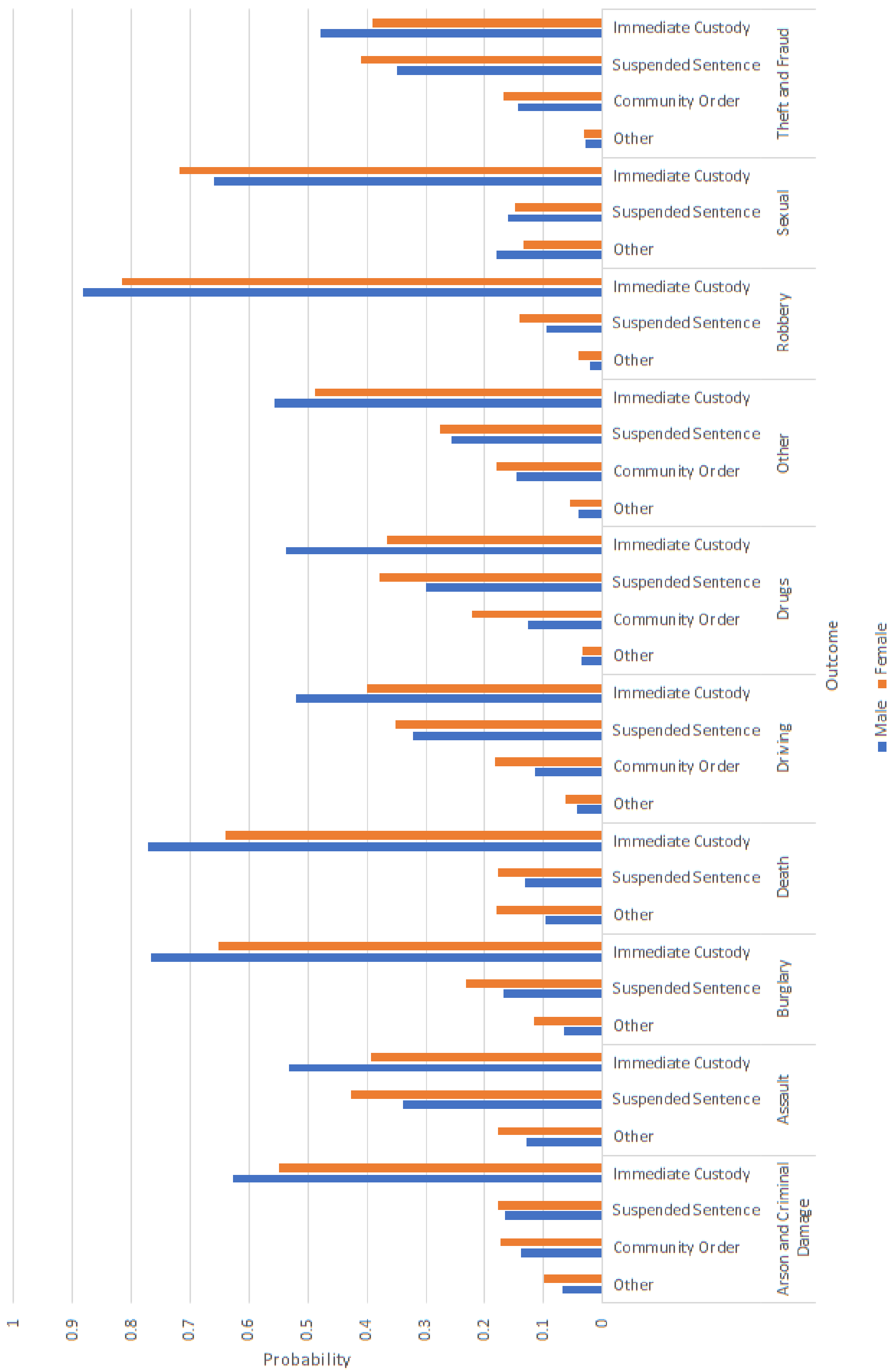
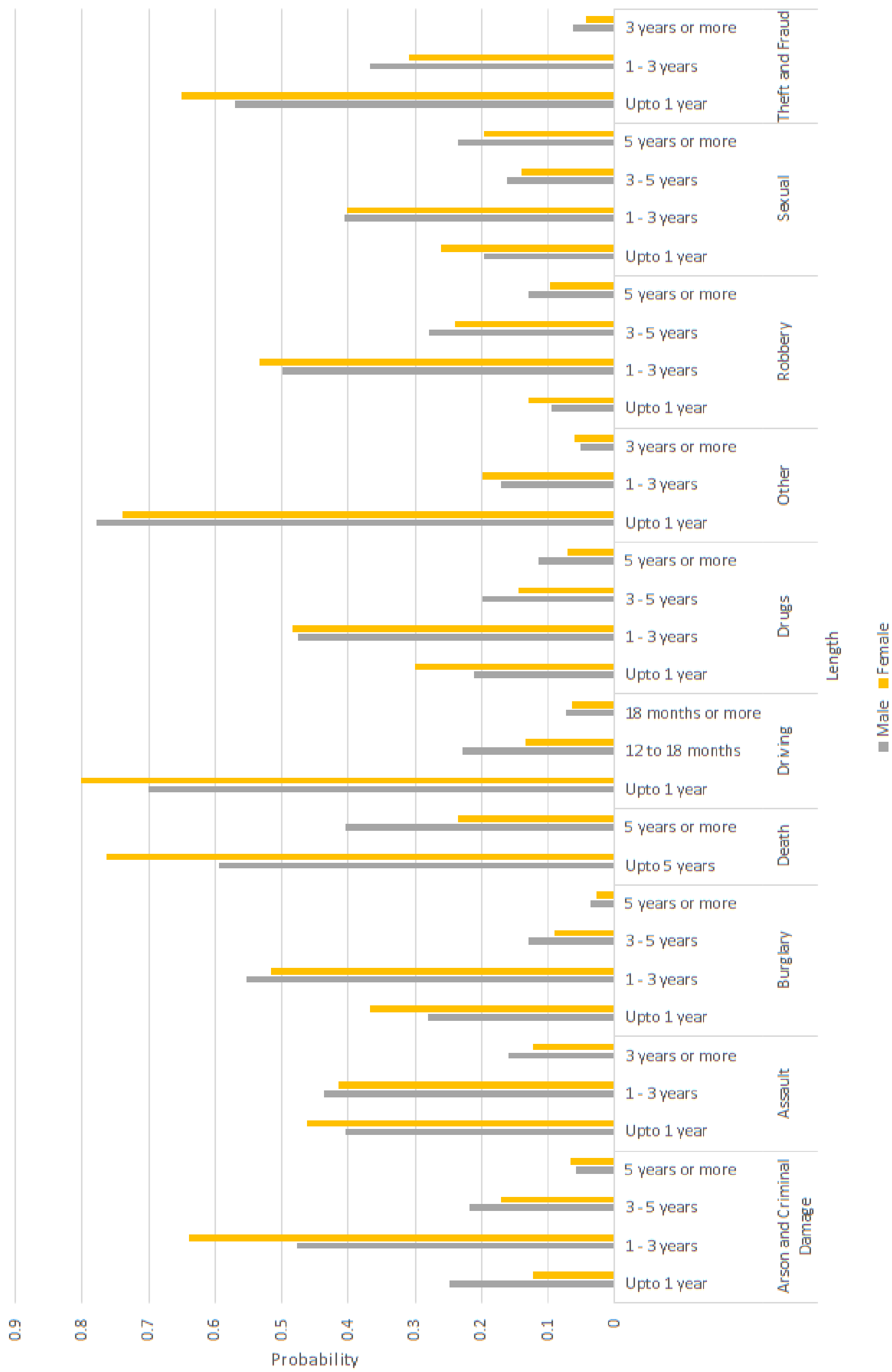


Figure 8: Marginal Effects for Sentence Lengths



3.6 Conclusions

Despite the various guidelines and regulations surrounding sentencing in England and Wales, there is still much discussion around gender gaps and bias. Some argue that women commit the least serious offences and pose the least risk to the public, so they should therefore be offered non-custodial sentences, particularly in cases where the offender is solely responsible for the care of children. However, others claim that women are already treated more leniently than men by the Criminal Justice system, citing the vast gender gap in the prison population and various news stories which show women receiving low sentences.

The aim of the analysis in this chapter is to use a unique, individual-level dataset to find whether there really is a gender gap in sentencing in England and Wales when other relevant factors are accounted for. Controlling for other explanatory variables which influence the Judge's decision allows us to isolate the gender gap. For example, if women are generally awarded shorter sentences because they care for children, controlling for this should eliminate the gender gap in the estimated model.

Others in economics have explored the possibility of disparities in sentencing choices between male and female offenders but none have used the far-reaching list of relevant factors used by the Crown Court Sentencing Survey. The econometric analysis method used is a regression model designed specifically for ordered categorical dependent variables and this model is employed to find the effect of gender on sentence outcome and length.

The results from the empirical analysis show that the sentencing outcomes for women are lower than for men even when aggravating and mitigating factors are controlled for. Conditional on an offender receiving a custodial sentence, women generally receive significantly shorter sentences than men. The probability of receiving a custodial sentence is significantly higher for men than for women for all offence types except sexual offences. Conditional on receiving a custodial sentence, the probability of receiving a longer sentence is significantly higher for men than women for all offence types except sexual and other offences.

Unlike a simple comparison of sentence lengths, these probabilities control for all other factors and are calculated assuming mean values of all other variables. It is important to keep in mind that there may be unobserved variables not considered here which contribute to the gap in sentencing, but the variables outlined in this dataset are representative of all factors indicated on the Sentencing Survey by the Judge.

There is potential for future work on this topic since the established gender gap in sentencing naturally leads to the question of what causes the gap. Judicial bias is one potential explanation which has been explored in some of the literature using US data. In order to investigate this further for England and Wales, individual cases would need to be grouped by Judge in order to analyse any sentencing patterns and to determine whether taste-based or statistical discrimination exists, similar to the analysis carried out by Park [2014].

4 THE SEVERITY OF ROAD TRAFFIC ACCIDENTS

4.1 Introduction

In 2015, 22,144 people were seriously injured and 1,730 people were killed in road traffic accidents in Great Britain [Department for Transport, 2015]. In addition to the human cost of road traffic accidents, there are medical costs and lost output which contributed to a total cost of £35,550 million. Table 9 shows the breakdown of this cost by severity and cost type.

The aim of the research in this chapter is to analyse factors which affect severity of accidents in Norfolk and Suffolk by using individual-level accident data to identify groups of drivers who are most at risk of being involved in a traffic accident. Several variables are found to significantly effect the severity of an accident, including gender, wearing a seatbelt and visibility. Drivers found to be most at risk of being involved in an accident are those aged 17-39 and female drivers over the age of 70.

These results would be useful for policy-makers when considering which groups and situations to target with new interventions. It allows us to focus on drivers most at risk as well as run campaigns increasing awareness of behaviour that is likely to increase the severity of road traffic accidents.

Table 9: Reproduced from Department for Transport Statistics

£ million (2015 prices)

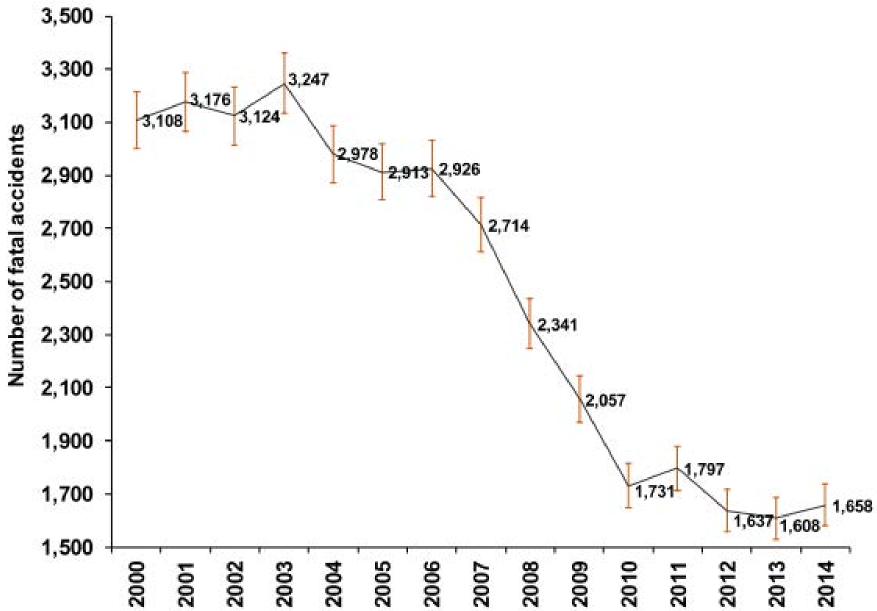
	Cost Elements						
	Casualty related costs			Accident related costs			
Accident Severity	Lost Output	Medical and Ambulance	Human Costs	Police Costs	Insurance and Admin	Damage to Property	Total
Fatal	1,073	9	2,107	32	1	19	3,241
Serious	528	317	3,599	46	4	110	4,604
Slight	387	164	1,846	70	15	383	2,865
All injury accidents	1,989	491	7,552	147	19	512	10,710
Damage only accidents	0	0	0	81	125	4,370	4,577
Not reported	2,534	1,323	14,964	0	53	1,388	20,263
All accidents	4,523	1,814	22,516	229	197	6,270	35,550

4.2 Literature Review

Road traffic accidents have featured in many papers in the economics literature; why they happen and what we can do to prevent them are key questions for policy-makers. There are several topics within the literature which are relevant to this chapter, some which have received very little attention from Economists. Firstly, in order to decide how best to prevent car accidents, it is important to understand what causes them in the first place. The word "accident" implies that there is a certain degree of randomness in their occurrence as they are not the result of an individual's behaviour but rather an unwanted side effect [Fridstrom et al., 1995]. However, there are many factors which are likely to contribute to the probability of having an accident and these are explored in the following papers.

In a recent government report, the authors analyse accident data for Great Britain during the years 2000 to 2014, in order to investigate whether there had been a significant change in the number of fatal accidents during these years [Department for Transport, 2014]. By comparing the observed frequency of fatal accidents with the frequencies expected under the Poisson distribution, they show that the number of "Fatal" road accidents in Great Britain does indeed follow this distribution, which allows them to use statistical tests when comparing observations in different years. The authors only show that fatal accidents follow this distribution, the distribution of all accidents is examined later in this chapter and are shown to also follow a Poisson distribution, making it possible to test the significance in frequency changes by calculating Poisson confidence intervals. Given this information, the authors are able to calculate 95% confidence intervals for the fatal accident frequency in each year. Whether these confidence intervals overlap or not is a fair estimation of the significance or insignificance in the change in the number of accidents.

Figure 9: "Fatal" Road Accidents in Great Britain, 2000-2014, *Reproduced from Department for Transport [2014]*



The objective for policy-makers in this area is to find ways to prevent traffic accidents, or at least to reduce their severity. Several different interventions are used across the country and many have been evaluated in the literature. Most of these interventions can be placed in one of the following three categories; enforcement, engineering and education and each attempts to reduce traffic accidents.

Paola et al. [2010] use road accident data for Italy to estimate the effectiveness of the introduction of a penalty point system in 2003 on reducing traffic offences and accidents. They take advantage of this national policy change by using a regression discontinuity design to measure the causal effect. They find that this change in policy was very effective in reducing both accidents and offences.

In a similar study, Hashimoto [1979] aims to measure the impact of police surveillance on the prevention of road traffic accidents. They divide accidents into two categories, one where the driver was at fault and one where the pedestrian was at fault, they then discard accidents where the pedestrian was at fault in order to focus on the effects of surveillance on driver behaviour. He calculates the probability of vulnerable behaviour by drivers (behaviour which could result in a collision) and the conditional probability of a collision occurring, given that vulnerable behaviour occurs. He finds that the effectiveness of police surveillance depends on the type of collision. It has a positive impact for collisions which occur at pedestrian crossings between a vehicle and a pedestrian and collisions that occur when the driver is turning right. It is less effective for rear-end collisions, where the cause is often carelessness by the driver. The author identifies two positive outcomes of police surveillance, (i) improving the behaviour of the driver (which is temporary) and (ii) increasing the driver's alertness (which is a more durable effect).

Van Houten and Nau [1981] compare the effectiveness of two highway interventions on reducing speeding on two Highways in Nova Scotia. The first is a large sign which displays the percentage of drivers who didn't speed during the previous week. The second was increased police surveillance and ticketing, which included visible road-side radars. The analysis finds that the first intervention is very successful in reducing occurrences of speeding, while the second is not successful. This is an interesting result as police surveillance and ticketing required far more time and investment than displaying the percentage of drivers who didn't speed.

Drink-driving is also in important topic in accident prevention, Mercer [1985] looks at data from British Columbia to analyse the relationship between the number of drink-driving road checks, the number of drivers impaired and the number of alcohol-related traffic accidents. By looking at correlations between these three events, he finds that drink-driving road checks are only successful in reducing the number of drink-driving related traffic accidents when they are accompanied by wide-spread media coverage. However, they do not find the extent of media coverage necessary to achieve this reduction in accidents. Goss et al. [2008] find that police patrols have a slight impact on reducing collisions caused by drink driving, although the report highlights that academic literature in this area is very poor.

Speeding is often cited as a main cause of traffic accidents, Van Benthem [2015] examines the choice of a rational driver when deciding at which speed to travel, by considering both the private and public costs of increased speed. By considering many factors such as fuel cost, pollution and the risk of an accident, he concludes that the optimal speed of travel on highways in the US is around 55mph and therefore recommends that lowering the speed limit would reduce the burden of these social costs.

There are also several papers which study the effectiveness of speed cameras in reducing the frequency of road traffic accidents, which generally agree that they are successful. Hirst et al. [2005] calculate that for every 1mph speed reduction resulting for a speed camera, there are 4% fewer accidents on that road for roads where the speed limit is between 30-35mph, while Mountain et al. [2004] finds that a speed camera on a 30mph road can reduce accidents for a distance of 1km in each direction by 20%. Corrall, however, suggests that speed cameras should be installed more widely in order to assess their effectiveness more accurately, rather than focusing on "Killed and Seriously Injured" (KSI) hotspots. These papers only analyse the relationship for a narrow range of speed limits so the results cannot be generalised to roads where vehicles are travelling more quickly.

Li et al. [2013] use propensity score matching, rather than a simple before-after analysis, for the introduction of speed cameras in the UK in 1991. They find that speed cameras are most effective in reducing accidents up to to a distance of 200m from the site. They also find a lack of increased accidents before and after the speed camera, which they take to imply that drivers are consistently reducing their speed, rather than decelerating and accelerating again after the camera.

However, Blincoe et al. [2006] investigate attitudes towards speed cameras in Norfolk by analysing questionnaire results from drivers, split into four groups according to their driving and speeding behaviour. The research found that their is a prevalent speeding culture and many drivers are unhappy about the use of speed cameras. A limitation of the use of speed cameras seems to be that, according to UK guidelines, cameras must be highly visible to drivers, so it is very easy to slow down as you approach and speed up again when you have passed. There may, of course, have been a self-selection issue here as only 31% of drivers who were sent applications chose to respond to the questionnaire.

A common complaint by the driving population is that speed cameras are mainly used to generate revenue for the government. Tay [2010] tries to find whether this is the case by analysing the effectiveness of ticketing over and above police presence, to find whether raising revenue is really effective in reducing traffic accidents. The author estimates a Poisson regression model and tests whether the reduction in accidents due to ticketing is statistically significant. He finds that both speed camera operating times and issuing speeding tickets significantly reduce the number of accidents and advises policy makers to expand these programmes in order to reduce them further.

It is generally assumed that increasing car safety reduces the severity of car accidents. But Peltzman [1975] argues that drivers' lives saved by these improved safety devices are offset by an increase in the number of pedestrian deaths and non fatal accidents. The mechanism here is that, due to increased safety measures in cars, drivers are willing to take more risks when driving and there are an increased number of young and drunk drivers, both of whom have an increased probability of having an accident.

However, Peltzman's findings are rebutted two years later by Robertson [1977], who claims that Peltzman failed to consider data which support the success of car safety measures in reducing the number of accident related deaths. He corrects several apparent flaws in Peltzman's analysis and comes to the conclusion that the introduction of increased safety measures in the US in 1965 did reduce fatal accidents for drivers and motorcyclists and had no effect on pedestrian deaths. More recently, Traynor [2009] confirms this result by showing that policies concerning restricted teen driving and drink-driving significantly reduced the number of fatal accidents in the US between the years 1999 and 2003. He does, however, find that enforced wearing of seatbelts has a statistically insignificant effect on accident fatalities which may be due to the fact that wearing a seatbelt protects drivers and passengers whilst involved in a crash rather than preventing a crash. This result is not confirmed by the analysis in this chapter, it is found that wearing a seatbelt significantly reduces the probability of being involved in a fatal accident.

4.3 Data

The analysis in this chapter utilises an accident-level dataset for road traffic accidents in Norfolk and Suffolk in the years 2005 to 2014. The datasets are extracted from Accident Report Forms and contain information recorded at the time of an accident. These data are converted to driver-level for some parts of the analysis in order to identify driver characteristics. For every accident which is attended by the police, an Accident Report Form is completed which gives characteristics of the driver(s) involved and a detailed description of the environmental factors such as the type of road and weather conditions, these are outlined in Appendix C.1.

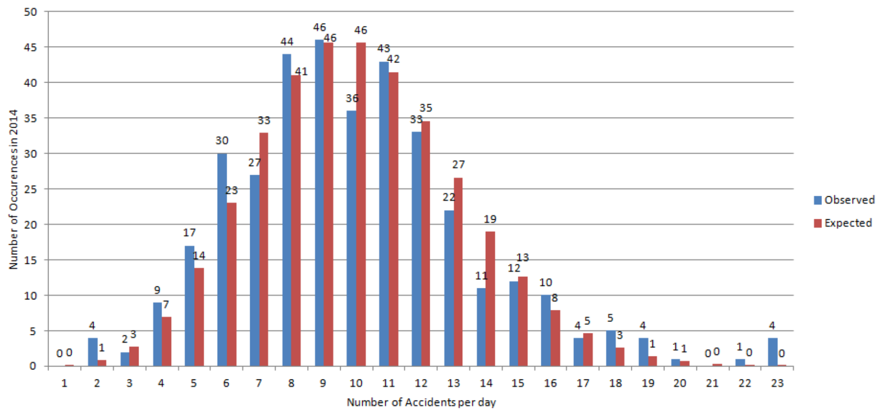
Driver characteristics included in the dataset are gender, age, ethnicity, breath test result, whether they were wearing a seatbelt and whether or not they hold a UK driving licence. In addition to these are several variables describing the environment and characteristics of the accident itself including the condition of the road, visibility, number of casualties, time of day, day of the week, road class and type, speed limit and weather conditions. The outcome variable for this analysis is the severity of an accident, which is recorded categorically as 1, 2 or 3 for the categories slight, severe and fatal respectively.

In order to find whether the number of accidents has changed over time, it is not enough to say that the number of accidents is lower in one year than the previous year, it is necessary to identify significant differences so as not to confuse natural variation in the data with changes in accident frequency.

In a report by the Department for Transport [Department for Transport, 2014], it is shown that the occurrence of fatal accidents follows a Poisson distribution by calculating the expected number of fatal accidents per day and comparing this with the observed number of accidents. A similar result can be shown for all road accidents in Norfolk and Suffolk by calculating the expected number of occurrences under the Poisson distribution and testing the statistical significance of the deviations, the distribution for which is illustrated in Figure 10.[8]

[8] See Table of χ^2 significance tests in Appendix C.2

Figure 10: Comparison of Simulated Poisson distribution with Observed Accidents

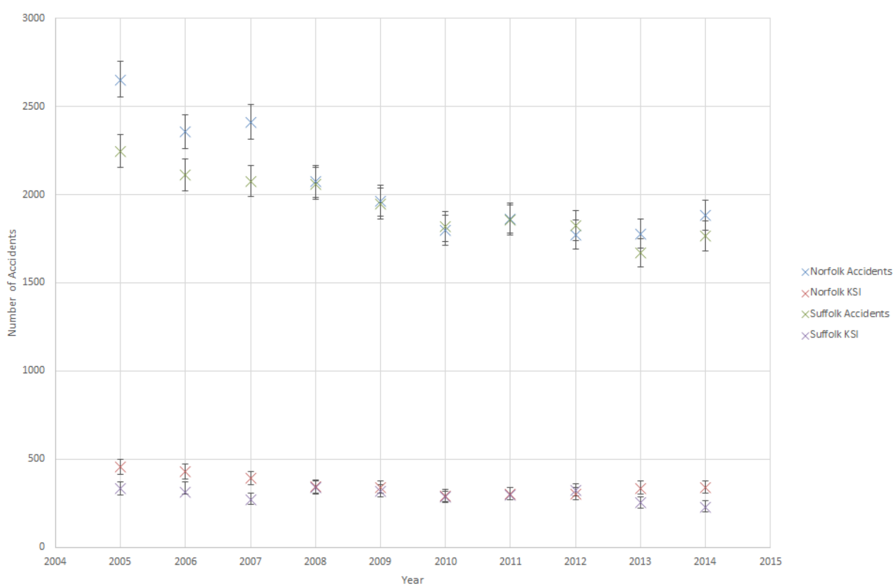


The Poisson Distribution is a discrete probability distribution, for a fixed time period it gives the probability that an event will occur a given number of times. In this case, it gives the probability that a certain number of accidents will occur during one day. An assumption is made that the occurrence of one accident does not affect the probability of a future accident occurring. The Poisson confidence interval is given in Equation 32 where k is the observed number of accidents and α is the significance level.

$$\begin{aligned} \frac{1}{2}\chi^2\left(\frac{\alpha}{2};2k\right) &\leq \mu \\ &\leq \frac{1}{2}\chi^2\left(\frac{1-\alpha}{2};2k+2\right) \end{aligned} \tag{32}$$

Given the number of accidents that actually occurred during a particular year, month or day, 95% confidence intervals can be calculated. We can be 95% sure that the number of accidents would fall within this range for this particular time period. It follows then that if these confidence intervals overlap for two time periods, we cannot say for certain that there is a statistically significant difference between the two frequencies. The frequency and Poisson confidence intervals for Norfolk and Suffolk are shown in Figure 11 and it is clear that there are very few statistically significant changes from one year to the next for accidents or KSI accidents.

Figure 11: Road Traffic Accidents in Norfolk and Suffolk



4.4 Estimation Methodology

The severity level of an accident is recorded by the police as one of three categories: slight, severe or fatal and this severity level is coded accordingly in the dataset. As such, the outcome variable for this analysis is categorical rather than continuous. Further than this, the numerical value assigned to each severity level is used to maintain order rather than magnitude, a severe accident with a value of 2 is not twice as bad as a slight accident with a value of 1. Therefore, the outcome variable is an ordinal categorical variable and must be treated differently to a continuous variable.

The OLS model is inappropriate for this type of variable since it requires that the error terms are normal with constant variance. This is not the case for ordered categorical variables as they are non-continuous, bounded and cannot be measured on an interval or ratio scale, leading to biased estimates as shown by Winship, C. & Mare [1984] McKelvey and Zavoina [1975]. It is, therefore, necessary to use a group of models specifically designed for this type of dependent variable, known as Ordered Logistic Regression Models (OLOGIT) or Proportional Odds Models. For this type of model, the dependent variable has M discrete levels and $M - 1$ binary logistic regressions are estimated using grouped values of the dependent variables. Table 10 illustrates an example where the dependent variable has three levels and the model estimates two logistic regressions. For the first regression, the dependent variable is equal to 0 when $Y = 1$ and one when $Y = 2$ *or* 3, for the second it is equal to 0 when $Y = 1$ *or* 2 and one when $Y = 3$. Accordingly, the estimated coefficients are the effect of a change in the confounding factors on the odds that $Y = 2\, or\, 3$ in the first regression and $Y = 3$ in the second.When using this type of model, it is necessary to first estimate an OLOGIT model and test whether heteroskedasticity in the errors is causing the proportional odds assumption to be violated. If heteroskedasticity is present, the next step is to estimate a Heterogeneous Choice Model (HCM). However, if the proportional odds assumption is still violated when the HCM is estimated, the Generalised Ordered Logistic Regression model (GOLOGIT) should be used to account for this.

Table 10: Ordered Logistic Regression Example

Logistic Regression	Dep. var = 0 if...	Dep. var = 1 if...
1	Y = 1	Y = 2 or 3
2	Y = 1 or 2	Y = 3

Ordered Logistic Regression Model

Equation 33 gives the OLOGIT model as described by Williams [2016], where $j = 1, ..., M$ is again the number of levels for the dependent variable and X_i is a vector of observed explanatory variables. As described in Table 10, a set of logistic regressions is used to find the effect on the odds that the dependent variable has a value higher or lower than a particular cut off. To estimate an OLOGIT model, the proportional odds assumption must be met, which requires that the estimated log odds are the same regardless of the cut off. As can be seen in Equation 33, α and β do not have a j subscript, since only one set of coefficients is estimated.

$$P(Y_i > j) = \frac{exp\,(\alpha + X_i\beta)}{1 + [exp\,(\alpha + X_i\beta)]},\ j = 1, 2, ..., M-1 \tag{33}$$

In practice, this assumption is very strong and violation can be tested for in two ways. The Brant test uses the idea that, if the proportional odds assumption holds, the estimated coefficients should not be significantly different from one logistic regression to the next. It calculates a χ^2 statistic for each explanatory variable and one for all, showing which, if any, violate the assumption. If the overall test is insignificant, OLOGIT is the appropriate model. If, however, the overall test is significant, the individual tests can be used to identify which explanatory variables do not meet the assumption, that is their estimated coefficients differ across regressions. The second way to test for violation is the Likelihood Ratio Test, which is similar to the Brant test but tests the assumption by estimating the model given by Equation 33 and the generalised model given by Equation 34 and comparing the goodness of fit for each. Again, an insignificant test statistic implies that the proportional odds assumption is not violated.

Estimated coefficients for factor variables are interpreted as the change in log odds when the value of the variables deviates from the baseline. For example, the baseline for Ethnicity is “White”, so the estimated coefficient for “Asian” represents how much more or less likely an Asian driver is to have a serious accident than a White driver.

Generalised Ordered Logistic Regression Model

Equation 34 gives the model for the Generalised Ordered Logit Regression (GOLOGIT), of which OLOGIT is a special case. As can be seen in the model, the proportional odds assumption is relaxed and a separate set of coefficients is estimated for each logistic regression. This model is suitable when the above tests indicate that the assumption for OLOGIT is not met.

$$P(Y_i > j) = \frac{exp\,(\alpha_j + X_i\beta_j)}{1 + [exp\,(\alpha_j + X_i\beta_j)]}, j = 1, 2, ..., M - 1 \tag{34}$$

The disadvantage of this, however, is that there are now $M - 1$ sets of coefficients to interpret and work with, rather than just one. The most commonly used model is a mixture of the two discussed called the Partially Constrained Generalised Ordered Logit Regression (PC-GOLOGIT) [Williams, 2006, 2016, Long, 1997]. Equation 35 gives an example of the model given by Williams [2016], where the estimated coefficients β_1 and β_2 are constrained and β_3 is unconstrained. This model is often the most appropriate in practice as it can accommodate both OLOGIT and GOLOGIT models where needed.

$$P(Y_i > j) = \frac{exp\,(\alpha_j + X1_i\beta 1 + X2_i\beta 2 + X3_i\beta 3_j)}{1 + [exp\,(\alpha_j + X1_i\beta 1 + X2_i\beta 2 + X3_i\beta 3_j)]}, j = 1, 2, ..., M - 1 \quad (35)$$

Heterogeneous Choice Model

The Heterogeneous Choice Model (HCM) is explained in Williams [2006], Williams [2010] and Williams [2016] and allows for heteroskedasticity. If an OLOGIT model is estimated and the error variances are assumed to be constant, the standard errors will be wrong and the estimated parameters will be biased [Williams, 2010]. Williams [2010] explains that an ordered categorical dependent variable in an OLOGIT regression is treated as a collapsed version of an underlying continuous variable, which means that the estimated coefficients are approximations of the true coefficients and have a relationship with the error variance. When this variance is constant, we have homoskedasticity and proportional odds, as the coefficients are the same for pairs of outcomes. However, when these error variances are not constant, there is heteroskedasticity and the proportional odds assumption may no longer be met.

One way of dealing with this problem is to estimate an HCM, using the variables highlighted by the LR and Brant tests. By specifying which variables are likely to be causing heteroskedasticity, the estimated model allows the error variance for these variables to vary.

4.5 Results

Ordered Logistic Regression

The first estimated regression model is the OLOGIT model, the results for which are shown in Table 55 in Appendix C.3. Factors estimated to increase the odds of having a more serious accident at the 5% significance level are a positive breath test, not wearing a seatbelt, mixed ethnicity, driving in the dark, increased number of casualties, driving on a road with a higher speed limit, driving on a weekday, increased number of OAPS and increased number of pedestrians. Factors estimated to reduce the odds of having a more serious accident are being female, not having a driving licence, frosty or icy conditions, being on a slip road and raining with high winds.

However, for these estimated coefficients to carry any weight, the proportional odds assumption must not be violated. Table 56 in Appendix C.3 shows the Likelihood ratio test, which indicates that the assumption is violated since the χ^2 statistic is statistically significant at 34.71 with a p-value equal to 0.015. Similarly, the Brant test in Table 57 has a significant χ^2 test statistic of 30.77 with the p-value 0.014, also indicating that at least one variable in the OLOGIT model violates the assumption.

These results imply that the effect of a change in the independent variable is not the same for an increase from slight to severe as it is for severe to fatal, all pairs of groups do not have the same relationship. The Brant test can be used to identify which independent variables may be violating this assumption, as these too will have significant estimates. These variables can be specified in an estimation model which allows heteroskedasticity, the Heterogeneous Choice Model. Table 57 shows that the variables with a significant test statistic are the number of vehicles, gender, breath test result and visibility.

Heterogeneous Choice Model

The regression output for the HCM model is shown in Table 58 in Appendix C.3[9]. Driver age and number of casualties involved in the accident have significant Lnsigma estimates, which indicates that these independent variables may be causing heteroskedasticity in the OLOGIT model. This result differs from the results found by Quddus et al. [2010], who conclude that the number of vehicles and

[9]The variable "Road Class" has been dropped for both the HCM and GOLOGIT models as it prevented them from running

the number of casualties are likely to be causing heteroskedasticity in the model. However, although this estimation model controls for heteroskedasticity, it does not control for other possible causes of violation of the proportional odds assumption. In order to account for this violation, it is necessary to use a GOLOGIT model.

Generalised Ordered Logistic Regression Model

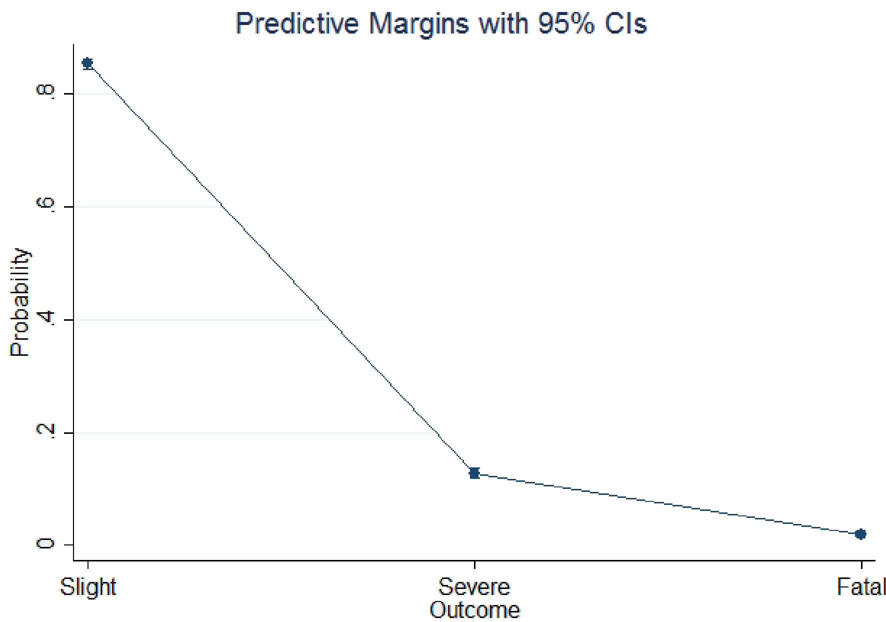
The OLOGIT model is a special case of the GOLOGIT model when every independent variable is constricted to meeting the proportional odds assumption, therefore the results of the constrained GOLOGIT model in Table 59 in Appendix C.3 are almost identical to the OLOGIT results in Table 55. As shown by the LR and Brant tests, this model is too restrictive and does not allow for different estimated coefficients at different levels. To allow for this, the proportional odds assumption is relaxed in an unconstrained model and the estimated coefficients are allowed to differ between levels for the same variable, the results for which are shown in Table 60. The ideal model, however, is a mixture of the two where estimates are constrained when the difference between them is not statistically significant and the results for this partially constrained model are presented in Table 61 in Appendix C.3. An LR test now gives an insignificant test statistic indicating that the final model does not violate the parallel lines assumption.

The independent variables for which the estimated coefficient is statistically significant at the 5% level are the number of gender, a positive breath test, wearing a seatbelt, not having a driving licence, mixed ethnicity[10], frost or ice on the road, dark with no lighting, a higher number of casualties, having an accident on a slip road, increased speed limit, rain with high winds and having an accident on a weekday, higher number of OAPs and higher number of pedestrians.

The estimated coefficients are in the form of log odds, which can be difficult to interpret . Table 64 in Appendix C.3 presents the marginal effects which give the probability of a particular outcome occurring for each possible value of a variable. Figure 12 shows that for any individual who has a road accident, there is an 85% chance that it will be slight, 13% chance that it will be severe and 2% chance that it will be fatal.

[10]It is possible that the small sample size for this group has resulted in a significant estimate rather than a causal relationship as it seems unlikely that mixed ethnicity would affect the severity of an accident.

Figure 12: Marginal Effects



The robustness of the model is tested by estimating the partially constrained GOLOGIT model for the years 2005-2009 and 2010-2014. The main drawback of the GOLOGIT model is that it can estimate unusually high coefficients and negative p-values when a category has very few observations. This problem has occurred in the 2005 - 2009 model and has affected the estimates for several variables, for example, the second estimated coefficient for Ethnicity = Black is 5,870,000,000. However, the majority of estimates appear to be robust to the time sample. The estimated coefficients are also similar to those estimated in previous models, as shown in Table 62 in Appendix C.3.

While the analysis in this chapter identifies factors which have an effect in isolation, the analysis in the next chapter uses classification and regression tree analysis (CART) to identify groups of factors which interact to have a significant effect. Due to the large number of identified groups, it is not possible to include all of them as interactions in the PC-Gologit model, however Table 65 illustrates the estimated coefficients when two of the groups are included. These are “OAPS & casualties” and “OAPs, casualties and pedestrians”. The addition of these interactions does not affect the sign or significance of the other estimations. The estimated coefficient for the first group is insignificant, whereas the estimated coefficient for the second group is negative and significant. The proportional odds assumption does not hold for this group as the estimated coefficient for the first model is -4.778 and for the second model is -7.698.

Chi Squared Goodness of Fit Test

The Chi-squared Goodness of Fit test is used to test whether the observed proportion of accidents involving a particular age group is significantly different from the proportion expected. However, simply comparing the driving age distribution in Norfolk and Suffolk with the age distribution of drivers involved in accident may give misleading results. Drivers in different age groups also tend to drive different average distances over a year and it is reasonable to expect that the chances of having an accident are higher when a driver spends more time driving.

It is necessary, therefore, to weight the expected proportion of accidents by the average distance driven by an age group in that year. The figures for England for the years 2005 - 2013 are given in Table 11[11].

Table 11: Average Annual Distances in England

Average miles/ person/year	17-20	21-29	30-39	40-49	50-59	60-69	70+
2005	1,801	3,805	5,745	6,307	5,871	4,114	1,861
2006	1,595	3,943	5,943	6,235	5,891	4,110	1,741
2007	1,828	4,235	5,579	6,428	5,795	4,115	1,702
2008	1,820	3,563	5,373	6,208	5,611	3,846	1,776
2009	1,719	3,608	4,874	5,593	5,295	3,957	1,783
2010	1,399	3,338	5,237	5,992	5,662	4,004	1,765
2011	1,338	3,375	5,045	5,987	5,585	4,158	1,872
2012	1,508	2,968	4,837	5,795	5,521	4,314	1,957
2013	1,249	3,274	4,643	5,659	5,321	4,116	1,905

Assuming that these distances are a good approximation for the average distances driven in Norfolk and Suffolk, the driving population is used to calculate the total number of miles driven per year for each age group, and this, as a proportion of the total number of miles driven in Norfolk and Suffolk, is the proportion of accidents which would be expected to involve a driver from that age group.

Figure 22a shows that drivers aged 17-20 have consistently been

[11] These distances are calculated using the National Travel Survey dataset which can be found at https://www.gov.uk/government/collections/national-travel-survey-statistics.

involved in far more accidents than expected given the average distance driven by this age group. These large differences are shown to be statistically significant in Table 67, as the χ^2 statistic is extremely large and the p-value is equal to zero for every year. Figure 22b shows the expected and observed accident percentages for male and female drivers aged 17-20. Observed accidents are, again, shown to be significantly higher each year than expected, but it is also clear that female drivers in this age group are expected to and observed to have fewer accidents than males. This is due to the fact that female drivers in this age group drive for a shorter distance, on average, during each year than male drivers in the same age group.

Figure 23a shows that drivers aged 21-29 have also consistently exceeded the expected proportion of accidents during the years 2005 - 2013 and the difference is significant. Figure 23b shows that, again, observed accidents are shown to be higher each year than expected, but it is also clear that female drivers in this age group are expected to and observed to have fewer accidents than males. This is due to the fact that female drivers in this age group drive for a shorter distance, on average, during each year than male drivers in the same age group. Table 68 shows the χ^2 statistics are statistically significant for all years.

Figure 24a shows that drivers aged 30-39 are also involved in a larger proportion of road accidents than expected, but by less than the younger age groups. The differences shown in Figure 24b are larger, now, for male drivers than for females in this age group. The χ^2 tests in Table 69 show that the difference between observed and expected accident proportions are not statistically significant in 2005, 2006 or 2008 for all drivers. The difference for male drivers is not statistically significant in 2006 and for female drivers, the difference is insignificant in 2006 and 2008.

Figure 25a shows that drivers aged 40-49 are involved in fewer accidents than expected, given the average annual distance driven in this age group. Again, female driver are expected to and observed to be involved in fewer accidents than males, this is shown in Figure 25b. Despite being relatively small, Table 70 shows that all differences between expected and observed proportions are statistically significant for this age group.

Drivers aged 50-59 are shown in Figure 26a to have far fewer accidents than expected, this trend is consistent throughout the time period. Figure 26b shows that these differences are larger for male drivers than for females, but all differences are, again, shown to be statistically significant in Table 71.

Figure 27a shows that drivers aged 60-69 have fewer accidents than expected, but it is clear from Figure 27b that this difference is driven by male drivers in this age group. Female drivers show very little difference between expected and observed accidents and these differences are shown to be insignificant for the years 2005 - 2011. The difference for females is, however, significant for the years 2012 and 2013, mainly due to the factor that the number of female drivers in this age group increased during these years and, therefore, the number of expected accidents has also increased. All other differences in Table 72 are significant.

Figure 28a misleadingly implies that drivers over 70 years old consistently have more than the expected number of accidents each year. However, Figure 28b shows that this is solely driven by the number of accidents involving female drivers in this age group. The differences for male drivers are shown in Table 73 to be insignificant for most years, but significant for female drivers.

4.6 Conclusions

This chapter uses econometric analysis to study the severity of traffic accidents in Norfolk and Suffolk during the years 2005 - 2014. OLOGIT models are used to estimate the effect of independent variables on the severity of an accident, as severity is an ordered, categorical variable. Using a Likelihood Ratio test and a Brant test, it is shown that the baseline model violates the proportional odds assumption and contains variables which may be causing heteroskedasticity. These problems can be tackled by using a Heterogeneous

Choice Model or a Partially-Constrained Generalised Ordered Logistic Model, which estimate similar coefficients. These results show that several independent variables have a statistically significant effect on the severity of accidents, the largest of which are a positive breath test, not wearing a seatbelt and crashing on a slip road. Three variables are found to have different effects depending on the level of severity: number of vehicles, driving in the dark with lights lit and driving in the dark with no lights. Log odds and marginal effects are reported in order to facilitate interpretation of the magnitude of these estimates.

Two example interaction terms are introduced to the estimated model in Table 65 to illustrate the relationship between this regression model and the CART analysis presented in the next chapter. The first of these two groups involves the factors "OAPs" and "casualties". The results from the CART analysis, illustrated on page 118, show that for the groups of drivers involved in accidents in which the number of OAPs is fewer than or equal to 1 and the number of casualties is more than 5, 37 out of 54 were severely injured as opposed to slightly or fatally. When the interaction between OAPs and casualties is included in the regression in Table 65, the estimated coefficient is insignificant. This is not surprising, since the interaction term is showing the combined effect of a higher number of OAPs and a higher number of casualties, which is not what is being shown in the CART analysis. For the second group, the results from CART show that for groups of drivers involved in accidents in which the number of OAPs is fewer than or equal to 1, the number of casualties is fewer than or equal to 5 and the number of pedestrians is greater than 0, 32 out of 47 are severely injured rather than slightly or fatally. It is also important to remember that not all of the interactions found by CART are included in the regression, which may affect the estimated coefficients. These results highlight the difference between these two methods of analysis and this is discussed further in the next chapter.

In addition to regression analysis, a Chi-Squared Goodness of fit test is carried out for the age group and gender of drivers involved in traffic accidents. By using the average distances driven by certain groups, it is possible to estimate the proportion of accidents that are expected to involve certain drivers and compare this with the observed proportions in order to identify those groups that are most at risk of being involved in a crash. Simply looking at an age group as a whole can be misleading, particularly in the case of drivers aged 70+, who appear at first to exceed their expected accident proportion every year in the analysis period. However, when split by gender, it becomes clear that male drivers in this age group actually have fewer accidents than expected, while female drivers are involved in more than double the number of accidents expected given their average distance driven. The groups identified as most at risk are all drivers aged 17 to 39 and female drivers aged 70+.

All other groups are involved in significantly fewer accidents given their average distances.

The results found in this chapter indicate that there is a strong need for targetted policies to discourage drivers from acting in such a way that increases the probability of being killed or seriously injured if they are involved in an accident. The second part of the analysis identifies specific demographic groups, namely those ages 17 to 39 and females over the age of 70, who are involved in the highest number of accidents per mile driven. These groups could also be targetted by campaigns to increase their awareness of road safety in the hope that this reduces the number of road traffic accidents.

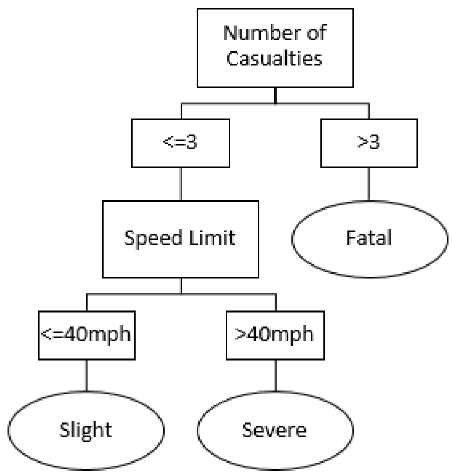
5 IDENTIFYING GROUPS AT RISK OF KSI ACCIDENTS USING CLASSIFICATION AND REGRESSION TREE ANALYSIS

5.1 Introduction

The aim of the research in this chapter is to identify driver characteristics and environmental factors which affect the severity of a road traffic accident. Several combinations of variables are found to make severe and fatal accidents more likely and these are represented as leaves on the final decision tree. This type of model would be beneficial to the police as they can identify groups of drivers and circumstances which are most at risk of KSI accidents. Accident prevention policy can then be targetted and may be more cost effective than blanket policies.

Classification and regression tree (CART) analysis has been adopted by other disciplines such as medicine and psychology, but is not explored much in economics. This type of analysis lends itself well to datasets with several categorical variables and is also good at handling interactions in addition to heterogeneity. CART are, therefore, well suited to economic analysis since these are often traits of real-world datasets. The output of CART is an intuitive decision tree which can be easily interpreted by both academics and non-academics, an example of which is shown in Figure 13. The analysis contributes to the economics literature by employing a novel technique, which is not used in many papers, to identify groups who are most are at risk of KSI accidents.

Figure 13: Example Classification Tree



5.2 Literature Review

The use of CART relies heavily on the algorithm with which the model is built and the analysis in this chapter uses the C5.0 algorithm developed by Quinlan [1993], which is discussed in detail in subsection 5.4. This method of analysis, however, seems to be missing from the economics literature and has not been used in any similar analyses.

Due to the versatility and user-friendly nature of this method of analysis, it is used in the medical literature, particularly when designing predictive models of mortality rates. Austin [2007] uses a dataset for hospital patients in Ontario and compares the results of CART analysis with a traditional logistic regression. He finds, however, that the predictions of the logistic regression are more accurate than the classification tree.

In Psychology, Tonkin et al. [2012] compare the use of CART with logistic regressions when building predictive models for case linkage. They use two different samples, one for burglaries in Finland and one for car thefts in the UK, and run a binary logistic regression and CART for each. Like Austin [2007], the authors also find that the tree is less robust than the logistic regression when discriminating between outcome types for the burglary dataset, but the predictive accuracy of the tree model performs well when compared to logistic regression for car theft data.

5.3 Data

This chapter uses a driver-level dataset compiled by Norfolk and Suffolk police from road traffic accidents during the years 2005 to 2014. Police reports are used to record the driver characteristics and environmental factors surrounding an accident and the details of these variables are described in Appendix C.1. The dataset includes driver characteristics such as gender, age, ethnicity, breath test result, whether they were wearing a seatbelt and whether or not they hold a UK driving licence. Additionally, several variables are included to describe the environment and characteristics of the accident itself including the condition of the road, visibility, number of casualties, time of day, day of the week, road class and type, speed limit and weather conditions. The outcome variable for this analysis is the severity of an accident, which is recorded categorically as 1, 2 or 3 for the categories slight, severe and fatal respectively.

Figures 14 to 17 show the location of KSI accidents in Norfolk and Suffolk for these years. These maps have been produced using Geographical Systems Information (GIS) software.

Figure 14: Fatal Accidents in Norfolk: 2005 - 2014 (Accident data for Norfolk and Suffolk)

Figure 15: Serious Accidents in Norfolk: 2005 - 2014 (Accident data for Norfolk and Suffolk)

Figure 16: Fatal Accidents in Suffolk: 2005 - 2014 (Accident data for Norfolk and Suffolk)

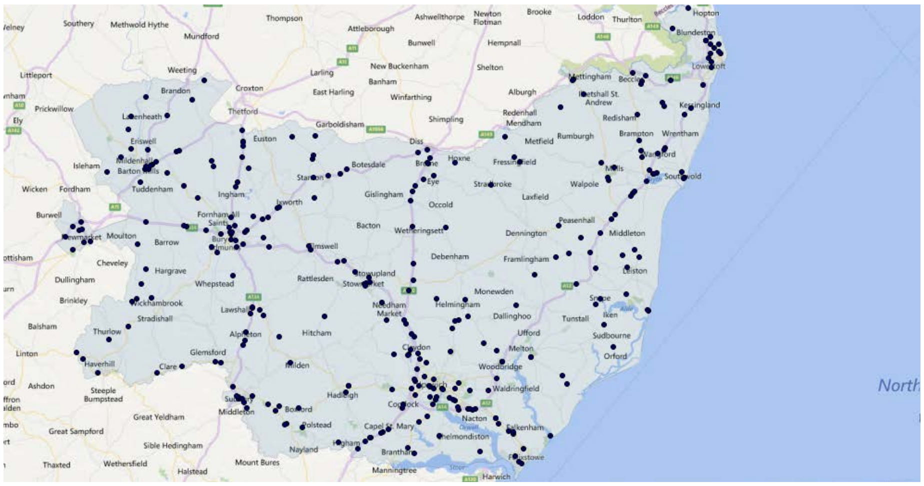
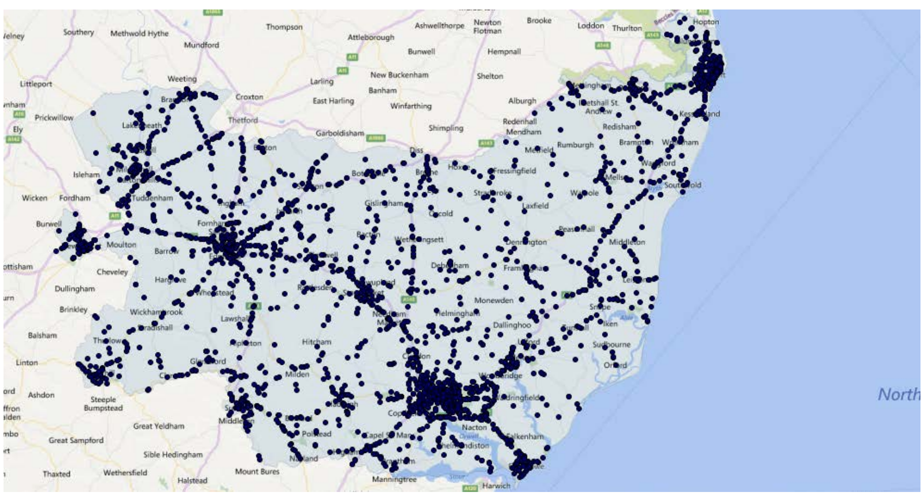


Figure 17: Serious Accidents in Suffolk: 2005 - 2014 (Accident data for Norfolk and Suffolk)



5.4 Estimation Methodology

Classification and regression tree modelling (CART) is a method of analysis used in data-mining which seeks to predict the outcome of future events based on the characteristics of past events. There are several algorithm options, all of which create splits in the group of observations and create a tree. There are several advantages to using this type of model, particularly when analysing categorical variables, as is often the case when the data are derived from forms or surveys. The repeated splitting of observations into groups is most natural for categorical variables which can be split into clearly defined categories. An additional benefit of CART over traditional regressions is that they work well with interactions, for example a fatal accident may be more likely if a driver is both a certain age and driving a certain speed. These interactions can be difficult to include in regressions since there are so many possibilities to consider. In a CART model, however, the nature of the tree means that interactions are automatically included if they are optimal without having to specify them.

The classification tree estimated in this chapter uses the C5.0 algorithm, a descendant of C4.5 which was developed by Quinlan [1993]. Like many decision tree algorithms, it is based on information entropy and each node contains a test which splits the data in such a way that the normalised information gain is maximised. As explained by Salzberg [1994], in his review of Quinlan's book, the optimal test at any node in the tree would be one which splits the observations so that each subset contains only one class. However, since this almost never occurs in practice, the role of the splitting algorithm is to come as close to this as possible.

The C5.0 algorithm is explained by Quinlan [1993] and starts with a training dataset, a subset of the original dataset on which the model is built in order to predict the outcomes of the remaining observations. Any random subset of this training set is denoted S and the class of an observation is C_j. Information theory states that the information conveyed by a message is measured in bits, as in Equation 36. If $|S|$ denotes the number of observations in subset S and $freq(C_j, S)$ denotes the number of observations in subset S which belong to class C_j, then the information conveyed by any random observation is given by Equation 37.

$$-log_2(probability)\ bits \tag{36}$$

$$-log_2 \frac{freq(C_j, S)}{|S|}\ bits \tag{37}$$

The aim of the splitting algorithm is to maximise the information gain in the subset, S, with respect to the class, C_j, to which observations belong. This is also known as the entropy of the subset and is calculated using Equation 38, weighting the information conveyed by an observation by the probability that it belongs to class C_j.

Therefore, if the training set, T, is split into subsets using a test, X, the expected information requirement can be found as a weighted sum over all subsets, which is given by Equation 39. Equation 40 illustrates the overall information gain from splitting T into subsets using test X.

$$info(S) = \sum_{j=1}^{k} \frac{freq(C_j, S)}{|S|} \times -log_2 \frac{freq(C_j, S)}{|S|} \; bits \tag{38}$$

$$info_X(T) = \sum_{i=1}^{n} \frac{|T_i|}{T} \times info(T) \tag{39}$$

$$gain(X) = info(T) - info_X(T) \tag{40}$$

The role of the C5.0 algorithm is to calculate this information gain for all possible tests, X, and choose the one by which it is maximised. Once a test has been chosen, the algorithm will repeat this process at each node until the subset contains observations from a single class or a single observation. Since this may result in a very large tree, various rules are often put in place to halt the algorithm at an earlier stage.

For this analysis, the accident-level dataset was converted to driver-level in order to analyse the effect of driver characteristics in addition to factors surrounding the accident on the severity of an accident. Although classification trees are used to predict the outcome of future events, the primary purpose of the analysis in this chapter is to identify groups and situations where the risk of more severe road traffic accidents is higher and this is done by building a model which has predictive power for a test dataset.

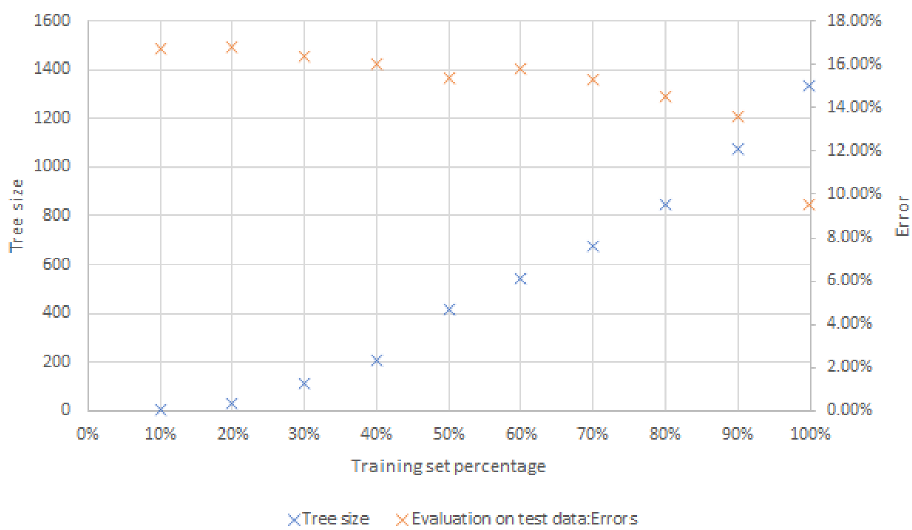
The driver-level dataset for Norfolk and Suffolk contains 76, 334 records, for which the accident severity, driver characteristics and other confounding factors are recorded. In order to avoid selection bias, half of the dataset is selected at random to create the training set on which to build the model and the other half is reserved for testing when the model is complete. The training set, therefore, contains 38,167 observations which belong to one of three severity classes: slight, severe or fatal.

As discussed by Quinlan [1993], the test which maximises the information gain is chosen at each stage in order to create the next set of branches. A test for a continuous variable often comes in the form of an inequality and for a categorical variable, branches represent one or more classes within that category.

5.5 Results

One of the advantages of the software is the choice of settings and adjustments which can be made when designing the tree. The first decision to make is the proportion of the data which should be used as the training set and consequently how large the test set should be, which is the set of observations on which the model will be tested. A percentage error will be calculated to show the proportion of observations for which the outcome variable class in incorrectly predicted. As shown in Figure 18, the percentage of observations used in the training set affects both the size of the tree produced and the percentage of incorrectly predicted observation classes. As the size of the training set increases, so too does the size of the tree and the percentage error falls. There is, therefore, a trade-off between producing a small tree which is easy to interpret and over-fitting the model to the dataset in order to achieve a low percentage error. The majority of papers who use this analysis method settle on using 50% of the dataset in order to strike a balance between these two objectives and this is the training set size which is used in this chapter.

Figure 18: The Effect of the Size of the Training Set



Quinlan [1993] explains that repetitive splitting of observations in a non-trivial way will always result in single-class leaves eventually, but the aim of designing a model is to produce a tree which is small enough to interpret with the minimum percentage error possible. With this in mind, Table 12 illustrates the tree size and the percentage error when the minimum number of cases per leaf is restricted. This is a form of "pre-pruning", where a decision is made about the size of the tree before it is run. By increasing the number of minimum cases to eight, the size of the tree becomes manageable without much of an effect on the percentage error.

Table 12: Minimum Cases

Minimum Cases	Tree Size	Percentage Error
1	273	16.1%
2	210	16.1%
3	198	16.2%
4	108	16.1%
5	95	15.9%
6	57	16.4%
7	46	16.5%
8	27	16.3%

The result of this process is the classification tree on page 119 which, when tested on the test dataset, correctly predicts the severity of accidents for 83.7% of observations. At each node is a variable, either categorical or continuous, and it is followed by a set of branches, each with a result of the test which leads to the next node. For example, the first node is concerned with the number of old aged pensioners (OAPs) involved in the accident and splits into two branches depending whether there were fewer than or equal to one OAP involved or more than one. These tests continue along each branch until a leaf is reached, which is represented as an oval. Each leaf gives the class to which the majority of observations in the subset belong and the ratio to other classes. The interpretation for the first branch is as follows: in an accident where 0 or 1 OAPs are involved and there are more than 5 casualties, 37 out of 54 drivers were involved in a severe accident and therefore any drivers who meet these conditions in the future would also be predicted to have a severe accident rather than slight or fatal.

Table 13: Evaluation on Test Data

		Classified as		
		Fatal	Severe	Slight
	Fatal	19	20	772
Actual	Severe	6	120	5465
	Slight	13	61	32481

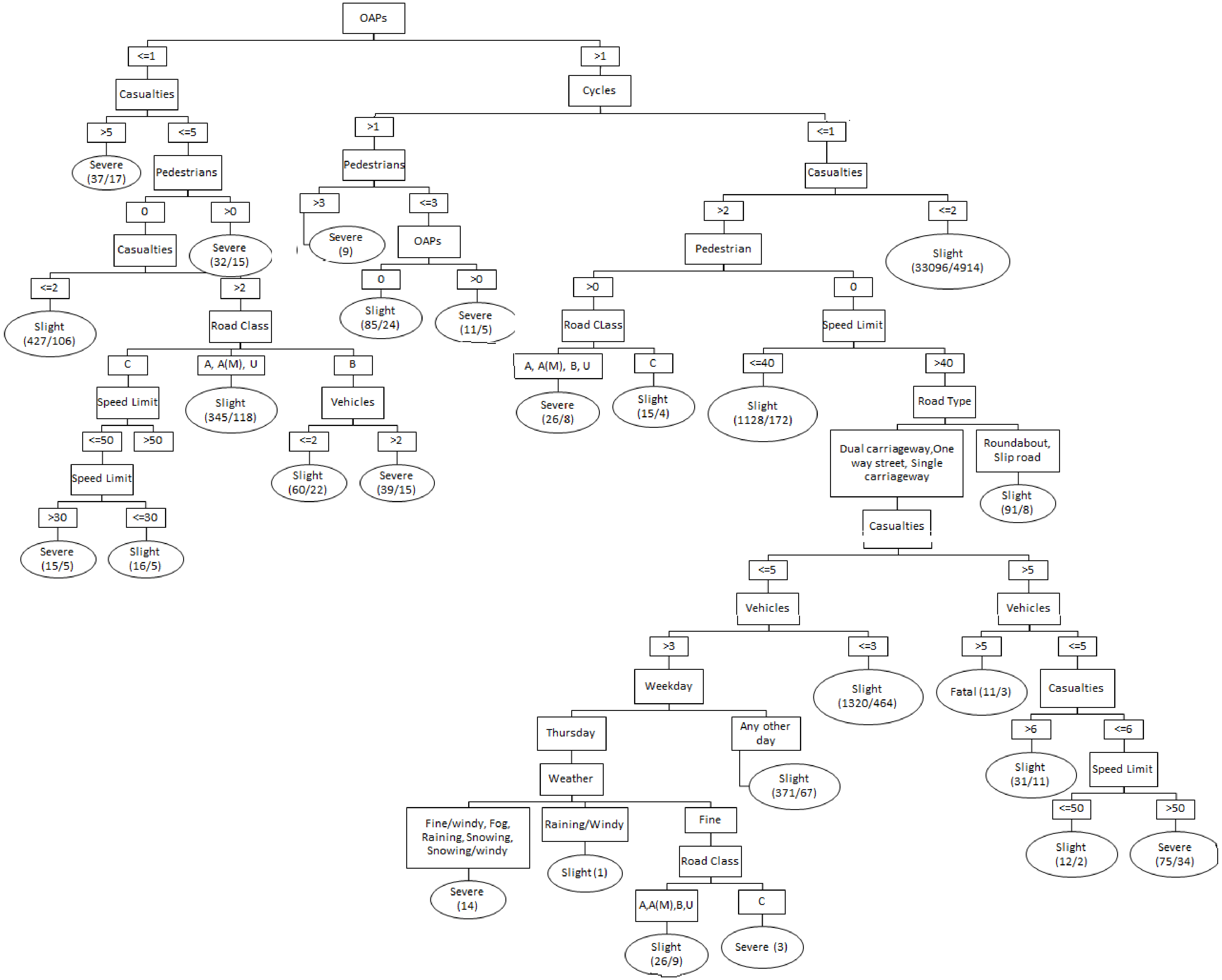
Errors: 16.3%

The accuracy of the decision tree is illustrated in Table 13 which shows that, when used to predict the outcome of the observations in the test dataset, 16.3% are incorrectly predicted. The observations along the diagonal show cases which are correctly predicted, for example there were 19 cases which were classified as fatal and were, in fact, fatal. Off the diagonal are the cases which are incorrectly predicted, for example there are 13 cases which were classified as fatal but were really slight. The error percentage is calculated as the percentage of total predictions which are incorrect, which is illustrated in Equation 41.

$$\left[\frac{(20+772+6+5465+13+61)}{(19+20+772+6+120+5465+13+61+32481)}\right] * 100 \tag{41}$$
$$= 16.3\%$$

As mentioned above, the predictive power of this type of model is useful in some contexts, however the objective of this chapter is to identify those groups of characteristics and environmental factors which put drivers at most risk of being involved in a KSI accident. As illustrated by the classification tree, there is only one leaf contain-ing observations where the majority are fatal and these observations share the following characteristics: more than one OAP involved, one or no cycles involved, more than two casualties, zero pedestrians, speed limit over 40mph, either a dual carriageway, one way street or a single carriageway, more than five casualties and more than five vehicles.

There are several important points to keep in mind when interpreting the results of a classification tree. The first is that this is not the only combination of factors which will result in a fatal accident, there may be drivers in other leaves who fall into the same class. The classification merely shows that the majority of this group were involved in a fatal accident and also share these characteristics. Another point to keep in mind is that the results do refer to combinations of factors, which is different to the marginal effects of individual variables found in regression analysis. For example, a speed limit which is over 40mph only increases the probability of an accident being fatal when combined with the other relevant factors.


OAPs
<=1
>1
Casualties
Cycles
>5
<=5
Severe (37/17)
Pedestrians
0
>0
Casualties
Severe (32/15)
<=2
>2
Slight (427/106)
Road Class
C
A, A(M), U
B
Speed Limit
Slight (345/118)
Vehicles
<=50
>50
<=2
>2
Speed Limit
Slight (60/22)
Severe (39/15)
>30
<=30
Severe (15/5)
Slight (16/5)
>1
Pedestrians
>3
<=3
Severe (9)
OAPs
0
>0
Slight (85/24)
Severe (11/5)
<=1
Casualties
>2
<=2
Pedestrian
Slight (33096/4914)
>0
0
Road CLass
Speed Limit
A, A(M), B, U
C
Severe (26/8)
Slight (15/4)
<=40
>40
Slight (1128/172)
Road Type
Dual carriageway,One way street, Single carriageway
Roundabout, Slip road
Slight (91/8)
Casualties
<=5
>5
Vehicles
Vehicles
>3
<=3
>5
<=5
Weekday
Slight (1320/464)
Fatal (11/3)
Casualties
Thursday
Any other day
>6
<=6
Weather
Slight (371/67)
Slight (31/11)
Speed Limit
Fine/windy, Fog, Raining, Snowing, Snowing/windy
Raining/Windy
Fine
<=50
>50
Severe (14)
Slight (1)
Road Class
Slight (12/2)
Severe (75/34)
A,A(M),B,U
C
Slight (26/9)
Severe (3)

5.6 Conclusions

The aim of this analysis is to identify combinations of factors which increase the risk that drivers will experience severe or fatal injuries when involved in an accident. This is a novel approach in the economics literature and has not been used to identify factors which affect the severity of road accidents.

Unlike the regression analysis in the previous chapter, CART identifies groups of factors which affect the severity of accidents rather than single variables in isolation. As discussed in the literature, this method is not necessarily superior to traditional regression analysis, rather it offers a different way of looking at the relationships by classifying them in a non-linear way.

Several combinations of factors are found to increase the risk of severe or fatal accidents, indicating that there are specific groups which could be the focus of policies aimed at reducing the severity of future accidents.

Expansion of this work could be carried out if similar datasets are made available for other police force areas in England and Wales. The model can be tested on them to find whether it works well when applied to other geographical areas. A more generalisable model could be developed if the training set included observations from different areas.

References

George A. Akerlof and Rachel E. Kranton. Economics and Identity. *The Quarterly Journal of Economics*, 115(3):715–753, 2000.

Peter C. Austin. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Statistics in Medicine*, 26:2937–2957, 2007.

Prof Mirko Bagaric and Brienna Bagaric. Mitigating the Crime that is the Over-Imprisonment of Women : Why Orange Should Not Be the New Black. *Deakin Law School Legal Studies Research Paper*, (16), 2016.

Ann P. Bartel. Women and Crime: An Economic Analysis. *Economic Inquiry*, 17(1):29 – 51, 1979.

BBC. Is it time for Scotland to stop jailing so many women?, 2016. URL `http://www.bbc.co.uk/news/uk-scotland-36502851`.

Gary Becker. Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(1968):169–217, 1968.

Kate M. Blincoe, Andrew P. Jones, Violet Sauerzapf, and Robin Haynes. Speeding drivers' attitudes and perceptions of speed cameras in rural England. *Accident Analysis and Prevention*, 38(2):371–378, 2006.

Alan S Blinder. Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4):436–455, 1973.

Holger Bonin, Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. Cross-sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes. *Labour Economics*, 14(6):926–937, 2007.

Kristin F Butcher and Robert J. Lalonde. Female Offenders Use of Social Welfare Programs Before and After Jail and Prison: Does Prison Cause Welfare Dependency? *Federal Reserve Bank of Chicago*, 2006.

Kristin F Butcher and Kyung H Park. Comparing Apples to Oranges: Differences in Women's and Men's Incarceration and Sentencing Outcomes. 2017.

Rosa Minhyo Cho. Impact of maternal imprisonment on children's probability of grade retention. *Journal of Urban Economics*, 65(1):11–23, 2009.

Keith B. Crew. Sex differences in criminal sentencing: Chivalry or patriarchy? *Justice Quarterly*, 8(1):59–83, 1991.

Department for Transport. Testing for statistically significant changes in road accident and casualty numbers. Technical Report April 2004, Department for Transport, 2014.

Department for Transport. Reported Road Casualties in Great Britain: Main Results 2015. Statistical Release. Technical Report 25 June, 2015.

Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, Jurgen Schupp, and Gert G. Wagner. Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. *IZA Discussion Paper*, (1730):1–56, 2005.

Edna Erez. Dangerous men, evil women: Gender and parole decisionmaking. *Justice Quarterly*, 9(1):105–126, 1992.

Helga Fehr-Duda, Manuele De Gennaro, and Renate Schubert. Gender, Financial Risk, and Probability Weights. *Theory and Decision*, 60:283–313, 2006.

Richard Freeman. The Economics of Crime. In *Handbook of Labour Economics*, volume 3, chapter 52, pages 3529–3563. Elsevier B.V., 1999.

L Fridstrom, J Ifver, S Ingebrigtsen, R Kulmala, and L K Thomsen. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident, analysis and prevention*, 27(1):1–20, 1995.

Evelina Gavrilova and Nadia Campaniello. Uncovering the Gender Participation Gap in the Crime Market. *IZA Discussion Paper*, (8982):1–37, 2015.

Loraine Gelsthorpe and Allison Morris. Women's imprisonment in England and Wales: A Penal Paradox. *Criminal Justice*, 2(3):277–301, 2002.

Cynthia W Goss, Lisa D Van Bramer, Jeffrey A Gliner, Todd R Porter, Ian G Roberts, and Carolyn Diguiseppi. Increased police patrols for preventing alcohol-impaired driving. *Cochrane database of systematic reviews*, (4), 2008.

E D Gould, B a Weinberg, and David Mustard. Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997. *Review of Economics and Statistics*, 84(1):45–61, 2002.

Jennifer Hart. Why Do Women Commit Less Crime? *New Society*, 73:298 – 300, 1985.

Akihiro Hashimoto. Measuring the effect of police surveillance on the prevention of traffic accidents. *Accident Analysis & Prevention*, 11(4):261–270, 1979.

W. M. Hirst, L. J. Mountain, and M. J. Maher. Are speed enforcement cameras more effective than other speed management measures?: An evaluation of the relationship between speed and accident reductions. *Accident Analysis and Prevention*, 37(4):731–741, 2005.

Ben Jann. The Blinder-Oaxaca Decomposition for Linear Regression. *Stata Journal*, 8(4):453–479, 2008.

Robert J. Lalonde and Rosa M. Cho. The impact of incarceration in state prison on the employment prospects of women. *Journal of Quantitative Criminology*, 24(3):243–265, 2008.

Haojie Li, Daniel J. Graham, and Arnab Majumdar. The impacts of speed cameras on road accidents: An application of propensity score matching methods. *Accident Analysis and Prevention*, 60(2013):148–157, 2013.

Caesar Lombrosso and William Ferraro. *The Female Offender*. D. Appleton, New York, 1895.

J Scott Long. Regression Models for Categorical and Limited Dependent Variables. *Advanced Quantitative Techniques in the Social Sciences*, pages 114–147, 1997.

S Machin and C Meghir. Crime and Economic Incentives. *Journal of Human Resources*, 34(4):958–979, 2004.

Kimmarie McGoldrick. Do Women Receive Compensating Wages for Earnings Uncertainty? *Southern Economic Journal*, 62(1):210–222, 1995.

Richard D. McKelvey and William Zavoina. A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *The Journal of Mathematical Sociology*, 4(1):103–120, 1975.

G W Mercer. The relationships among driving while impaired charges, police drinking-driving roadcheck activity, media coverage and alcohol-related casualty traffic accidents. *Accid Anal Prev*, 17(6):467–474, 1985.

Jacob A Mincer. Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4):281–302, 1958.

Jacob A Mincer. Schooling, Experience and Education. *NBER Working Paper Series*, I:41–63, 1974.

Ministry of Justice. Statistics on Women and the Criminal Justice System 2015. Technical Report November, 2016.

Linda Mountain, William Hurst, and Mike Maher. Costing lives or saving lives: a detailed evaluation of the impact of speed cameras. *Traffic Engineering & Control*, 45(8):280–287, 2004.

David B. Mustard. Racial, Ethnic, and Gender Disparities in Sentencing : Evidence From. *The Journal of Law & Economics*, 44(1):285–314, 2001.

Ronald Oaxaca. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709, 1973.

Maria De Paola, Vincenzo Scoppa, and Mariatiziana Falcone. The Deterrent Effects of Penalty Point System in Driving Licenses: A Regression Discontinuity Approach. 2010.

Kyung H Park. Do Judges Have Tastes for Racial Discrimination ? Evidence from Trial Judges. *The Review of Economics and Statistics*, Accepted f:1–34, 2014.

S Peltzman. The Effects of Autombile Safety Regulation. *The Journal of Political Economy*, 83(4):677–726, 1975.

Mari B. Pierce and Tina L. Freiburger. Assessing the Influence of Familial Paternalism on Child Neglect Sentencing Decisions. *American Journal of Criminal Justice*, 36(4):421–433, 2011.

Mohammed a. Quddus, Chao Wang, and Stephen G. Ison. Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models. *Journal of Transportation Engineering*, 136(5):424–435, 2010.

Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

L.S. Robertson. A Critical Analysis of Peltzman's" The Effects of Automobile Safety Regulation". *Journal of Economic Issues*, 9(3):587 – 600, 1977.

S Fernando Rodriguez, Theodore R Curry, Gang Lee, and Fernando Rodriguez. Gender Differences in Criminal Sentencing: Do Effects Vary Across Violent, Property, and Drug Offenses? *Social Science Quarterly*, 87(2), 2006.

Steven Salzberg. Book Review: C4.5: Programs for Machine Learning. *Machine Learning*, 1(16):235–240, 1994.

Renate Schubert, Martin Brown, Matrhias Gysler, and Hans Wolfgang Brachinger. Financial Desision-Making: Are Women Really More Risk-Averse? *American Economic Association*, 89(2):381–385, 1999.

Todd Andrew Sorensen, Surpriya Sarnikar, and Ronald L. Oaxaca. Do you receive a lighter prison sentence because you are a woman or a white? An economic analysis of the federal criminal sentencing guidelines. *B.E. Journal of Economic Analysis and Policy*, 14(1):1–54, 2014.

Sonja B. Starr. Estimating gender disparities in federal criminal cases. *American Law and Economics Review*, 17(1):127–159, 2015.

Darrell J. Steffensmeier and Robert M. Terry. Institutional Sexism in the Underworld: A View From the Inside. *Sociological Inquiry*, 56(3):304–323, 1986.

Richard Tay. Speed Cameras: Improving Safety or Raising Revenue? *Journal of Transport Economics and Policy*, 44(2):247–257, 2010.

Rob Tillyer, Richard D. Hartley, and Jeffrey T. Ward. Differential Treatment of Female Defendants. *Criminal Justice and Behavior*, 42(7):703–721, 2015.

Matthew Tonkin, Jessica Woodhams, Ray Bull, John W. Bond, and Pekka Santtila. A Comparison of Logistic Regression and Classification Tree Analysis for Behavioural Case Linkage. *Journal of Investigative Psychology and Offender Profiling*, 9(3):235–258, 2012.

Thomas L. Traynor. The impact of state level behavioral regulations on traffic fatality rates. *Journal of Safety Research*, 40(6):421–426, 2009.

Arthur Van Benthem. What Is the Optimal Speed Limit on Freeways? *Journal of Public Economics*, 124:44–62, 2015.

Ron Van Houten and Paul A. Nau. A Comparison of the Effects of Posted Feedback and Increased Police Surveillance on Highway Speeding. *Journal of Applied Behaviour Analysis*, 14(3):261–271, 1981.

Christy A. Visher. Gender, Police Arrest Decisions, and Notions of Chivalry. *Criminology*, 21(1):5–28, 1983.

Richard Williams. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, 6(1):58–82, 2006.

Richard Williams. Fitting heterogeneous choice models with oglm. *Stata Journal*, 10(3):288–308, 2010.

Richard Williams. Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, 5874(February):1–14, 2016.

R. Winship, C. & Mare. Regression Models with Ordinal Variables. *American Sociological Review*, 49(4):512–525, 1984.

Robert Witt, Alan Clarke, and Nigel Fielding. Crime, Earnings Inequality and Unemployment in England and Wales. *Applied Economics Letters*, 5:265–267, 1998.

A RISK PREFERENCES OF FEMALE OFFENDERS AND THE GENDER PARTICIPATION GAP IN CRIME

A.1 Variable Descriptions

Table 14: Offence Categories

Offence Group	Includes
Robbery	Robbery Attempted robbery
Snatch Theft	Snatch theft from the person
Theft from the person (TFTP)	Other theft from the person Attempted theft from the person
Domestic Burglary	Attempted burglary to non-connected domestic garage/outhouse Burglary in a dwelling (nothing taken) Burglary in a dwelling (something taken) Attempted burglary in a dwelling
Theft from a dwelling (TFAD)	Theft in a dwelling Theft from a meter
Domestic burglary in an outhouse (DBOH)	Burglary from non-connected domestic garage/outhouse - nothing taken Burglary from non-connected domestic garage/outhouse - something taken
Vehicle-related thefts	Theft of car/van Theft from car/van Theft of motorbike, motor scooter or moped Theft from motorbike, motorscooter or moped Attempted theft of/from car/van Attempted theft of/from motorcycle, motorscooter or moped Theft of pedal cycle
Theft from outside dwelling (TFOD)	Theft from outside dwelling
Other personal theft	Other personal theft
Criminal Damage	Arson Criminal damage to a motor vehicle (£20 or under) Criminal damage to a motor vehicle (over £20) Criminal damage to the home (£20 or under) Criminal damage to the home (over £20) Other criminal damage (£20 or under) Other criminal damage (over £20)

A.2 Regression Output

Table 16: Earnings Risk Regression

Earnings Risk	Coef.	Std. Err.	t	P>t
Female	-0.213***	0.004	-58.95	0.000
Age=16-24	-0.398***	0.006	-67.95	0.000
Age=25-39	-0.390***	0.006	-64.02	0.000
Age=40+	0.042***	0.006	6.60	0.000
Drug Influence	-0.161***	0.004	-45.52	0.000
Drink Influence	0.276***	0.004	74.41	0.000
Rel=Family	-1.107***	0.007	-152.23	0.000
Rel=Friend	-1.381***	0.007	-201.30	0.000
Rel=Acquain.	-0.542***	0.007	-81.25	0.000
Rel=F. Spouse	-1.306***	0.007	-196.20	0.000
Race=Black	-0.239***	0.006	-38.56	0.000
Race=Asian	-0.365***	0.007	-51.18	0.000
Weapon	0.803***	0.007	119.02	0.000
Contact	-0.148***	0.005	-29.91	0.000
Force	-0.149***	0.004	-36.15	0.000
Threaten	0.534***	0.004	135.20	0.000
Sexual	-0.351***	0.009	-39.40	0.000
Knew Offender	0.036***	0.007	5.47	0.000
Street Gang	-0.013*	0.007	-1.79	0.074
Time=Night	0.537***	0.003	172.55	0.000
Weekend	-0.235***	0.003	-72.31	0.000
Constant	-144.529***	1.687	-85.65	0.000
Observations	1,648,827			
Prob>F	0.000			
R Squared	0.1306			
Year Dummies	Yes			

Table 15: Crime Survey for England and Wales

Name	Description	Obs	Mean	S.D	Min	Max
Survey	Survey Year	289,440	2002.5	8.0	1981	2014
Contact	Did you have contact with the person?	252,113	0.311	0.463	0	1
Force	Did the person use any force or violence?	257,831	0.114	0.318	0	1
Threaten	Did the offender(s) threaten to use force or violence?	251,837	0.154	0.361	0	1
Sexual	Was there a sexual element to the offence?	252,296	0.009	0.093	0	1
Weapon	Did the person have a weapon?	92,524	0.092	0.289	0	1
Female	Was the person who did it male or female?	53,081	0.161	0.367	0	1
Age	How old was the person who did it?	51,965	3.591	0.917	1	5
Drug Influence	Was the person under the influence of drugs?	33,997	0.207	0.405	0	1
Drink Influence	Had the person been drinking?	40,506	0.355	0.479	0	1
Relationship	What was the person's relationship to you?	25,053	3.391	1.186	1	5
Race	What was the ethnic origin of the person who did it?	57,120	1.151	0.444	1	3
Knew Offender	Was it done by someone you knew?	52,046	0.548	0.498	0	1
Street Gang	Was the person who did it a member of a known street gang?	10,132	0.036	0.185	0	1
Time	At what time of day did it happen?	244,385	0.632	0.482	0	1
Weekend	Did it happen during the week or at a weekend?	246,406	0.352	0.478	0	1
Offence	Offence Code	211,196	64.337	11.291	41	80
Arrest	What action, if any, did the police take?	289,440	0.092	0.289	0	1
Earnings	What was the total replacement value of what was stolen?	95,761	405.094	1924.346	0	99,000

A.3 Tests for Multicollinearity

Table 17: Variance Inflation Factors

Variable	Log Earnings		Earnings Risk	
	VIF	1/VIF	VIF	1/VIF
Female	1.26	0.80	1.12	0.90
Age = 16 - 24	3.17	0.32	3.62	0.28
Age = 25 - 39	3.25	0.31	3.94	0.25
Age = 40+	2.43	0.41	3.39	0.30
Drug Influence	1.34	0.74	1.29	0.77
Drink Influence	1.34	0.75	1.36	0.73
Rel=Family	4.81	0.21	2.35	0.43
Rel=Friend	7.15	0.14	2.87	0.35
Rel=Acquain.	9.88	0.10	4.5	0.22
Rel=F. Spouse	3.82	0.26	2.99	0.33
Race=Black	1.12	0.89	1.05	0.95
Race=Asian	1.12	0.89	1.05	0.95
Weapon	1.31	0.76	1.11	0.90
Contact	1.16	0.86	1.08	0.93
Force	3.34	0.30	1.63	0.61
Threaten	2.78	0.36	1.62	0.62
Sexual	1.22	0.82	1.07	0.94
Knew Offender	1.18	0.85	1.12	0.90
Street Gang	1.16	0.86	1.11	0.90
Time=Night	1.38	0.72	1.24	0.81
Weekend	1.29	0.77	1.2	0.83
Mean VIF	2.69		1.87	

[12]

[12] If VIF < 10, multicollinearity does not require further investigation.

A.4 Robustness Checks

Table 18: Blinder-Oaxaca Decomposition

Log Crime	2001 - 2007		2008 - 2014	
Differential				
Prediction (Male)	11.418***	(0.204)	11.131	(0.189)
Prediction (Female)	9.791***	(0.220)	9.493	(0.201)
Difference	1.627***	(0.300)	1.638	(0.276)
Decomposition				
Endowments	-0.160***	(0.297)	-0.054	(0.256)
Coefficients	1.806***	(0.094)	1.725	(0.121)
Interaction	-0.019***	(0.084)	-0.033	(0.108)

A.5 Residual Distribution for Linear Regressions

Figure 19: Residuals for Gender Earnings Gap Regression in Table 4

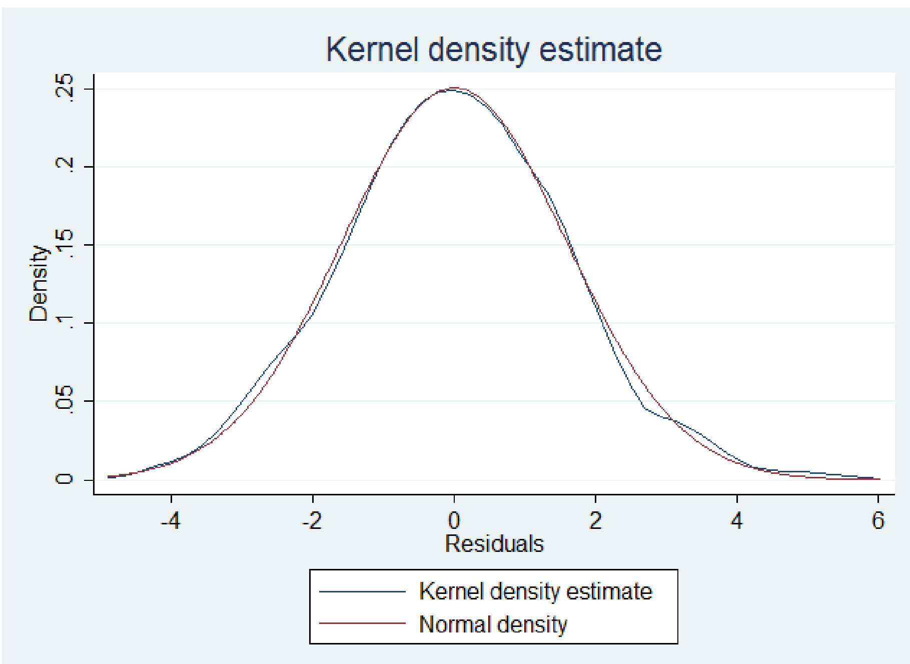


Figure 20: Residuals for Blinder-Oaxaca Decomposition - Male

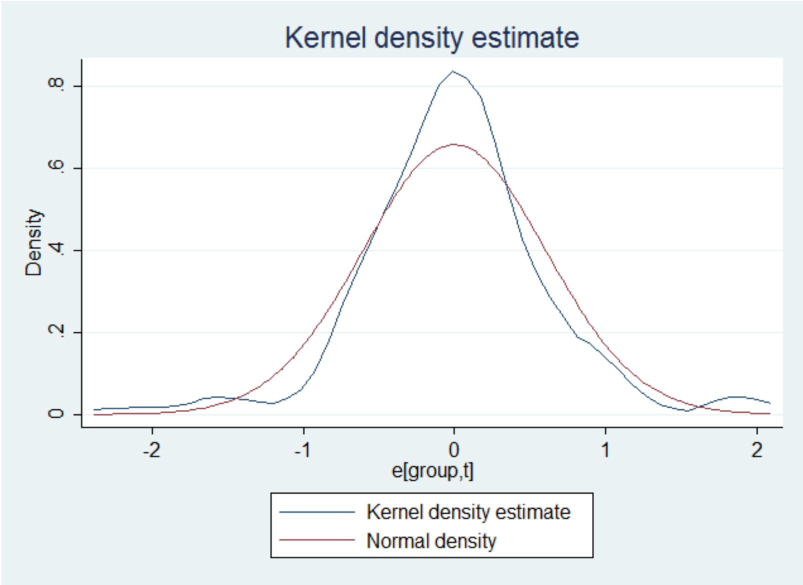
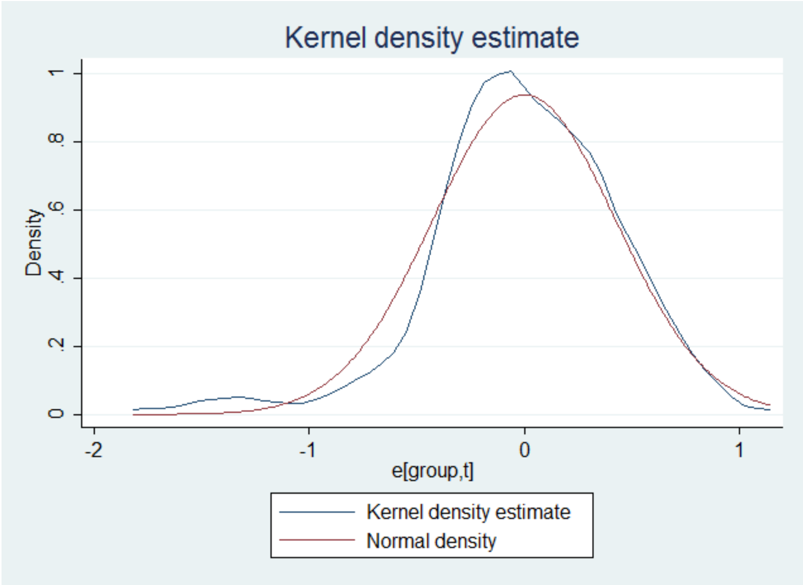


Figure 21: Residuals for Blinder-Oaxaca Decomposition - Female



B THE GENDER GAP IN SENTENCING

B.1 Variable Descriptions

Table 19: Arson and Criminal Damage

Gender	Frequency	Percentage
Male	2871	85.02%
Female	506	14.98%
Total	3377	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	1,002	118	1,120	89.46%	10.54%	33.17%
25 to 34	864	131	995	86.83%	13.17%	29.46%
35 to 44	553	123	676	81.80%	18.20%	20.02%
45 to 54	337	101	438	76.94%	23.06%	12.97%
Type of Offence						
Arson endangering life	946	277	1223	77.35%	22.65%	36.22%
Arson not endangering life	620	125	745	83.22%	16.78%	22.06%
Criminal damage endangering life (excl. arson)	310	31	341	90.91%	9.09%	10.10%
Other	756	59	815	92.76%	7.24%	24.13%
Sentence Outcome						
Other	337	68	405	83.21%	16.79%	11.99%
Community Order	478	114	592	80.74%	19.26%	17.53%
Suspended Sentence	454	96	550	82.55%	17.45%	16.29%
Immediate Custody	1,602	228	1,830	87.54%	12.46%	54.19%
Sentence Length						
Up to 1 year	431	22	453	95.14%	4.86%	25.80%
1 - 3 years	656	147	803	81.69%	18.31%	45.73%
3 - 5 years	328	37	365	89.86%	10.14%	20.79%
years or more	120	15	135	88.89%	11.11%	7.69%
Previous Convictions						
None	1,206	302	1,508	79.97%	20.03%	44.66%
1 to 3	668	83	751	88.95%	11.05%	22.24%
4 to 9	496	39	535	92.71%	7.29%	15.84%
Aggravating Factors						
Pre-planning or premeditation	565	69	634	89.12%	10.88%	18.77%
Member of a group or gang	233	20	253	92.09%	7.91%	7.49%

Damage of high value	497	76	573	86.74%	13.26%	16.97%
Offence motivated by/ demonstrating hostility to race/religion	50	4	54	92.59%	7.41%	1.60%
Victim particularly vulnerable	238	22	260	91.54%	8.46%	7.70%
Offender was under the influence of alcohol/drugs	827	152	979	84.47%	15.53%	28.99%
Act of revenge	505	69	574	87.98%	12.02%	17.00%
Damage to emergency equipment or a public amenity	115	3	118	97.46%	2.54%	3.49%
Significant public or private fear caused	346	48	394	87.82%	12.18%	11.67%
Offender was on bail	255	21	276	92.39%	7.61%	8.17%
More than one victim	258	29	287	89.90%	10.10%	8.50%
Mitigating Factors						
Age	705	132	837	84.23%	15.77%	24.79%
Genuine remorse	947	219	1,166	81.22%	18.78%	34.53%
Offender responding well to existing order/sentence	181	43	224	80.80%	19.20%	6.63%
Offender can/is addressing needs/addiction	397	123	520	76.35%	23.65%	15.40%
Offender is main carer/has responsibilities	86	46	132	65.15%	34.85%	3.91%
Currently in, or prospects of work/training	224	13	237	94.51%	5.49%	7.02%
Loss of job or reputation	120	11	131	91.60%	8.40%	3.88%
Physical or mental illness	627	228	855	73.33%	26.67%	25.32%
Difficult/deprived background	295	131	426	69.25%	30.75%	12.61%
Offence out of character	567	162	729	77.78%	22.22%	21.59%
Cooperation with authorities	279	78	357	78.15%	21.85%	10.57%
Provocation	90	27	117	76.92%	23.08%	3.46%
Guilty plea discount						
None	53	5	58	91.38%	8.62%	1.72%
1% - 10%	147	19	166	88.55%	11.45%	4.92%
11% - 20%	134	13	147	91.16%	8.84%	4.35%
21% - 32%	252	52	304	82.89%	17.11%	9.00%
33% or more	1,589	294	1,883	84.39%	15.61%	55.76%

Table 20: Assault

Gender	Frequency	Percentage
Male	45,667	90.21%
Female	4,956	9.79%
Total	50,623	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	18,515	1,777	20,292	91.24%	8.76%	40.08%
25 to 34	15,293	1,637	16,930	90.33%	9.67%	33.44%
35 to 44	7,257	982	8,239	88.08%	11.92%	16.27%
45 to 54	3,602	446	4,048	88.98%	11.02%	8.00%
Over 54	1,000	114	1,114	89.77%	10.23%	2.20%
Type of offence						
Affray	7,759	594	8,353	92.89%	7.11%	16.50%
Assault	5,053	677	5,730	88.18%	11.82%	11.32%
Cruelty/Neglect of a child	421	552	973	43.27%	56.73%	1.92%
Harassment	1,225	92	1,317	93.01%	6.99%	2.60%
GBH	12,456	1,170	13,626	91.41%	8.59%	26.92%
ABH	15,848	1,634	17,482	90.65%	9.35%	34.53%
Violent Disorder	1,011	29	1,040	97.21%	2.79%	2.05%
Sentence Outcome						
Other	8,439	1,646	10,085	83.68%	16.32%	19.92%
Suspended Sentence	14,434	2,017	16,451	87.74%	12.26%	32.50%
Immediate Custody	22,794	1,293	24,087	94.63%	5.37%	47.58%
Sentence Length						
Up to 1 year	9,724	580	10,304	94.37%	5.63%	42.78%
1 - 3 years	9,179	497	9,676	94.86%	5.14%	40.17%
3 years or more	3,570	200	3,770	94.69%	5.31%	15.65%
Previous convictions						
None	20,077	2,866	22,943	87.51%	12.49%	45.32%
1 to 3	11,877	879	12,756	93.11%	6.89%	25.20%
4 to 9	5,409	343	5,752	94.04%	5.96%	11.36%
10 or more	1,948	118	2,066	94.29%	5.71%	4.08%
Factors indicating greater harm						
Injury/fear of injury which is serious in context of the offence	8,529	632	9,161	93.10%	6.90%	18.10%
Victim particularly vulnerable	4,832	490	5,322	90.79%	9.21%	10.51%
Sustained or repeated assault on same person	8,556	678	9,234	92.66%	7.34%	18.24%

Factors indicating lesser harm						
Injury/fear of injury which is less serious in context of the offence	5,581	767	6,348	87.92%	12.08%	12.54%
Factors indicating higher culpability						
Race/religion	606	84	690	87.83%	12.17%	1.36%
Disability	36	10	46	78.26%	21.74%	0.09%
Sexual orientation	70	4	74	94.59%	5.41%	0.15%
Transgender identity	5	0	5	100.00%	0.00%	0.01%
Other aggravating factors						
Significant degree of premeditation	2,999	257	3,256	92.11%	7.89%	6.43%
Threatened/actual use of a weapon	14,015	1,462	15,477	90.55%	9.45%	30.57%
Intention to cause more serious harm	1,703	105	1,808	94.19%	5.81%	3.57%
Deliberately causes more harm than necessary	1,520	107	1,627	93.42%	6.58%	3.21%
Targetting of vulnerable victim(s)	2,703	195	2,898	93.27%	6.73%	5.72%
Leading role in group or gang	2,038	164	2,202	92.55%	7.45%	4.35%
Offence motivated by/demonstrating hostility to age or sex	141	13	154	91.56%	8.44%	0.30%
Factors indication lower culpability						
Subordinate role in group or gang	1,110	165	1,275	87.06%	12.94%	2.52%
Greater degree of provocation	2,374	349	2,723	87.18%	12.82%	5.38%
Lack of premeditation	6,672	833	7,505	88.90%	11.10%	14.82%
Mental disorder/learning disability where linked to the commission of the offence	811	197	1,008	80.46%	19.54%	1.99%
Excessive self defence	1,443	186	1,629	88.58%	11.42%	3.22%
Factors increasing seriousness						
Location	14,014	1,044	15,058	93.07%	6.93%	29.74%
Timing	7,762	521	8,283	93.71%	6.29%	16.36%
On-going effect on victim	7626	631	8257	92.36%	7.64%	16.31%
Offence against those in the public sector/service to public	2,146	179	2,325	92.30%	7.70%	4.59%
Presence of others	7,852	592	8,444	92.99%	7.01%	16.68%

Gratuitous degradation	817	72	889	91.90%	8.10%	1.76%
Victim compelled to leave home (domestic violence in particular)	750	55	805	93.17%	6.83%	1.59%
Failure to comply with current court orders	2,764	183	2,947	93.79%	6.21%	5.82%
On licence	940	29	969	97.01%	2.99%	1.91%
Attempt to conceal/dispose of evidence	285	39	324	87.96%	12.04%	0.64%
Failure to respond to warnings/concerns	684	66	750	91.20%	8.80%	1.48%
Offender was under the influence of alcohol/drugs	9,140	926	10,066	90.80%	9.20%	19.88%
Abuse of power/trust	1,083	148	1,231	87.98%	12.02%	2.43%
Exploiting contact arrangements	75	2	77	97.40%	2.60%	0.15%
Previous violence/threats	2,280	130	2,410	94.61%	5.39%	4.76%
Established evidence of community impact	257	15	272	94.49%	5.51%	0.54%
Steps taken to prevent reporting/assisting prosecution	204	17	221	92.31%	7.69%	0.44%
TICs	28	1	29	96.55%	3.45%	0.06%
Factors reducing seriousness						
No previous relevant convictions	8,818	1,434	10,252	86.01%	13.99%	20.25%
Single blow	5,458	695	6,153	88.70%	11.30%	12.15%
Remorse	12,141	1,571	13,712	88.54%	11.46%	27.09%
Good character/exemplary conduct	5,429	943	6,372	85.20%	14.80%	12.59%
Determination/demonstration to address addiction/behaviour	2,893	502	3,395	85.21%	14.79%	6.71%
Serious medical conditions	923	215	1,138	81.11%	18.89%	2.25%
Isolated incident	5,588	801	6,389	87.46%	12.54%	12.62%
Age/lack of maturity affecting responsibility	2,964	397	3,361	88.19%	11.81%	6.64%
Lapse of time not fault of offender	1,054	118	1,172	89.93%	10.07%	2.32%
Mental disorder/learning disability where not linked to the commission of the offence	1,193	296	1,489	80.12%	19.88%	2.94%
Sole/primary carer for dependent relative	941	597	1,538	61.18%	38.82%	3.04%

Guilty plea discount						
None	387	63	450	86.00%	14.00%	0.89%
1% - 10%	1,777	153	1,930	92.07%	7.93%	3.81%
11% - 20%	1,503	140	1,643	91.48%	8.52%	3.25%
21% - 32%	3,196	298	3,494	91.47%	8.53%	
33% or more	16,601	1,766	18,367	90.38%	9.62%	36.28%

Table 21: Burglary

Gender	Frequency	Percentage
Male	25,614	95.45%
Female	1,222	4.55%
Total	26,836	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	9,727	361	10,088	96.42%	3.58%	37.59%
25 to 34	9,171	508	9,679	94.75%	5.25%	36.07%
35 to 44	5,079	267	5,346	95.01%	4.99%	19.92%
45 to 54	1,431	76	1,507	94.96%	5.04%	5.62%
Over 54	206	10	216	95.37%	4.63%	0.80%
Type of offence						
Aggravated Burglary	597	26	623	95.83%	4.17%	2.32%
Domestic Burglary	19,161	1,005	20,166	95.02%	4.98%	75.15%
Non-domestic Burglary	5,324	164	5,488	97.01%	2.99%	20.45%
Other burglary	532	27	559	95.17%	4.83%	2.08%
Sentence outcome						
Other	2,156	209	2,365	91.16%	8.84%	8.81%
Suspended Sentence	4,146	318	4,464	92.88%	7.12%	16.63%
Immediate Custody	19,312	695	20,007	96.53%	3.47%	74.55%
Sentence length						
Up to 1 year	5,031	202	5,233	96.14%	3.86%	26.16%
1 - 3 years	10,750	385	11,135	96.54%	3.46%	55.66%
3 - 5 years	2,620	81	2,701	97.00%	3.00%	13.50%
5 years or more	742	24	766	96.87%	3.13%	3.83%
Factors indicating greater harm						
Theft of/damage to property causing significant degree of loss	4,406	138	4,544	96.96%	3.04%	16.93%
Soiling/ransacking/vandalism of property	2,278	58	2,336	97.52%	2.48%	8.70%
Victim on/returns to premises while offender present	5,865	356	6,221	94.28%	5.72%	23.18%
Significant physical/psychological injury or trauma	1,630	97	1,727	94.38%	5.62%	6.44%
Violence used/ threatened particularly involving a weapon	937	47	984	95.22%	4.78%	3.67%

Context of general public disorder	231	14	245	94.29%	5.71%	0.91%
Factors indicating lesser harm						
No physical/psychological injury or trauma	2,391	103	2,494	95.87%	4.13%	9.29%
No violence used/threatened and a weapon not produced	3,137	132	3,269	95.96%	4.04%	12.18%
Nothing stolen or of very low value	2,689	145	2,834	94.88%	5.12%	10.56%
Limited damage/disturbance to property	2,880	142	3,022	95.30%	4.70%	11.26%
Factors indicating higher culpability						
Deliberately targeted	4,608	319	4,927	93.53%	6.47%	18.36%
Significant degree of planning	3,724	149	3,873	96.15%	3.85%	14.43%
Equipped for burglary	3,361	42	3,403	98.77%	1.23%	12.68%
Weapon present on entry or carried	715	20	735	97.28%	2.72%	2.74%
Member of group or gang	5,074	184	5,258	96.50%	3.50%	19.59%
Factors indicating lower culpability						
Offender exploited by others	370	68	438	84.47%	15.53%	1.63%
Offence committed on impulse/limited intrusion	1,857	102	1,959	94.79%	5.21%	7.30%
Mental disorder/learning disability where linked to the commission of the offence	217	18	235	92.34%	7.66%	0.88%
Factors increasing seriousness						
Previous relevant convictions	13,379	484	13,863	96.51%	3.49%	51.66%
Offence committed on bail	1,133	63	1,196	94.73%	5.27%	4.46%
Other aggravating factors						
Child at home/returns	939	41	980	95.82%	4.18%	3.65%
Committed at night	4,939	164	5,103	96.79%	3.21%	19.02%
Abuse of power/trust	483	58	541	89.28%	10.72%	2.02%
Gratuitous degradation	130	13	143	90.91%	9.09%	0.53%
Steps taken to prevent reporting/assisting prosecution	92	3	95	96.84%	3.16%	0.35%
Victim compelled to leave home (domestic violence in particular)	168	16	184	91.30%	8.70%	0.69%
Established evidence of community impact	395	9	404	97.77%	2.23%	1.51%

Offender was under the influence of alcohol/drugs	2,813	170	2,983	94.30%	5.70%	11.12%
Failure to comply with current court orders	2,017	93	2,110	95.59%	4.41%	7.86%
On licence	2,039	36	2,075	98.27%	1.73%	7.73%
TICs	1,282	30	1,312	97.71%	2.29%	4.89%
Factors reducing seriousness or reflecting personal mitigation						
Subordinate role in group or gang	927	111	1,038	89.31%	10.69%	3.87%
Injuries caused recklessly	32	3	35	91.43%	8.57%	0.13%
Nothing stolen or of very low value	1,649	91	1,740	94.77%	5.23%	6.48%
Made voluntary reparation	158	4	162	97.53%	2.47%	0.60%
No previous relevant convictions	1,524	156	1,680	90.71%	9.29%	6.26%
Remorse	3,731	232	3,963	94.15%	5.85%	14.77%
Good character/exemplary conduct	629	64	693	90.76%	9.24%	2.58%
Determination/demonstration to address addiction/behaviour	1,604	140	1,744	91.97%	8.03%	6.50%
Serious medical conditions	242	24	266	90.98%	9.02%	0.99%
Age/lack of maturity affecting responsibility	1,260	60	1,320	95.45%	4.55%	4.92%
Lapse of time not fault of offender	159	15	174	91.38%	8.62%	0.65%
Mental disorder/learning disability where not linked to the commission of the offence	384	44	428	89.72%	10.28%	1.59%
Sole/primary carer for dependent relatives	228	65	293	77.82%	22.18%	1.09%
Guilty plea discount						
None	189	8	197	95.94%	4.06%	0.73%
1% - 10%	1332	57	1389	95.90%	4.10%	5.18%
11% - 20%	2351	86	2437	96.47%	3.53%	9.08%
21% - 32%	3086	136	3222	95.78%	4.22%	12.01%
33% or more	13403	625	14028	95.54%	4.46%	52.27%

Table 22: Death

Gender	Frequency	Percentage
Male	2,503	90.95%
Female	249	9.05%
Total	2,752	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	721	49	770	93.64%	6.36%	27.98%
25 to 34	735	71	806	91.19%	8.81%	29.29%
35 to 44	488	60	548	89.05%	10.95%	19.91%
45 to 54	364	42	406	89.66%	10.34%	14.75%
Type of offence						
Attempted murder	186	14	200	93.00%	7.00%	7.27%
Causing death by careless driving when under the influence of drink or drugs	348	62	410	84.88%	15.12%	14.90%
Causing death by dangerous driving	272	18	290	93.79%	6.21%	10.54%
Making threats to kill	498	32	530	93.96%	6.04%	19.26%
Manslaughter	316	30	346	91.33%	8.67%	12.57%
Murder of persons aged one year or over	712	63	775	91.87%	8.13%	28.16%
Sentence Outcome						
Other	241	57	298	80.87%	19.13%	10.83%
Suspended Sentence	252	42	294	85.71%	14.29%	10.68%
Immediate Custody	2,010	150	2,160	93.06%	6.94%	78.49%
Sentence Length						
Up to 5 years	762	57	819	93.04%	6.96%	37.92%
5 years or over	1,162	90	1,252	92.81%	7.19%	57.96%
Seriousness						
1	390	29	419	93.08%	6.92%	15.23%
2	433	42	475	91.16%	8.84%	17.26%
3	273	45	318	85.85%	14.15%	11.56%
Previous Convictions						
None	1,556	200	1,756	88.61%	11.39%	63.81%
1 to 3	144	7	151	95.36%	4.64%	5.49%
3 or more	431	17	448	96.21%	3.79%	16.28%
Aggravating Factors						
Pre-planning or premeditation	526	43	569	92.44%	7.56%	20.68%
Victim particularly vulnerable	507	62	569	89.10%	10.90%	20.68%

Mental or physical suffering inflicted on the victim	274	25	299	91.64%	8.36%	10.86%
Abuse of power/trust	109	21	130	83.85%	16.15%	4.72%
Use of weapon	72	5	77	93.51%	6.49%	2.80%
Concealment, destruction or dismemberment of the body	263	14	277	94.95%	5.05%	10.07%
Driving off to avoid detection or apprehension	60	0	60	100.00%	0.00%	2.18%
Serious injury to others in addition to the death(s)	195	17	212	91.98%	8.02%	7.70%
Offender was on bail or licence	194	1	195	99.49%	0.51%	7.09%
More than one victim	144	12	156	92.31%	7.69%	5.67%
Mitigating factors						
Age	734	82	816	89.95%	10.05%	29.65%
Genuine remorse	897	117	1,014	88.46%	11.54%	36.85%
Offender responding well to existing order/sentence	27	3	30	90.00%	10.00%	1.09%
Provocation	155	16	171	90.64%	9.36%	6.21%
Good driving record	296	56	352	84.09%	15.91%	12.79%
Lack of premeditation	548	83	631	86.85%	13.15%	22.93%
Acted to an extent in self-defence	47	5	52	90.38%	9.62%	1.89%
Mental disorder/learning disability	277	59	336	82.44%	17.56%	12.21%
Giving assistance	50	9	59	84.75%	15.25%	2.14%
Effect on the offender	233	49	282	82.62%	17.38%	10.25%
Actions of the victim or a third party	69	10	79	87.34%	12.66%	2.87%
Guilty plea discount						
1% - 10%	144	9	153	94.12%	5.88%	5.56%
11% - 20%	205	16	221	92.76%	7.24%	8.03%
21% - 32%	227	8	235	96.60%	3.40%	8.54%
33% or more	737	80	817	90.21%	9.79%	29.69%

Table 23: Driving

Gender	Frequency	Percentage
Male	7,493	95.68%
Female	338	4.32%
Total	7,831	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	3,195	97	3,292	97.05%	2.95%	42.04%
25 to 34	2,674	115	2,789	95.88%	4.12%	35.61%
35 to 44	1,019	70	1,089	93.57%	6.43%	13.91%
45 to 54	446	41	487	91.58%	8.42%	6.22%
Over 54	159	15	174	91.38%	8.62%	2.22%
Type of Offence						
Aggravated vehicle taking	1,213	47	1,260	96.27%	3.73%	16.09%
Careless driving	322	25	347	92.80%	7.20%	4.43%
Dangerous driving	5,119	210	5,329	96.06%	3.94%	68.05%
Making false statements to obtain or failure to produce revoked license	34	2	36	94.44%	5.56%	0.46%
Other driving offence	805	54	859	93.71%	6.29%	10.97%
Sentence Outcome						
Other	444	36	480	92.50%	7.50%	6.13%
Community Order	1,022	85	1,107	92.32%	7.68%	14.14%
Suspended Sentence	2,349	142	2,491	94.30%	5.70%	31.81%
Immediate Custody	3,678	75	3,753	98.00%	2.00%	47.92%
Sentence Length						
Up to 1 year	2,517	56	2,573	97.82%	2.18%	68.56%
12 to 18 months	820	8	828	99.03%	0.97%	10.57%
18 months or more	281	10	291	96.56%	3.44%	3.72%
Previous Convictions						
None	2,559	180	2,739	93.43%	6.57%	34.98%
1 to 3	2,221	75	2,296	96.73%	3.27%	29.32%
4 to 9	951	15	966	98.45%	1.55%	12.34%
10 or more	633	12	645	98.14%	1.86%	8.24%
Aggravating Factors						
Offender was under the influence of alcohol/drugs	1,959	112	2,071	94.59%	5.41%	26.45%
Disregard of warnings	1,394	46	1,440	96.81%	3.19%	18.39%
Aggressive driving	2,684	87	2,771	96.86%	3.14%	35.39%
Carrying out other tasks while driving	84	9	93	90.32%	9.68%	1.19%
Injury to others	1,092	73	1,165	93.73%	6.27%	14.88%

Damage to other vehicles or property	2,200	100	2,300	95.65%	4.35%	29.37%
Poorly maintained or dangerously loaded vehicle	98	3	101	97.03%	2.97%	1.29%
Tiredness	54	7	61	88.52%	11.48%	0.78%
Driving when knowingly suffering from a medical condition which significantly impairs driving	62	7	69	89.86%	10.14%	0.88%
Offender was on bail or licence	911	18	929	98.06%	1.94%	11.86%
More than one victim	450	22	472	95.34%	4.66%	6.03%
Mitigating Factors						
Age	1,961	105	2,066	94.92%	5.08%	26.38%
Genuine remorse	2,506	160	2,666	94.00%	6.00%	34.04%
Offender responding well to existing order/sentence	367	18	385	95.32%	4.68%	4.92%
Offender can/is addressing needs/addiction	599	47	646	92.72%	7.28%	8.25%
Offender is main carer/has responsibilities	659	105	764	86.26%	13.74%	9.76%
Currently in, or prospects of work/training	1,374	51	1,425	96.42%	3.58%	18.20%
Loss of job or reputation	688	25	713	96.49%	3.51%	9.10%
Physical or mental illness	451	58	509	88.61%	11.39%	6.50%
Difficult/deprived background	295	33	328	89.94%	10.06%	4.19%
Offence out of character	1,298	117	1,415	91.73%	8.27%	18.07%
Co-operation with authorities	604	37	641	94.23%	5.77%	8.19%
Good driving record	817	65	882	92.63%	7.37%	11.26%
Genuine emergency	37	6	43	86.05%	13.95%	0.55%
Guilty Plea Discount						
None	95	3	98	96.94%	3.06%	1.25%
1% - 10%	383	16	399	95.99%	4.01%	5.10%
11% - 20%	331	15	346	95.66%	4.34%	4.42%
21% - 32%	846	36	882	95.92%	4.08%	11.26%
33% or more	4414	186	4600	95.96%	4.04%	58.74%

Table 24: Drugs

Gender	Frequency	Percentage
Male	33,925	90.92%
Female	3,388	9.08%
Total	37,313	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	10,656	762	11,418	93.33%	6.67%	30.60%
25 to 34	12,413	1,262	13,675	90.77%	9.23%	36.65%
35 to 44	6,525	833	7,358	88.68%	11.32%	19.72%
45 to 54	3,362	428	3,790	88.71%	11.29%	10.16%
Over 54	969	103	1,072	90.39%	9.61%	2.87%
Type of Offence						
Bringing in/taking out controlled drug	1,038	215	1,253	82.84%	17.16%	3.36%
Conspiracy to supply	2,197	160	2,357	93.21%	6.79%	6.32%
Other drug offences	190	102	292	65.07%	34.93%	0.78%
Permitting premises to be used	378	333	711	53.16%	46.84%	1.91%
Possession	2,699	291	2,990	90.27%	9.73%	8.01%
Possession with intent to supply	13,065	976	14,041	93.05%	6.95%	37.63%
Production/being concerned in production/cultivation	8,322	630	8,952	92.96%	7.04%	23.99%
Supply	6,036	681	6,717	89.86%	10.14%	18.00%
Sentence outcome						
Other	1,167	192	1,359	85.87%	14.13%	3.64%
Community Order	4,114	884	4,998	82.31%	17.69%	13.39%
Suspended Sentence	9,809	1,248	11,057	88.71%	11.29%	29.63%
Immediate Custody	18,835	1,064	19,899	94.65%	5.35%	53.33%
Sentence length						
Up to 1 year	3,848	267	4,115	93.51%	6.49%	11.03%
1 - 3 years	8,944	525	9,469	94.46%	5.54%	25.38%
3 to 5 years	3,684	168	3,852	95.64%	4.36%	10.32%
More than 5 years	2,180	93	2,273	95.91%	4.09%	6.09%
Previous convictions						
None	19,165	2,398	21,563	88.88%	11.12%	57.79%
1 to 3	8,639	503	9,142	94.50%	5.50%	24.50%
4 to 9	2,221	116	2,337	95.04%	4.96%	6.26%
5 or more	914	56	970	94.23%	5.77%	2.60%
Drug/Class of drug associated with the offence						

Cannabis or Cannabis	14,079	1,312	15,391	91.48%	8.52%	41.25%
Cocaine	8,222	620	8,842	92.99%	7.01%	23.70%
Heroin	4,365	542	4,907	88.95%	11.05%	13.15%
Other Class A	981	84	1,065	92.11%	7.89%	2.85%
Other Class B	1,983	269	2,252	88.06%	11.94%	6.04%
Class C	336	74	410	81.95%	18.05%	1.10%
Factors increasing seriousness						
Previous relevant convictions	9,039	520	9,559	94.56%	5.44%	25.62%
Permitted under 18 year old to deliver etc.	43	4	47	91.49%	8.51%	0.13%
Offence committed on bail	663	26	689	96.23%	3.77%	1.85%
18 years or over supplies in the vicinity of school etc.	10	1	11	90.91%	9.09%	0.03%
Other aggravating factors						
Sophisticated nature of concealment/attempts to avoid detection	849	48	897	94.65%	5.35%	2.40%
Attempt to conceal/dispose of evidence	567	34	601	94.34%	5.66%	1.61%
Exposure of others to more than usual danger	81	14	95	85.26%	14.74%	0.25%
Presence of weapon	203	5	208	97.60%	2.40%	0.56%
High purity or high potential yield	2,726	149	2,875	94.82%	5.18%	7.71%
Failure to comply with current court orders	991	48	1,039	95.38%	4.62%	2.78%
On licence	672	15	687	97.82%	2.18%	1.84%
Targeting premises of vulnerable people	44	3	47	93.62%	6.38%	0.13%
On-going/large scale evidenced by specialist equipment	1,399	62	1,461	95.76%	4.24%	3.92%
Presence of others, especially children and/or non-users	507	125	632	80.22%	19.78%	1.69%
Use of premises with unlawful access to utility supply	850	59	909	93.51%	6.49%	2.44%
Level of profit element	2,367	127	2,494	94.91%	5.09%	6.68%
Premises adapted to facilitate drug activity	1,139	56	1,195	95.31%	4.69%	3.20%
Location of premises	318	36	354	89.83%	10.17%	0.95%
Length of time premises used	407	36	443	91.87%	8.13%	1.19%
Charged as importation of very small amount	11	0	11	100.00%	0.00%	0.03%

Nature of likely supply	879	66	945	93.02%	6.98%	2.53%
Possession of drug in school/licensed premises	50	3	53	94.34%	5.66%	0.14%
Possession of drug in prison	116	75	191	60.73%	39.27%	0.51%
Volume of activity permitted	427	40	467	91.43%	8.57%	1.25%
Established evidence of community impact	389	25	414	93.96%	6.04%	1.11%
Factors reducing seriousness or reflecting personal mitigation						
Lack of sophistication as to nature of concealment	2,511	330	2,841	88.38%	11.62%	7.61%
Involvement due to pressure/intimidation/coercion	1,761	472	2,233	78.86%	21.14%	5.98%
Mistaken belief regarding type of drug	101	17	118	85.59%	14.41%	0.32%
Isolated incident	2,130	410	2,540	83.86%	16.14%	6.81%
Low purity	782	65	847	92.33%	7.67%	2.27%
No previous relevant convictions	7,450	1,006	8,456	88.10%	11.90%	22.66%
Offender's vulnerability exploited	1402	490	1892	74.10%	25.90%	5.07%
Remorse	5,744	838	6,582	87.27%	12.73%	17.64%
Good character/exemplary conduct	3,433	587	4,020	85.40%	14.60%	10.77%
Determination/demonstration to address addiction/behaviour	3,037	313	3,350	90.66%	9.34%	8.98%
Serious medical conditions	888	132	1,020	87.06%	12.94%	2.73%
Age/lack of maturity affecting responsibility	1,947	171	2,118	91.93%	8.07%	5.68%
Mental disorder/learning disability	442	67	509	86.84%	13.16%	1.36%
Sole/primary carer for dependent relatives	911	458	1,369	66.54%	33.46%	3.67%
Offender addicted to same drug	3,787	365	4,152	91.21%	8.79%	11.13%
Offender using cannabis to help diagnosed medical condition	457	33	490	93.27%	6.73%	1.31%
Guilty plea discount						
None	1817	182	1999	90.90%	9.10%	5.36%
1% - 10%	1351	101	1452	93.04%	6.96%	3.89%
11% - 20%	1835	124	1959	93.67%	6.33%	5.25%
21% - 32%	4247	390	4637	91.59%	8.41%	12.43%

33% or more	20176	2039	22215	90.82%	9.18%	59.54%

Table 25: Other

Gender	Frequency	Percentage
Male	16,083	88.50%
Female	2,089	11.50%
Total	18,172	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	4,535	517	5,052	89.77%	10.23%	27.80%
25 to 34	5,263	679	5,942	88.57%	11.43%	32.70%
35 to 44	3,262	487	3,749	87.01%	12.99%	20.63%
45 to 54	2,116	310	2,426	87.22%	12.78%	13.35%
Over 54	907	96	1,003	90.43%	9.57%	5.52%
Type of offence						
Absconding from lawful custody	527	12	539	97.77%	2.23%	2.97%
Blackmail	330	30	360	91.67%	8.33%	1.98%
Breach of ASBO	409	45	454	90.09%	9.91%	2.50%
Breach of protective order	2,591	73	2,664	97.26%	2.74%	14.66%
False imprisonment	385	35	420	91.67%	8.33%	2.31%
Intimidating a juror or witness	506	82	588	86.05%	13.95%	3.24%
Kidnapping	365	50	415	87.95%	12.05%	2.28%
Other offences	2,313	594	2,907	79.57%	20.43%	16.00%
Perverting the course of public justice	1,659	668	2,327	71.29%	28.71%	12.81%
Possession of offensive weapons without lawful authority or reasonable excuse	4,451	366	4,817	92.40%	7.60%	26.51%
Possession/distribution of prohibited weapons or ammunition	2,313	101	2,414	95.82%	4.18%	13.28%
Unauthorised use of trademark	234	33	267	87.64%	12.36%	1.47%
Sentence outcome						
Other	838	140	978	85.69%	14.31%	5.38%
Community Order	2,053	329	2,382	86.19%	13.81%	13.11%
Suspended Sentence	4,293	840	5,133	83.64%	16.36%	28.25%
Immediate Custody	8,899	780	9,679	91.94%	8.06%	53.26%
Sentence length						
Up to 1 year	5,252	459	5,711	91.96%	8.04%	59.00%
1 to 3 years	2,232	244	2,476	90.15%	9.85%	25.58%

More than 3 years	1,286	68	1,354	94.98%	5.02%	13.99%
Seriousness						
1	525	41	566	92.76%	7.24%	3.11%
2	1,006	90	1,096	91.79%	8.21%	6.03%
3	958	86	1,044	91.76%	8.24%	5.75%
4	822	73	895	91.84%	8.16%	4.93%
Previous convictions						
None	5,882	1,262	7,144	82.33%	17.67%	39.31%
1 to 3	4,277	261	4,538	94.25%	5.75%	24.97%
4 to 9	2,095	132	2,227	94.07%	5.93%	12.26%
10 or more	1,190	86	1,276	93.26%	6.74%	7.02%
Aggravating factors						
Pre-planning or premeditation	2,303	337	2,640	87.23%	12.77%	14.53%
Member of group or gang	1,194	146	1,340	89.10%	10.90%	7.37%
Intimidation or force used	1,874	120	1,994	93.98%	6.02%	10.97%
Victim particularly vulnerable	1,689	180	1,869	90.37%	9.63%	10.29%
Use of drugs, alcohol or another substance to facilitate the offence	1,063	89	1,152	92.27%	7.73%	6.34%
Background of intimidation or coercion	1,062	57	1,119	94.91%	5.09%	6.16%
Threats to prevent victim reporting the incident	328	22	350	93.71%	6.29%	1.93%
Financial or other gain	850	118	968	87.81%	12.19%	5.33%
Professionalism	273	16	289	94.46%	5.54%	1.59%
Motivated by hostility towards an individual/group	605	105	710	85.21%	14.79%	3.91%
Detrimental impact on the administration of justice	910	224	1,134	80.25%	19.75%	6.24%
Offender was on bail or licence	941	61	1,002	93.91%	6.09%	5.51%
Mitigating factors						
Age	3,326	588	3,914	84.98%	15.02%	21.54%
Genuine remorse	4,236	824	5,060	83.72%	16.28%	27.85%
Offender responding well to existing order/sentence	671	67	738	90.92%	9.08%	4.06%
Offender can/is addressing needs/addiction	1,383	185	1,568	88.20%	11.80%	8.63%
Offender is main carer/has responsibilities	913	478	1,391	65.64%	34.36%	7.65%
Currently in, or prospects of work/training	1,516	163	1,679	90.29%	9.71%	9.24%
Loss of job or reputation	938	221	1,159	80.93%	19.07%	6.38%

Physical or mental illness	1,557	389	1,946	80.01%	19.99%	10.71%
Difficult/deprived background	890	302	1,192	74.66%	25.34%	6.56%
Offence out of character	2,714	716	3,430	79.13%	20.87%	18.88%
Co-operation with authorities	1,450	295	1,745	83.09%	16.91%	9.60%
Guilty plea discount						
None	273	23	296	92.23%	7.77%	1.63%
1% - 10%	988	101	1089	90.73%	9.27%	5.99%
11% - 20%	800	87	887	90.19%	9.81%	4.88%
21% - 32%	1628	182	1810	89.94%	10.06%	9.96%
33% or more	8616	1124	9740	88.46%	11.54%	53.60%

Table 26: Robbery

Gender	Frequency	Percentage
Male	11,548	93.47%
Female	807	6.53%
Total	12,355	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	6,556	372	6,928	94.63%	5.37%	56.07%
25 to 34	3,318	273	3,591	92.40%	7.60%	29.07%
35 to 44	1,268	126	1,394	90.96%	9.04%	11.28%
45 to 54	355	35	390	91.03%	8.97%	3.16%
Over 54	51	1	52	98.08%	1.92%	0.42%
Type of offence						
Assault with intent to rob	161	19	180	89.44%	10.56%	1.46%
Other robbery	43	4	47	91.49%	8.51%	0.38%
Robbery	11,344	784	12,128	93.54%	6.46%	98.16%
Sentence outcome						
Other	371	59	430	86.28%	13.72%	3.48%
Suspended Sentence	1,132	173	1,305	86.74%	13.26%	10.56%
Immediate Custody	10,045	575	10,620	94.59%	5.41%	85.96%
Sentence length						
Up to 1 year	911	67	978	93.15%	6.85%	9.21%
1 to 3 years	4,376	323	4,699	93.13%	6.87%	44.25%
3 to 5 years	2,601	130	2,731	95.24%	4.76%	25.72%
More than 5 years	1,925	45	1,970	97.72%	2.28%	15.94%
Seriousness						
1	1,389	87	1,476	94.11%	5.89%	11.95%
2	5,558	380	5,938	93.60%	6.40%	48.06%
3	2,500	208	2,708	92.32%	7.68%	21.92%
Previous convictions						
None	3,328	302	3,630	91.68%	8.32%	29.38%
1 to 3	3,498	210	3,708	94.34%	5.66%	30.01%
4 to 9	1,964	106	2,070	94.88%	5.12%	16.75%
10 or more	1,191	75	1,266	94.08%	5.92%	10.25%
Aggravating factors						
Member of group or gang	5,656	370	6,026	93.86%	6.14%	48.77%
Targeting of vulnerable victim(s)	5,238	402	5,640	92.87%	7.13%	45.65%
Use of a weapon (including body and shod feet)	4,543	230	4,773	95.18%	4.82%	38.63%
Offender was under the influence of alcohol/drugs	2,780	221	3,001	92.64%	7.36%	24.29%

Offender was on bail or licence	1,665	64	1,729	96.30%	3.70%	13.99%
Degree of force or violence	2,819	171	2,990	94.28%	5.72%	24.20%
Wearing of a disguise	2,232	43	2,275	98.11%	1.89%	18.41%
Value of items taken	1,462	46	1,508	96.95%	3.05%	12.21%
More than one victim	1,808	81	1,889	95.71%	4.29%	15.29%
Offence committed at night/hours of darkness	3,783	211	3,994	94.72%	5.28%	32.33%
Mitigating factors						
Age	3,993	263	4,256	93.82%	6.18%	34.45%
Genuine remorse	3,592	291	3,883	92.51%	7.49%	31.43%
Offender responding well to existing order/sentence	391	67	458	85.37%	14.63%	3.71%
Offender can/is addressing needs/addiction	969	130	1,099	88.17%	11.83%	8.90%
Offender is main carer/has responsibilities	292	71	363	80.44%	19.56%	2.94%
Currently in, or prospects of work/training	627	35	662	94.71%	5.29%	5.36%
Loss of job or reputation	221	11	232	95.26%	4.74%	1.88%
Physical or mental illness	665	87	752	88.43%	11.57%	6.09%
Difficult/deprived background	1,197	209	1,406	85.14%	14.86%	11.38%
Offence out of character	1,674	169	1,843	90.83%	9.17%	14.92%
Co-operation with authorities	891	69	960	92.81%	7.19%	7.77%
Voluntary return of stolen property	134	7	141	95.04%	4.96%	1.14%
Unplanned/opportunistic	1,872	167	2,039	91.81%	8.19%	16.50%
Peripheral involvement	295	48	343	86.01%	13.99%	2.78%
Guilty plea discount						
None	139	11	150	92.67%	7.33%	1.21%
1% - 10%	857	61	918	93.36%	6.64%	7.43%
11% - 20%	692	741	1433	48.29%	51.71%	11.60%
21% - 32%	1477	1571	3048	48.46%	51.54%	24.67%
33% or more	5878	406	6284	93.54%	6.46%	50.86%

Table 27: Sexual

Gender	Frequency	Percentage
Male	6,712	98.65%
Female	92	1.35%
Total	6,804	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	1,272	15	1,287	98.83%	1.17%	18.92%
25 to 34	1,551	22	1,573	98.60%	1.40%	23.12%
35 to 44	1,333	33	1,366	97.58%	2.42%	20.08%
45 to 54	1,264	13	1,277	98.98%	1.02%	18.77%
Over 54	1,292	9	1,301	99.31%	0.69%	19.12%
Type of offence						
Other sexual offences	1,959	57	2,016	97.17%	2.83%	29.63%
Rape	1,393	5	1,398	99.64%	0.36%	20.55%
Sexual activity with a child	827	5	832	99.40%	0.60%	12.23%
Sexual assault	1,663	19	1,682	98.87%	1.13%	24.72%
Indecent photos of children	724	6	730	99.18%	0.82%	10.73%
Sentence outcome						
Other	1,201	16	1,217	98.69%	1.31%	17.89%
Suspended Sentence	917	17	934	98.18%	1.82%	13.73%
Immediate Custody	4,594	59	4,653	98.73%	1.27%	68.39%
Sentence length						
Up to 1 year	700	16	716	97.77%	2.23%	15.39%
1 to 3 years	1,404	24	1,428	98.32%	1.68%	30.69%
3 to 5 years	698	4	702	99.43%	0.57%	15.09%
More than 5 years	1,485	12	1,497	99.20%	0.80%	32.17%
Seriousness						
1	1,970	30	2,000	98.50%	1.50%	29.39%
2	1,837	24	1,861	98.71%	1.29%	27.35%
3	1,652	28	1,680	98.33%	1.67%	24.69%
4	1,253	10	1,263	99.21%	0.79%	18.56%
Previous convictions						
None	4,755	75	4,830	98.45%	1.55%	70.99%
1 to 3	1,003	4	1,007	99.60%	0.40%	14.80%
4 to 9	227	0	227	100.00%	0.00%	3.34%
10 or more	100	0	100	100.00%	0.00%	1.47%
Aggravating factors						
Planning or pre-meditation	1,222	18	1,240	98.55%	1.45%	18.22%
A sustained assault or repeated assaults on the same victim	1,909	18	1,927	99.07%	0.93%	28.32%

Offender was under the influence of alcohol/drugs	800	8	808	99.01%	0.99%	11.88%
Offender was on bail or license	298	1	299	99.67%	0.33%	4.39%
Abuse of power/trust	2,310	28	2,338	98.80%	1.20%	34.36%
Victim was particularly vulnerable	2,343	26	2,369	98.90%	1.10%	34.82%
Background of intimidation or coercion	517	7	524	98.66%	1.34%	7.70%
Abduction or detention	296	2	298	99.33%	0.67%	4.38%
Threats to prevent victim reporting the incident	417	7	424	98.35%	1.65%	6.23%
Physical harm caused	447	3	450	99.33%	0.67%	6.61%
More than one victim	1,011	7	1,018	99.31%	0.69%	14.96%
Mitigating factors						
Age	1,986	17	2,003	99.15%	0.85%	29.44%
Genuine remorse	1,934	32	1,966	98.37%	1.63%	28.89%
Offender responding well to exiting order/sentence	90	1	91	98.90%	1.10%	1.34%
Offender can/is addressing needs/addiction	648	2	650	99.69%	0.31%	9.55%
Offender is main carer/has responsibilities	255	12	267	95.51%	4.49%	3.92%
Currently in, or prospects of, work/training	444	3	447	99.33%	0.67%	6.57%
Loss of job or reputation	997	12	1,009	98.81%	1.19%	14.83%
Physical or mental illness	647	10	657	98.48%	1.52%	9.66%
Difficult/deprived background	364	15	379	96.04%	3.96%	5.57%
Offence out of character	1,681	23	1,704	98.65%	1.35%	25.04%
Co-operation with authorities	783	13	796	98.37%	1.63%	11.70%
Victim engaged in consensual sexual activity	567	11	578	98.10%	1.90%	8.50%
Reasonable belief that the victim was ages 16 or over	25	0	25	100.00%	0.00%	0.37%
Minimal contact	365	2	367	99.46%	0.54%	5.39%
Guilty plea discount						
None	176	1	177	99.44%	0.56%	2.60%
1% - 10%	360	7	367	98.09%	1.91%	5.39%
11% - 20%	374	5	379	98.68%	1.32%	5.57%
21% - 32%	616	624	1240	49.68%	50.32%	18.22%
33% or more	2907	47	2954	98.41%	1.59%	43.42%

Table 28: Theft and Fraud

Gender	Frequency	Percentage
Male	12,161	71.87%
Female	4,759	28.13%
Total	16,920	100.00%

	Male	Female	Total	Male %	Female %	% of total offences
Age						
18 to 24	2,388	499	2,887	82.72%	17.28%	17.06%
25 to34	4,012	1,324	5,336	75.19%	24.81%	31.54%
35 to 44	2,820	1,380	4,200	67.14%	32.86%	24.82%
45 to 54	1,922	1,059	2,981	64.48%	35.52%	17.62%
Over 55	1,019	497	1,516	67.22%	32.78%	8.96%
Type of offence						
Dishonest representation for obtaining benefit	1,280	1,685	2,965	43.17%	56.83%	17.52%
Other fraud	3,840	1,028	4,868	78.88%	21.12%	28.77%
Other theft, dishonesty and fraud	1,056	129	1,185	89.11%	10.89%	7.00%
Receiving stolen goods	541	66	607	89.13%	10.87%	3.59%
Theft from person	1,531	393	1,924	79.57%	20.43%	11.37%
Theft from shops and stalls	1,564	444	2,008	77.89%	22.11%	11.87%
Theft in breach of trust	2,056	963	3,019	68.10%	31.90%	17.84%
With intent knowingly possess false/improperly obtained passport/another ID document	293	51	344	85.17%	14.83%	2.03%
Sentence outcome						
Other	447	191	638	70.06%	29.94%	3.77%
Community Order	1,946	840	2,786	69.85%	30.15%	16.47%
Suspended Sentence	3,710	2,300	6,010	61.73%	38.27%	35.52%
Immediate Custody	6,058	1,428	7,486	80.92%	19.08%	44.24%
Sentence length						
Up to 1 year	3,365	868	4,233	79.49%	20.51%	56.55%
1 to 3 years	2,148	490	2,638	81.43%	18.57%	35.24%
3 to 5 years	493	62	555	88.83%	11.17%	7.41%
Seriousness						
1	1,510	404	1,914	78.89%	21.11%	11.31%
2	3,364	1,274	4,638	72.53%	27.47%	27.41%
3	3,519	1,855	5,374	65.48%	34.52%	31.76%
4	2,282	822	3,104	73.52%	26.48%	18.35%
5	1,486	404	1,890	78.62%	21.38%	11.17%
Previous convictions						

None	5,854	3,220	9,074	64.51%	35.49%	53.63%
1 to 3	2,310	543	2,853	80.97%	19.03%	16.86%
4 to 9	1,182	234	1,416	83.47%	16.53%	8.37%
10 or more	1,550	310	1,860	83.33%	16.67%	10.99%
Aggravating factors						
Pre-planning or premeditation	4,358	1,267	5,625	77.48%	22.52%	33.24%
Member of a group or gang	2,233	440	2,673	83.54%	16.46%	15.80%
Targeting of vulnerable victim(s)	2,233	440	2,673	83.54%	16.46%	15.80%
Offender was under the influence of alcohol/drugs	634	144	778	81.49%	18.51%	4.60%
Offender was on bail	1,158	210	1,368	84.65%	15.35%	8.09%
Victim particularly vulnerable	995	557	1552	64.11%	35.89%	9.17%
High value (including sentimental value) of the property to the victim or substantial consequent	1967	683	2650	74.23%	25.77%	15.66%
High level of gain	1,935	1,034	2,969	65.17%	34.83%	17.55%
Intimidation or force	335	47	382	87.70%	12.30%	2.26%
More than one victim	1,286	341	1,627	79.04%	20.96%	9.62%
Mitigating factors						
Age	2,873	1,514	4,387	65.49%	34.51%	25.93%
Genuine remorse	3,776	2,209	5,985	63.09%	36.91%	35.37%
Offender responding well to exiting order/sentence	491	136	627	78.31%	21.69%	3.71%
Offender can/is addressing needs/addiction	1,070	405	1,475	72.54%	27.46%	8.72%
Offender is main carer/has responsibilities	1,179	1,570	2,749	42.89%	57.11%	16.25%
Currently in, or prospects of, work/training	1,444	459	1,903	75.88%	24.12%	11.25%
Loss of job or reputation	1,524	894	2,418	63.03%	36.97%	14.29%
Physical or mental illness	1,101	849	1,950	56.46%	43.54%	11.52%
Difficult/deprived background	502	500	1,002	50.10%	49.90%	5.92%
Offence out of character	2,606	1,543	4,149	62.81%	37.19%	24.52%
Co-operation with authorities	1,337	741	2,078	64.34%	35.66%	12.28%
Voluntary return of stolen items	571	389	960	59.48%	40.52%	5.67%
Impact on sentence of offender's dependency	302	226	528	57.20%	42.80%	3.12%
Offender motivated by desperation or need	739	432	1171	63.11%	36.89%	6.92%

Guilty plea discount						
None	149	56	205	72.68%	27.32%	1.21%
1% - 10%	804	283	1087	73.97%	26.03%	6.42%
11% - 20%	730	236	966	75.57%	24.43%	5.71%
21% - 32%	1466	501	1967	74.53%	25.47%	11.63%
33% or more	6601	2787	9388	70.31%	29.69%	55.48%

Table 29: Offenders with Sole Responsibility for Dependent Relatives

Offence	Survey Question	Male Offenders	Female Offenders	χ^2test p-value	Male prisoners	Female prisoners	χ^2test p-value
Arson	Offender is main carer/has responsibilities	86	46	0.000	32	10	0.000
		3.00%	9.09%		3.01%	4.39%	
Assault	Sole/primary carer for dependent relatives	941	597	0.000	241	86	0.000
		2.06%	12.05%		1.06%	6.65%	
Burglary	Sole/primary carer for dependent relatives	228	65	0.000	103	19	0.000
		0.89%	5.32%		0.53%	2.73%	
Death	-	-	-	-	-	-	-
	-	-	-	-	-	-	-
Driving	Offender is main carer/has responsibilities	659	105	0.000	184	9	0.000
		8.79%	31.07%		5.00%	12.00%	
Drugs	Offender is main carer/has responsibilities	829	325	0.000	314	84	0.000
		2.44%	9.59%		1.67%	7.89%	
Other	Offender is main carer/has responsibilities	913	478	0.000	287	118	0.000
		5.68%	22.88%		3.23%	15.13%	
Robbery	Offender is main carer/has responsibilities	292	71	0.000	195	33	0.000
		2.53%	8.80%		1.94%	5.74%	
Sexual	Offender is main carer/has responsibilities	255	12	0.000	255	3	0.000
		3.80%	13.04%		5.55%	5.08%	
Theft and Fraud	Offender is main carer/has responsibilities	1179	1570	0.000	339	268	0.000
		9.69%	32.99%		5.60%	18.77%	

B.2 Generalised Ordered Logistic Regression Analysis

Table 30: Arson and Criminal Damage - Sentence Outcome

Variable	Outcome					
	Other		Community Order		Suspended Sentence	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.051	0.715	0.051	0.715	0.051	0.715
Age=35 to 44	0.175	0.278	0.175	0.278	0.175	0.278
Age=45 to 54	-0.091	0.622	-0.091	0.622	-0.091	0.622
Female	-0.555	0.000	-0.555	0.000	-0.555	0.000
Offence=Arson endangering life	0.774	0.093	-0.656	0.001	-1.394	0.000
Offence=Arson not endangering life	-1.461	0.000	-2.348	0.000	-2.917	0.000
Offence=Other	-2.761	0.000	-3.316	0.000	-3.316	0.000
Previous Convictions=1-3	1.222	0.000	0.626	0.000	0.828	0.000
Previous Convictions=4-9	1.330	0.000	1.330	0.000	1.330	0.000
Premeditation	1.262	0.000	1.262	0.000	1.262	0.000
Damage of High Value	0.885	0.000	0.885	0.000	0.885	0.000
Alcohol	0.576	0.000	0.576	0.000	0.576	0.000
Revenge	0.609	0.000	0.609	0.000	0.609	0.000
Fear Caused	0.962	0.000	0.962	0.000	0.962	0.000
Age	-0.147	0.278	-0.147	0.278	-0.147	0.278
Remorse	0.511	0.022	-0.152	0.306	-0.593	0.000
Addressing Addiction	0.924	0.008	-0.609	0.000	-1.387	0.000
Illness	-0.565	0.000	-0.565	0.000	-0.565	0.000
Difficult Background	1.737	0.005	-0.063	0.754	0.292	0.101
Out of Character	0.151	0.557	-0.220	0.214	-0.848	0.000
Cooperation	-0.292	0.067	-0.292	0.067	-0.292	0.067
GP Discount = 1% - 10%	1.593	0.000	1.593	0.000	1.593	0.000
GP Discount = 11% - 20%	1.933	0.000	1.933	0.000	1.933	0.000
GP Discount = 21% - 32%	1.756	0.000	1.756	0.000	1.756	0.000
GP Discount = 33% or more	1.675	0.000	1.675	0.000	1.675	0.000
Constant	1.135	0.005	0.821	0.031	0.211	0.568
Pseudo R squared	0.2938					
Obs	1985					

Table 31: Arson and Criminal Damage - Sentence Length

Variable	Length					
	Up to 1 year		1 - 3 years		3 - 5 years	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.325	0.032	0.325	0.032	0.325	0.032
Age=35 to 44	0.334	0.048	0.334	0.048	0.334	0.048
Age=45 to 54	0.461	0.020	0.461	0.020	0.461	0.020
Female	0.996	0.001	-0.211	0.329	0.186	0.625
Previous Convictions=1-3	-0.805	0.000	0.132	0.422	0.316	0.280
Previous Convictions=4-9	-0.893	0.000	-0.061	0.737	0.323	0.317
Premeditation	0.666	0.000	0.463	0.002	1.314	0.000
Damage of High Value	0.637	0.000	0.637	0.000	0.637	0.000
Alcohol	0.889	0.000	0.457	0.001	0.037	0.886
Revenge	0.538	0.000	0.538	0.000	0.538	0.000
Fear Caused	0.701	0.003	0.831	0.000	1.424	0.000
Age	0.211	0.163	0.211	0.163	0.211	0.163
Remorse	0.116	0.367	0.116	0.367	0.116	0.367
Addressing Addiction	-0.237	0.217	-0.237	0.217	-0.237	0.217
Illness	0.652	0.001	0.014	0.932	-0.465	0.146
Difficult Background	0.668	0.007	-0.087	0.665	0.159	0.655
Out of Character	0.089	0.576	0.089	0.576	0.089	0.576
Cooperation	-0.144	0.471	-0.144	0.471	-0.144	0.471
GP Discount = 1% - 10%	0.167	0.802	0.167	0.802	0.167	0.802
GP Discount = 11% - 20%	0.702	0.288	0.702	0.288	0.702	0.288
GP Discount = 21% - 32%	0.203	0.752	0.203	0.752	0.203	0.752
GP Discount = 33% or more	-0.134	0.832	-0.134	0.832	-0.134	0.832
Constant	0.306	0.637	-2.149	0.001	-4.649	0.000
Pseudo R squared	0.1229					
Obs	1255					

Table 32: Assault - Sentence Outcome

Variable	Outcome			
	Other		Suspended Sentence	
	Coef	P-value	Coef	P-value
Age=25 to 34	0.060	0.157	0.060	0.157
Age=35 to 44	0.219	0.006	-0.001	0.991
Age=45 to 54	-0.082	0.231	-0.082	0.231
Age = 54+	0.183	0.328	-0.415	0.006
Female	-0.571	0.000	-0.884	0.000
Offence = Assault	-1.307	0.000	-0.218	0.078
Offence = Cruelty/Neglect of a child	1.265	0.000	1.265	0.000
Offence = Harassment	-0.507	0.045	0.239	0.284
Offence = GBH	1.800	0.000	1.471	0.000
Offence = ABH	-0.018	0.871	-0.018	0.871
Offence = Violent Disorder	1.488	0.000	1.488	0.000
Seriousness = 2	-1.140	0.000	-0.758	0.000
Seriousness = 3	-2.277	0.000	-1.331	0.000
Previous Convictions=1-3	-0.856	0.202	1.084	0.019
Previous Convictions=4-9	-0.277	0.682	1.777	0.000
Previous Convictions = 10 or more	-0.075	0.915	2.299	0.000
GP Discount = 1% - 10%	0.302	0.097	-0.171	0.290
GP Discount = 11% - 20%	0.493	0.011	-0.085	0.604
GP Discount = 21% - 32%	0.398	0.022	-0.056	0.721
GP Discount = 33% or more	0.393	0.013	0.126	0.395
Injury	0.393	0.000	0.393	0.000
Vulnerable Victim	0.654	0.000	0.654	0.000
Repeated Assault	0.502	0.000	0.502	0.000
Injury Less Seriuos	-0.170	0.009	-0.002	0.975
Weapon	0.635	0.000	0.635	0.000
Lack of Premeditation	-0.309	0.000	-0.489	0.000
Previous Relevant Convictions	1.525	0.022	-0.400	0.386
Location	0.190	0.000	0.190	0.000
Timing	0.045	0.403	0.045	0.403
Ongoing Effect	0.533	0.000	0.533	0.000
Presence of Others	0.239	0.000	0.239	0.000
Alcohol	0.278	0.000	0.089	0.047
No Previous Convictions	-0.217	0.000	-0.217	0.000
Single Blow	-0.306	0.000	-0.306	0.000
Remorse	-0.204	0.001	-0.834	0.000
Character	-0.252	0.001	-0.600	0.000
Isolated Incident	-0.271	0.000	-0.610	0.000
Constant	1.868	0.000	-0.458	0.020
Pseudo R squared	0.2931			
Obs	15,996			

Table 33: Assault - Sentence Length

Variable	Length			
	Up to 1 year		1 - 3 years	
	Coef	P-value	Coef	P-value
Age=25 to 34	0.187	0.005	-0.036	0.649
Age=35 to 44	0.290	0.000	0.290	0.000
Age=45 to 54	0.333	0.001	0.333	0.001
Age = 54+	0.057	0.818	0.733	0.005
Female	-0.422	0.000	-0.422	0.000
Offence = Assault	-2.481	0.000	0.034	0.936
Offence = Cruelty/Neglect of a child	2.925	0.000	2.925	0.000
Offence = Harassment	-0.627	0.113	-0.627	0.113
Offence = GBH	2.985	0.000	3.953	0.000
Offence = ABH	-0.218	0.295	-0.218	0.295
Offence = Violent Disorder	1.906	0.000	1.906	0.000
Seriousness = 2	-1.826	0.000	-0.370	0.000
Seriousness = 3	-2.269	0.000	-1.133	0.000
Previous Convictions=1-3	-0.325	0.651	-0.769	0.285
Previous Convictions=4-9	0.226	0.754	-0.689	0.340
Previous Convictions = 10 or more	0.270	0.710	-0.825	0.260
GP Discount = 1% - 10%	-0.515	0.013	-0.515	0.013
GP Discount = 11% - 20%	-0.464	0.037	-0.960	0.000
GP Discount = 21% - 32%	-0.603	0.004	-0.931	0.000
GP Discount = 33% or more	-1.075	0.000	-1.075	0.000
Injury	0.314	0.000	-0.017	0.839
Vulnerable Victim	0.457	0.000	0.084	0.412
Repeated Assault	0.306	0.000	0.306	0.000
Injury Less Seriuos	0.004	0.965	0.436	0.000
Weapon	0.550	0.000	1.063	0.000
Lack of Premeditation	-0.299	0.000	-0.299	0.000
Previous Relevant Convictions	0.470	0.511	0.470	0.511
Location	0.080	0.204	0.080	0.204
Timing	0.121	0.112	-0.175	0.047
Ongoing Effect	0.280	0.000	0.531	0.000
Presence of Others	0.095	0.110	0.095	0.110
Alcohol	-0.030	0.570	-0.030	0.570
No Previous Convictions	-0.228	0.009	-0.228	0.009
Single Blow	-0.643	0.000	-0.643	0.000
Remorse	-0.019	0.740	-0.019	0.740
Character	-0.166	0.095	-0.166	0.095
Isolated Incident	-0.123	0.181	-0.123	0.181
Constant	0.438	0.149	-4.208	0.000
Pseudo R squared	0.3303			
Obs	8267			

Table 34: Burglary - Sentence Outcome

Variable	Outcome			
	Other		Suspended Sentence	
	Coef	P-value	Coef	P-value
Age=25 to 34	0.415	0.000	0.415	0.000
Age=35 to 44	0.249	0.000	0.249	0.000
Age=45 to 54	0.389	0.024	-0.030	0.776
Female	-0.793	0.000	-0.793	0.000
Offence = Domestic	-2.646	0.000	-2.646	0.000
Offence = Non-domestic	-3.512	0.000	-3.761	0.000
Offence = Other	-3.166	0.000	-3.166	0.000
Seriousness = 2	-1.168	0.000	-1.168	0.000
Seriousness = 3	-2.144	0.000	-1.500	0.000
Previous Convictions=1-3	0.951	0.000	0.951	0.000
Previous Convictions=4-9	1.171	0.000	1.516	0.000
Previous Convictions = 10 or more	1.311	0.000	1.714	0.000
GP Discount	-0.002	0.515	-0.002	0.515
Theft of High Value	0.617	0.000	0.251	0.000
Vandalism	0.361	0.000	0.361	0.000
Victim Returned	0.813	0.000	0.458	0.000
No Injury	-0.310	0.000	-0.310	0.000
No Violence	0.086	0.274	0.086	0.274
Low Value	-0.422	0.000	-0.422	0.000
Limited Damage	-0.236	0.000	-0.236	0.000
Deliberately Targetted	0.420	0.000	0.420	0.000
Significant Planning	0.868	0.000	0.469	0.000
Gang	0.065	0.265	0.065	0.265
Night	0.211	0.000	0.211	0.000
Alcohol	-0.017	0.792	-0.017	0.792
Failure to Comply	0.915	0.000	0.915	0.000
Remorse	1.387	0.000	1.387	0.000
Constant	5.536	0.000	3.701	0.000
Pseudo R squared	0.2337			
Obs	13,333			

Table 35: Burglary - Sentence Length

Variable	Length					
	Up to 1 year		1 - 3 years		3 - 5 years	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.297	0.000	0.585	0.000	0.712	0.000
Age=35 to 44	0.651	0.000	1.031	0.000	0.907	0.000
Age=45 to 54	0.502	0.000	0.976	0.000	1.065	0.000
Female	-0.619	0.000	-0.619	0.000	-0.619	0.000
Offence = Domestic	-3.529	0.000	-4.772	0.000	-5.532	0.000
Offence = Non-domestic	-5.499	0.000	-6.592	0.000	-7.078	0.000
Offence = Other	-4.034	0.000	-4.034	0.000	-4.034	0.000
Seriousness = 2	-2.457	0.000	-1.529	0.000	-1.887	0.000
Seriousness = 3	-2.966	0.000	-1.637	0.000	-1.984	0.006
Previous Convictions=1-3	0.581	0.000	0.179	0.070	0.480	0.009
Previous Convictions=4-9	1.040	0.000	1.040	0.000	1.040	0.000
Previous Convictions = 10 or more	1.228	0.000	1.228	0.000	1.228	0.000
GP Discount	-0.050	0.000	-0.019	0.000	-0.029	0.000
Theft of High Value	0.295	0.000	0.295	0.000	0.295	0.000
Vandalism	0.266	0.000	0.266	0.000	0.266	0.000
Victim Returned	0.398	0.000	0.398	0.000	0.398	0.000
No Injury	-0.361	0.000	-0.361	0.000	-0.361	0.000
No Violence	0.223	0.014	-0.252	0.039	-0.353	0.254
Low Value	-0.556	0.000	-0.556	0.000	-0.556	0.000
Limited Damage	-0.227	0.003	-0.227	0.003	-0.227	0.003
Deliberately Targetted	0.151	0.027	0.423	0.000	0.425	0.005
Significant Planning	0.310	0.000	0.597	0.000	0.908	0.000
Gang	-0.113	0.030	-0.113	0.030	-0.113	0.030
Night	0.181	0.000	0.181	0.000	0.181	0.000
Alcohol	-0.117	0.043	-0.117	0.043	-0.117	0.043
Failure to Comply	-0.210	0.006	-0.297	0.004	0.297	0.167
Remorse	-0.184	0.001	-0.184	0.001	-0.184	0.001
Constant	6.802	0.000	2.416	0.000	0.485	0.113
Pseudo R squared	0.2931					
Obs	10135					

Table 36: Death - Sentence Outcome

	Outcome			
	Other		Suspended Sentence	
	Coef	P-value	Coef	P-value
Age=25 to 34	-0.092	0.789	-0.092	0.789
Age=35 to 44	-0.227	0.520	-0.227	0.520
Age=45 to 54	-0.157	0.658	-0.157	0.658
Female	-1.002	0.003	-1.002	0.003
Seriousness = 2	-1.256	0.018	-2.363	0.000
Seriousness = 3	-3.037	0.000	-3.037	0.000
Previous Convictions=1-3	0.622	0.301	0.622	0.301
Previous Convictions=4-9	2.481	0.018	2.481	0.018
Premeditation	1.393	0.188	1.393	0.188
Vulnerable Victim	0.968	0.062	0.968	0.062
Suffering Inflicted	15.049	0.985	15.049	0.985
Concealment	3.073	0.004	3.073	0.004
Age	0.131	0.639	0.131	0.639
Remorse	-0.432	0.181	-0.432	0.181
Good Driving Record	-0.925	0.000	-0.925	0.000
Lack of Premeditation	-0.425	0.107	-0.425	0.107
Mental Illness	0.111	0.832	0.111	0.832
Effect on Offender	-0.197	0.449	-0.197	0.449
GP Discount = 11% - 20%	1.428	0.027	1.428	0.027
GP Discount = 21% - 32%	0.734	0.178	0.734	0.178
GP Discount = 33% or more	0.455	0.352	0.455	0.352
Constant	4.287	0.000	3.415	0.000
Pseudo R Squared	0.2877			
Obs	584			

Table 37: Death - Sentence Length

Variable	Length	
	5 years or more	
	Coef	P-value
Age=25 to 34	0.416	0.274
Age=35 to 44	-0.173	0.708
Age=45 to 54	-0.086	0.856
Female	-1.366	0.076
Seriousness = 2	-2.317	0.000
Seriousness = 3	-1.687	0.000
Previous Convictions=1-3	-1.455	0.009
Previous Convictions=4-9	0.340	0.351
Premeditation	2.577	0.000
Vulnerable Victim	0.160	0.725
Suffering Inflicted	2.609	0.010
Concealment	0.781	0.060
Age	-0.689	0.062
Remorse	-0.447	0.128
Good Driving Record	-1.391	0.000
Lack of Premeditation	0.622	0.092
Mental Illness	1.155	0.049
Effect on Offender	-0.430	0.256
GP Discount = 11% - 20%	-0.116	0.844
GP Discount = 21% - 32%	-0.496	0.373
GP Discount = 33% or more	-1.198	0.019
Constant	1.816	0.003
Pseudo R Squared	0.3707	
Obs	426	

Table 38: Driving - Sentence Outcome

	Outcome					
	Other		Community Order		Suspended Sentence	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	-0.047	0.525	-0.047	0.525	-0.047	0.525
Age=35 to 44	-0.050	0.836	0.035	0.791	-0.364	0.001
Age=45 to 54	-0.358	0.006	-0.358	0.006	-0.358	0.006
Age = 54+	-0.375	0.282	0.308	0.294	-0.369	0.167
Female	-0.758	0.000	-0.758	0.000	-0.758	0.000
Offence = Careless Driving	-5.455	0.000	-3.297	0.000	-2.446	0.000
Offence = Dangerous Driving	-0.053	0.892	0.490	0.000	-0.093	0.328
Offence = Other	-1.761	0.000	-0.425	0.005	-0.043	0.740
Previous Convictions=1-3	0.720	0.000	0.720	0.000	0.720	0.000
Previous Convictions=4-9	1.545	0.000	1.545	0.000	1.545	0.000
Previous Convictions=10 +	1.827	0.000	1.827	0.000	1.827	0.000
GP Discount	0.028	0.000	0.023	0.000	0.012	0.002
Alcohol	0.528	0.000	0.528	0.000	0.528	0.000
Disregard of Warnings	0.736	0.000	0.736	0.000	0.736	0.000
Aggressive Driving	0.884	0.000	0.884	0.000	0.884	0.000
Injury to Others	1.204	0.000	1.204	0.000	1.204	0.000
Damage	0.428	0.000	0.428	0.000	0.428	0.000
On Bail	0.197	0.596	0.540	0.001	0.953	0.000
Age	-0.305	0.000	-0.305	0.000	-0.305	0.000
Remorse	0.043	0.830	-0.295	0.002	-0.770	0.000
Prospect of Work	-0.074	0.725	-0.387	0.000	-1.333	0.000
Out of Character	-0.164	0.487	-0.203	0.074	-0.609	0.000
Good Driving Record	-0.873	0.000	-0.138	0.278	-0.270	0.050
Constant	2.867	0.000	0.050	0.776	-1.018	0.000
Pseudo R-squared	0.2764					
Obs	5533					

Table 39: Driving - Sentence Length

	Length			
	Up to 1 year		1 - 3 years	
	Coef	P-value	Coef	P-value
Age=25 to 34	-0.013	0.904	0.426	0.009
Age=35 to 44	-0.097	0.503	-0.097	0.503
Age=45 to 54	-0.029	0.897	-0.029	0.897
Age = 54+	0.035	0.938	0.035	0.938
Female	-0.622	0.082	-0.146	0.721
Offence = Careless Driving	-0.545	0.503	1.240	0.251
Offence = Dangerous Driving	-0.292	0.007	-0.292	0.007
Offence = Other	0.109	0.526	1.997	0.000
Previous Convictions=1-3	0.160	0.256	0.160	0.256
Previous Convictions=4-9	0.638	0.000	0.638	0.000
Previous Convictions=10 +	1.066	0.000	1.066	0.000
GP Discount	-0.031	0.000	-0.031	0.000
Alcohol	0.243	0.009	-0.034	0.821
Disregard of Warnings	0.150	0.127	0.150	0.127
Aggressive Driving	0.724	0.000	0.724	0.000
Injury to Others	0.864	0.000	1.341	0.000
Damage	0.189	0.041	0.189	0.041
On Bail	0.177	0.088	0.177	0.088
Age	-0.327	0.011	0.266	0.197
Remorse	0.091	0.428	0.091	0.428
Prospect of Work	-0.409	0.022	-1.080	0.000
Out of Character	-0.328	0.151	-0.328	0.151
Good Driving Record	-0.420	0.151	-0.420	0.151
Constant	-0.883	0.000	-3.440	0.000
Pseudo R Squared	0.1157			
Obs	2821			

Table 40: Drugs - Sentence Outcome

	Outcome					
	Other		Community Order		Suspended Sentence	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.152	0.000	0.152	0.000	0.152	0.000
Age=35 to 44	0.457	0.000	0.200	0.000	0.175	0.000
Age=45 to 54	0.219	0.000	0.219	0.000	0.219	0.000
Age = 54+	0.139	0.436	0.357	0.001	0.093	0.299
Female	0.088	0.427	-0.714	0.000	-1.021	0.000
Offence = Possession	2.760	0.000	0.735	0.000	-0.482	0.000
Offence = Production	1.821	0.000	0.414	0.000	-0.494	0.000
Offence = Supplying	2.059	0.000	0.473	0.000	-0.462	0.000
Previous Convictions=1-3	0.476	0.000	0.476	0.000	0.476	0.000
Previous Convictions=4-9	0.730	0.000	0.730	0.000	0.730	0.000
Previous Convictions=10 +	1.004	0.000	1.004	0.000	1.004	0.000
Cannabis	0.167	0.099	0.167	0.099	0.167	0.099
Class C	1.272	0.000	1.704	0.000	2.231	0.000
Other Class B	1.445	0.000	1.702	0.000	2.350	0.000
Other Class A	0.742	0.000	0.432	0.000	0.506	0.000
Cocaine	0.901	0.000	1.021	0.000	1.309	0.000
Heroin	-0.075	0.637	0.112	0.343	0.323	0.005
Drug Category = 1	-1.354	0.000	-1.354	0.000	-1.354	0.000
Drug Category = 2	-2.016	0.000	-2.016	0.000	-2.016	0.000
Drug Category = 3	-1.175	0.000	-1.255	0.000	-1.705	0.000
Previous Relevant Convictions	0.040	0.693	0.040	0.693	0.040	0.693
High Purity	1.715	0.000	1.715	0.000	1.715	0.000
Lack of Sophistication	-0.666	0.000	-0.693	0.000	-0.913	0.000
No Previous Relevant Convictions	0.095	0.320	0.027	0.555	-0.172	0.000
Remorse	0.437	0.000	0.042	0.381	-0.470	0.000
Good Character	-0.257	0.023	0.001	0.982	-0.253	0.000
Determination to Address	0.243	0.064	-0.469	0.000	-1.369	0.000
Medical Conditions	-0.500	0.001	-0.194	0.028	-1.007	0.000
Constant	1.352	0.000	0.120	0.324	-0.805	0.000
Pseudo R squared	0.2274					
Obs	27243					

Table 41: Drugs - Sentence Length

	Length					
	Up to 1 year		1 - 3 years		3 - 5 years	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.300	0.000	0.664	0.000	1.249	0.000
Age=35 to 44	0.406	0.000	0.888	0.000	1.456	0.000
Age=45 to 54	0.709	0.000	1.034	0.000	1.513	0.000
Age = 54+	0.574	0.000	1.378	0.000	1.685	0.000
Female	-0.711	0.000	-0.711	0.000	-0.711	0.000
Offence = Possession	-0.232	0.002	-1.477	0.000	-1.921	0.000
Offence = Production	-0.064	0.435	-1.741	0.000	-1.754	0.000
Offence = Supplying	-0.291	0.002	-1.526	0.000	-1.737	0.000
Previous Convictions=1-3	-0.225	0.114	0.021	0.879	0.074	0.619
Previous Convictions=4-9	-0.458	0.003	-0.123	0.421	-0.005	0.977
Previous Convictions=10 +	-0.535	0.003	-0.134	0.442	-0.181	0.413
Cannabis	0.712	0.001	1.544	0.000	1.694	0.000
Class C	4.025	0.000	4.025	0.000	4.025	0.000
Other Class B	4.074	0.000	4.074	0.000	4.074	0.000
Other Class A	2.059	0.000	3.215	0.000	3.416	0.000
Cocaine	3.228	0.000	3.736	0.000	3.977	0.000
Heroin	0.782	0.000	1.809	0.000	1.927	0.000
Drug Category = 1	-5.084	0.000	-5.084	0.000	-5.084	0.000
Drug Category = 2	-4.750	0.000	-4.750	0.000	-4.750	0.000
Drug Category = 3	-4.915	0.000	-4.915	0.000	-4.915	0.000
Previous Relevant Convictions	0.247	0.077	0.247	0.077	0.247	0.077
High Purity	0.936	0.000	1.189	0.000	1.242	0.000
Lack of Sophistication	-0.540	0.000	-0.931	0.000	-0.901	0.000
No Previous Relevant Convictions	0.105	0.137	-0.273	0.000	-0.288	0.002
Remorse	-0.042	0.564	-0.447	0.000	-0.650	0.000
Good Character	0.027	0.657	0.027	0.657	0.027	0.657
Determination to Address	0.106	0.374	-0.495	0.000	-0.774	0.000
Medical Conditions	-0.178	0.311	-0.219	0.179	-1.286	0.000
Constant	-0.671	0.004	-3.437	0.000	-5.467	0.000
Pseudo R squared	0.2347					
Obs	14,116					

Table 42: Other - Sentence Outcome

	Outcome					
	Other		Community Sentence		Suspended Sentence	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.209	0.086	0.209	0.086	0.209	0.086
Age=35 to 44	-0.511	0.040	0.046	0.763	-0.123	0.362
Age=45 to 54	-0.194	0.165	-0.194	0.165	-0.194	0.165
Age = 54+	-1.233	0.000	0.053	0.842	-0.330	0.135
Female	-0.350	0.037	-0.350	0.037	-0.350	0.037
Offence = Breach of protective order	0.341	0.152	-0.356	0.016	-0.242	0.059
Offence = Perverting the course of justice	2.860	0.005	1.718	0.000	0.617	0.018
Offence = Possession of offensive weapons	1.705	0.000	0.267	0.107	-0.690	0.000
Offence = Possession/distribution of prohibited weapons	0.754	0.000	0.754	0.000	0.754	0.000
Previous Convictions=1-3	0.841	0.000	0.841	0.000	0.841	0.000
Previous Convictions=4-9	1.390	0.000	1.390	0.000	1.390	0.000
Previous Convictions=10 +	2.763	0.000	1.221	0.000	1.587	0.000
Seriousness	-1.042	0.000	-0.761	0.000	-0.417	0.000
GP Discount	0.007	0.120	0.007	0.120	0.007	0.120
Premeditation	2.525	0.011	1.596	0.000	0.741	0.000
Intimidation	0.487	0.001	0.487	0.001	0.487	0.001
Vulnerable Victim	0.417	0.001	0.417	0.001	0.417	0.001
Age	-0.541	0.037	0.133	0.385	-0.224	0.092
Remorse	-0.108	0.631	-0.348	0.006	-0.864	0.000
Illness	-0.596	0.000	-0.596	0.000	-0.596	0.000
Out of Character	-0.271	0.066	-0.271	0.066	-0.271	0.066
Constant	3.396	0.005	1.324	0.015	1.004	0.023
Pseudo R squared	0.1681					
Obs	2521					

Table 43: Other - Sentence Length

	Length			
	Up to 1 year		1 to 3 years	
	Coef	P-value	Coef	P-value
Age=25 to 34	0.227	0.285	0.227	0.285
Age=35 to 44	0.079	0.742	0.079	0.742
Age=45 to 54	0.391	0.132	0.391	0.132
Age = 54+	0.077	0.837	0.077	0.837
Female	0.291	0.375	0.291	0.375
Offence = Breach of protective order	-0.832	0.000	-3.340	0.000
Offence = Perverting the course of justice	-0.095	0.816	-0.095	0.816
Offence = Possession of offensive weapons	-0.367	0.173	-2.113	0.005
Offence = Possession/distribution of prohibited weapons	1.906	0.000	1.906	0.000
Previous Convictions=1-3	-0.504	0.042	-0.504	0.042
Previous Convictions=4-9	0.218	0.404	0.218	0.404
Previous Convictions=10 +	-0.211	0.473	-2.313	0.004
Seriousness	-0.842	0.000	-0.842	0.000
GP Discount	-0.032	0.000	-0.032	0.000
Premeditation	1.049	0.000	1.049	0.000
Intimidation	0.773	0.000	0.773	0.000
Vulnerable Victim	0.592	0.001	0.592	0.001
Age	0.001	0.998	0.001	0.998
Remorse	-0.079	0.720	-0.079	0.720
Illness	-0.230	0.455	-0.230	0.455
Out of Character	0.028	0.939	-1.052	0.048
Constant	-0.922	0.230	-1.106	0.209
Pseudo R squared	0.2641			
Obs	1379			

Table 44: Robbery - Sentence Outcome

	Outcome			
	Other		Suspended Sentence	
	Coef	P-value	Coef	P-value
Age=25 to 34	0.433	0.000	0.433	0.000
Age=35 to 44	0.069	0.810	0.717	0.000
Age=45 to 54	0.237	0.386	0.237	0.386
Age = 54+	-0.707	0.366	-0.707	0.366
Female	-0.805	0.000	-0.805	0.000
Offence = Other	1.013	0.249	1.013	0.249
Offence = Robbery	0.760	0.014	0.760	0.014
Previous Convictions=1-3	0.813	0.000	0.813	0.000
Previous Convictions=4-9	1.363	0.000	1.363	0.000
Previous Convictions=10 +	1.441	0.000	1.441	0.000
Seriousness	-0.381	0.000	-0.381	0.000
GP Discount	0.005	0.271	0.005	0.271
Gang	0.309	0.001	0.309	0.001
Vulnerable Victim	0.880	0.000	0.880	0.000
Weapon	1.054	0.000	1.719	0.000
Alcohol	0.199	0.053	0.199	0.053
On Bail	1.366	0.000	1.366	0.000
Degree of Violence	1.515	0.000	1.515	0.000
Disguise	0.919	0.000	0.919	0.000
Value of Items	0.836	0.000	0.836	0.000
Multiple Victims	0.581	0.000	0.581	0.000
Night	0.570	0.000	0.570	0.000
Age	-0.589	0.000	-0.589	0.000
Remorse	-0.060	0.728	-0.823	0.000
Difficult Background	-0.425	0.000	-0.425	0.000
Out of Character	-0.307	0.099	-0.729	0.000
Unplanned	-0.228	0.200	-0.657	0.000
Constant	2.426	0.000	0.776	0.040
Pseudo R squared	0.3108			
Obs	7402			

Table 45: Robbery - Sentence Length

	Length					
	Up to 1 year		1 - 3 years		3 - 5 years	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.476	0.000	0.476	0.000	0.476	0.000
Age=35 to 44	0.717	0.000	0.717	0.000	0.717	0.000
Age=45 to 54	0.688	0.000	0.688	0.000	0.688	0.000
Age = 54+	-0.779	0.303	1.153	0.029	1.258	0.016
Female	-0.443	0.000	-0.443	0.000	-0.443	0.000
Offence = Other	-0.817	0.170	-0.817	0.170	-0.817	0.170
Offence = Robbery	0.449	0.225	0.222	0.429	-0.665	0.060
Previous Convictions=1-3	0.162	0.020	0.162	0.020	0.162	0.020
Previous Convictions=4-9	0.486	0.000	0.486	0.000	0.486	0.000
Previous Convictions=10 +	0.559	0.000	0.559	0.000	0.559	0.000
Seriousness	-0.838	0.000	-0.811	0.000	-1.120	0.000
GP Discount	-0.025	0.000	-0.036	0.000	-0.046	0.000
Gang	0.227	0.000	0.227	0.000	0.227	0.000
Vulnerable Victim	0.318	0.000	0.318	0.000	0.318	0.000
Weapon	2.163	0.000	1.452	0.000	1.170	0.000
Alcohol	0.211	0.056	-0.140	0.045	-0.458	0.000
On Bail	0.445	0.000	0.445	0.000	0.445	0.000
Degree of Violence	1.122	0.000	0.600	0.000	0.919	0.000
Disguise	0.966	0.000	0.966	0.000	0.966	0.000
Value of Items	0.848	0.000	0.848	0.000	0.848	0.000
Multiple Victims	0.838	0.000	0.838	0.000	0.838	0.000
Night	0.190	0.070	0.158	0.017	-0.118	0.215
Age	-0.138	0.179	-0.574	0.000	-0.381	0.000
Remorse	-0.292	0.000	-0.292	0.000	-0.292	0.000
Difficult Background	-0.235	0.004	-0.235	0.004	-0.235	0.004
Out of Character	-0.357	0.000	-0.357	0.000	-0.357	0.000
Unplanned	-0.470	0.000	-0.825	0.000	-1.267	0.000
Constant	2.979	0.000	0.406	0.224	0.090	0.838
Pseudo R squared	0.2248					
Obs	6361					

Table 46: Sexual - Sentence Outcome

	Outcome			
	Other		Suspended Sentence	
	Coef	P-value	Coef	P-value
Age=25 to 34	0.245	0.047	0.245	0.047
Age=35 to 44	0.286	0.030	0.286	0.030
Age=45 to 54	0.584	0.000	0.584	0.000
Age = 54+	0.488	0.000	0.488	0.000
Female	0.415	0.164	0.415	0.164
Offence = Rape	4.237	0.000	4.237	0.000
Offence = Sexual Activity with a Child	0.944	0.000	0.944	0.000
Offence = Sexual Assault	0.159	0.138	0.159	0.138
Offence = Indecent Photos of Children	0.022	0.834	0.022	0.834
Previous Convictions=1-3	0.296	0.011	0.296	0.011
Previous Convictions=4-9	1.268	0.000	1.268	0.000
Previous Convictions=10 +	0.859	0.029	0.859	0.029
Seriousness	-0.523	0.000	-0.523	0.000
GP Discount	0.015	0.000	0.015	0.000
Premeditation	0.696	0.000	1.018	0.000
Repeated Assault	0.718	0.000	1.219	0.000
Alcohol	0.225	0.103	0.225	0.103
Abuse of Power	0.758	0.000	0.758	0.000
Vulnerable Victim	0.718	0.000	0.718	0.000
Multiple Victims	-0.039	0.750	-0.039	0.750
Age	-0.069	0.521	-0.282	0.004
Remorse	-0.293	0.005	-0.641	0.000
Loss of Job	0.042	0.684	0.042	0.684
Out of Character	-0.463	0.000	-0.463	0.000
Cooperation	-0.389	0.001	-0.670	0.000
Constant	1.705	0.000	0.718	0.000
Pseudo r squared	0.2502			
Obs	4017			

Table 47: Sexual - Sentence Length

	Length					
	Up to 1 year		1 to 3 years		3 to 5 years	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.093	0.501	0.093	0.501	0.093	0.501
Age=35 to 44	-0.099	0.599	0.607	0.001	0.384	0.073
Age=45 to 54	0.013	0.946	0.657	0.000	0.309	0.156
Age = 54+	-0.031	0.865	0.695	0.000	0.573	0.007
Female	-0.505	0.138	-0.505	0.138	-0.505	0.138
Offence = Rape	3.126	0.000	4.140	0.000	3.795	0.000
Offence = Sexual Activity with a Child	1.017	0.000	0.667	0.000	-0.285	0.145
Offence = Sexual Assault	-0.050	0.674	-0.050	0.674	-0.050	0.674
Offence = Indecent Photos of Children	-0.218	0.160	-0.218	0.160	-0.218	0.160
Previous Convictions=1-3	0.117	0.309	0.117	0.309	0.117	0.309
Previous Convictions=4-9	-0.114	0.588	-0.114	0.588	-0.114	0.588
Previous Convictions=10 +	0.013	0.968	0.013	0.968	0.013	0.968
Seriousness	-0.585	0.000	-0.800	0.000	-0.928	0.000
GP Discount	-0.015	0.000	-0.015	0.000	-0.015	0.000
Premeditation	0.378	0.000	0.378	0.000	0.378	0.000
Repeated Assault	0.698	0.000	1.103	0.000	1.187	0.000
Alcohol	-0.157	0.394	0.201	0.241	0.567	0.004
Abuse of Power	0.750	0.000	0.750	0.000	0.750	0.000
Vulnerable Victim	0.393	0.000	0.393	0.000	0.393	0.000
Multiple Victims	0.133	0.433	0.846	0.000	0.844	0.000
Age	-0.254	0.016	-0.254	0.016	-0.254	0.016
Remorse	-0.006	0.955	-0.006	0.955	-0.006	0.955
Loss of Job	-0.304	0.060	-0.837	0.000	-0.587	0.007
Out of Character	-0.650	0.000	-0.650	0.000	-0.650	0.000
Cooperation	0.041	0.765	0.041	0.765	0.041	0.765
Constant	2.527	0.000	-0.457	0.060	-1.454	0.000
Pseudo R squared	0.3090					
Obs	2470					

Table 48: Theft and Fraud - Sentence Outcome

	Outcome					
	Other		Community Order		Suspended Sentence	
	Coef	P-value	Coef	P-value	Coef	P-value
Age=25 to 34	0.130	0.412	0.325	0.000	0.151	0.028
Age=35 to 44	0.293	0.104	0.453	0.000	0.174	0.020
Age=45 to 54	0.139	0.479	0.533	0.000	0.223	0.006
Age = 54+	-0.464	0.034	0.759	0.000	0.154	0.131
Female	-0.149	0.261	-0.259	0.000	-0.575	0.000
Offence = Dishonest representation	0.662	0.001	0.504	0.000	-0.518	0.000
Offence = Receiving stolen goods	-0.328	0.003	-0.328	0.003	-0.328	0.003
Offence = Theft from person	-0.924	0.000	-0.669	0.000	-0.155	0.055
Offence = Theft from shops and stalls	-1.654	0.000	-1.053	0.000	-0.354	0.000
Offence = Theft in breach of trust	0.807	0.000	0.613	0.000	-0.015	0.828
Offence = With intent knowingly possess false	2.021	0.000	2.021	0.000	2.021	0.000
Previous Convictions=1-3	0.496	0.000	0.496	0.000	0.496	0.000
Previous Convictions=4-9	1.010	0.000	1.010	0.000	1.010	0.000
Previous Convictions=10 +	2.130	0.000	1.371	0.000	1.486	0.000
Seriousness	-0.853	0.000	-0.782	0.000	-0.530	0.000
GP Discount	0.020	0.000	0.020	0.000	0.020	0.000
Premeditation	0.733	0.000	0.733	0.000	0.733	0.000
Gang	0.980	0.001	0.057	0.516	0.161	0.020
Vulnerable Victims	1.180	0.000	1.180	0.000	1.180	0.000
High Value	0.916	0.000	0.916	0.000	0.916	0.000
High Gain	1.233	0.000	1.233	0.000	1.233	0.000
Age	-0.163	0.001	-0.163	0.001	-0.163	0.001
Remorse	0.172	0.181	-0.222	0.000	-0.618	0.000
Main Carer of Children	-0.070	0.675	0.069	0.372	-0.815	0.000
In Work	-0.214	0.200	-0.353	0.000	-1.670	0.000
Loss of Job	0.084	0.169	0.084	0.169	0.084	0.169
Illness	-0.147	0.380	-0.131	0.120	-0.592	0.000
Out of Character	-0.609	0.000	-0.115	0.093	-0.430	0.000
Cooperation	-0.386	0.000	-0.386	0.000	-0.386	0.000
Constant	5.672	0.000	2.702	0.000	0.506	0.000
Pseudo R squared	0.2764					
Obs	12,383					

Table 49: Theft and Fraud - Sentence Length

	Length			
	Up to 1 year		1 to 3 years	
	Coef	P-value	Coef	P-value
Age=25 to 34	0.670	0.000	0.670	0.000
Age=35 to 44	1.075	0.000	1.075	0.000
Age=45 to 54	1.199	0.000	1.199	0.000
Age = 54+	1.327	0.000	1.327	0.000
Female	-0.511	0.000	-0.511	0.000
Offence = Dishonest representation	-1.169	0.000	-1.169	0.000
Offence = Receiving stolen goods	-0.489	0.023	-0.489	0.023
Offence = Theft from person	-0.863	0.000	-1.727	0.000
Offence = Theft from shops and stalls	-2.007	0.000	-2.007	0.000
Offence = Theft in breach of trust	-0.144	0.133	-1.086	0.000
Offence = With intent knowingly possess false	-0.852	0.000	-0.852	0.000
Previous Convictions=1-3	-0.053	0.551	-0.053	0.551
Previous Convictions=4-9	-0.061	0.587	-0.061	0.587
Previous Convictions=10 +	0.039	0.719	0.039	0.719
Seriousness	-0.809	0.000	-1.119	0.000
GP Discount	-0.023	0.000	-0.023	0.000
Premeditation	0.405	0.000	0.405	0.000
Gang	0.052	0.518	0.052	0.518
Vulnerable Victims	0.492	0.000	0.492	0.000
High Value	0.879	0.000	0.879	0.000
High Gain	0.838	0.000	0.838	0.000
Age	-0.065	0.466	-0.065	0.466
Remorse	-0.004	0.962	-0.528	0.001
Main Carer of Children	-0.394	0.001	-0.394	0.001
In Work	-0.673	0.000	-0.673	0.000
Loss of Job	0.010	0.926	0.010	0.926
Illness	-0.346	0.006	-0.346	0.006
Out of Character	-0.264	0.009	-0.264	0.009
Cooperation	-0.030	0.795	-0.030	0.795
Constant	1.313	0.000	-1.089	0.000
Pseudo R squared	0.2651			
Obs	5609			

B.3 Marginal Effects

Table 50: Marginal Effects for Sentencing Outcomes

(a) Arson and Criminal Damage

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.067	0.005	0.000	0.058	0.077	10.430***	0.001
1#Female	0.099	0.011	0.000	0.078	0.120		
2#Male	0.140	0.007	0.000	0.126	0.154	13.580***	0.000
2#Female	0.174	0.011	0.000	0.153	0.195		
3#Male	0.165	0.008	0.000	0.150	0.180	12.340***	0.000
3#Female	0.177	0.008	0.000	0.160	0.193		
4#Male	0.627	0.009	0.000	0.610	0.645	13.020***	0.000
4#Female	0.550	0.020	0.000	0.511	0.589		

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.129	0.002	0.000	0.124	0.133	44.540***	0.000
1#Female	0.179	0.007	0.000	0.165	0.192		
2#Male	0.339	0.004	0.000	0.332	0.346	57.410***	0.000
2#Female	0.427	0.011	0.000	0.406	0.449		
3#Male	0.532	0.003	0.000	0.526	0.539	153.840***	0.000
3#Female	0.394	0.011	0.000	0.373	0.415		

(b) Assault

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.066	0.002	0.000	0.063	0.070	43.380***	0.000
1#Female	0.117	0.008	0.000	0.102	0.133		
2#Male	0.168	0.003	0.000	0.162	0.174	63.930***	0.000
2#Female	0.231	0.008	0.000	0.214	0.247		
3#Male	0.766	0.003	0.000	0.760	0.772	54.620***	0.000
3#Female	0.652	0.015	0.000	0.622	0.682		

(c) Burglary

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.097	0.011	0.000	0.075	0.119	6.850***	0.009
1#Female	0.181	0.032	0.000	0.119	0.243		
2#Male	0.132	0.013	0.000	0.106	0.157	9.520***	0.002
2#Female	0.179	0.021	0.000	0.139	0.220		
3#Male	0.771	0.015	0.000	0.741	0.801	8.290***	0.004
3#Female	0.640	0.043	0.000	0.555	0.724		

(d) Death

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.043	0.002	0.000	0.040	0.047	17.670***	0.000
1#Female	0.064	0.005	0.000	0.053	0.074		
2#Male	0.115	0.004	0.000	0.107	0.122	22.830***	0.000
2#Female	0.184	0.015	0.000	0.155	0.213		
3#Male	0.322	0.006	0.000	0.311	0.333	68.150***	0.000
3#Female	0.352	0.007	0.000	0.338	0.366		
4#Male	0.520	0.005	0.000	0.510	0.531	29.920***	0.000
4#Female	0.401	0.022	0.000	0.358	0.443		

(e) Driving

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.036	0.001	0.000	0.034	0.038	0.660	0.417
1#Female	0.034	0.003	0.000	0.028	0.039		
2#Male	0.126	0.002	0.000	0.122	0.130	150.860***	0.000
2#Female	0.222	0.008	0.000	0.207	0.237		
3#Male	0.301	0.003	0.000	0.296	0.306	72.400***	0.000
3#Female	0.379	0.009	0.000	0.362	0.397		
4#Male	0.537	0.003	0.000	0.532	0.542	380.180***	0.000
4#Female	0.365	0.008	0.000	0.349	0.382		

(f) Drugs

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.041	0.004	0.000	0.034	0.049	3.540*	0.060
1#Female	0.055	0.008	0.000	0.039	0.071		
2#Male	0.146	0.007	0.000	0.133	0.160	4.000**	0.046
2#Female	0.180	0.018	0.000	0.146	0.214		
3#Male	0.256	0.008	0.000	0.240	0.272	5.410**	0.020
3#Female	0.276	0.012	0.000	0.252	0.299		
4#Male	0.556	0.009	0.000	0.539	0.574	4.300**	0.038
4#Female	0.489	0.031	0.000	0.428	0.551		

(g) Other

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.021	0.002	0.000	0.018	0.024	21.770***	0.000
1#Female	0.042	0.005	0.000	0.033	0.052		
2#Male	0.096	0.003	0.000	0.090	0.102	31.930***	0.000
2#Female	0.142	0.008	0.000	0.125	0.158		
3#Male	0.883	0.003	0.000	0.877	0.890	29.330***	0.000
3#Female	0.816	0.012	0.000	0.792	0.839		

(h) Robbery

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.180	0.005	0.000	0.169	0.190	2.330	0.127
1#Female	0.135	0.029	0.000	0.079	0.192		
2#Male	0.162	0.005	0.000	0.151	0.172	1.400	0.237
2#Female	0.148	0.012	0.000	0.124	0.172		
3#Male	0.659	0.006	0.000	0.647	0.670	2.050	0.152
3#Female	0.717	0.040	0.000	0.638	0.796		

(i) Sexual

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.028	0.002	0.000	0.025	0.031	1.200	0.272
1#Female	0.032	0.003	0.000	0.026	0.037		
2#Male	0.144	0.003	0.000	0.137	0.150	12.500***	0.000
2#Female	0.169	0.006	0.000	0.157	0.180		
3#Male	0.350	0.004	0.000	0.341	0.359	39.240***	0.000
3#Female	0.409	0.008	0.000	0.394	0.425		
4#Male	0.479	0.004	0.000	0.470	0.487	105.570***	0.000
4#Female	0.391	0.007	0.000	0.377	0.405		

(j) Theft and Fraud

Table 51: Marginal Effects for Sentence Lengths

(a) Arson and Criminal Damage

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.248	0.012	0.000	0.225	0.271	15.520***	0.000
1#Female	0.123	0.029	0.000	0.065	0.180		
2#Male	0.478	0.015	0.000	0.449	0.506	13.550***	0.000
2#Female	0.639	0.041	0.000	0.559	0.719		
3#Male	0.217	0.012	0.000	0.193	0.241	1.940	0.164
3#Female	0.171	0.030	0.000	0.112	0.230		
4#Male	0.057	0.007	0.000	0.044	0.070	0.220	0.642
4#Female	0.067	0.020	0.001	0.028	0.106		

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.403	0.004	0.000	0.395	0.411	13.880***	0.000
1#Female	0.462	0.016	0.000	0.431	0.492		
2#Male	0.437	0.005	0.000	0.427	0.446	10.520***	0.001
2#Female	0.415	0.008	0.000	0.399	0.430		
3#Male	0.160	0.003	0.000	0.154	0.167	16.440***	0.000
3#Female	0.123	0.009	0.000	0.106	0.141		

(b) Assault

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.281	0.004	0.000	0.274	0.288	24.180***	0.000
1#Female	0.367	0.017	0.000	0.333	0.402		
2#Male	0.553	0.004	0.000	0.544	0.562	15.790***	0.000
2#Female	0.517	0.010	0.000	0.497	0.536		
3#Male	0.129	0.003	0.000	0.123	0.135	33.760***	0.000
3#Female	0.090	0.007	0.000	0.076	0.103		
4#Male	0.037	0.001	0.000	0.034	0.039	33.030***	0.000
4#Female	0.027	0.002	0.000	0.023	0.030		

(c) Burglary

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.595	0.019	0.000	0.559	0.632	4.230**	0.040
1#Female	0.764	0.080	0.000	0.608	0.921		
2#Male	0.405	0.019	0.000	0.368	0.441	4.230**	0.040
2#Female	0.236	0.080	0.003	0.079	0.392		

(d) Death

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.700	0.008	0.000	0.684	0.716	3.990**	0.046
1#Female	0.802	0.051	0.000	0.703	0.902		
2#Male	0.228	0.008	0.000	0.213	0.242	7.130***	0.008
2#Female	0.133	0.035	0.000	0.065	0.202		
3#Male	0.073	0.005	0.000	0.064	0.082	0.140	0.708
3#Female	0.065	0.022	0.003	0.022	0.107		

(e) Driving

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.212	0.003	0.000	0.206	0.217	68.750***	0.000
1#Female	0.300	0.011	0.000	0.279	0.321		
2#Male	0.475	0.004	0.000	0.468	0.483	34.670***	0.000
2#Female	0.485	0.004	0.000	0.477	0.493		
3#Male	0.198	0.003	0.000	0.192	0.205	82.350***	0.000
3#Female	0.145	0.006	0.000	0.133	0.157		
4#Male	0.115	0.002	0.000	0.110	0.119	106.190***	0.000
4#Female	0.070	0.004	0.000	0.062	0.079		

(f) Drugs

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.778	0.010	0.000	0.759	0.798	0.720	0.396
1#Female	0.740	0.044	0.000	0.654	0.826		
2#Male	0.171	0.009	0.000	0.152	0.189	0.720	0.395
2#Female	0.199	0.034	0.000	0.133	0.265		
3#Male	0.051	0.005	0.000	0.042	0.060	0.710	0.400
3#Female	0.061	0.012	0.000	0.037	0.084		

(g) Other

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.094	0.003	0.000	0.088	0.101	12.640***	0.000
1#Female	0.129	0.010	0.000	0.109	0.149		
2#Male	0.499	0.006	0.000	0.488	0.510	21.330***	0.000
2#Female	0.533	0.009	0.000	0.515	0.551		
3#Male	0.279	0.005	0.000	0.268	0.289	14.640***	0.000
3#Female	0.240	0.010	0.000	0.220	0.261		
4#Male	0.129	0.004	0.000	0.122	0.135	18.600***	0.000
4#Female	0.097	0.007	0.000	0.083	0.112		

(h) Robbery

	Margin	S.E	P Value	95% Conf.	Interval	Difference χ^2	P Value
1#Male	0.196	0.007	0.000	0.182	0.209	1.940	0.164
1#Female	0.261	0.047	0.000	0.169	0.352		
2#Male	0.407	0.009	0.000	0.389	0.424	0.350	0.553
2#Female	0.401	0.012	0.000	0.377	0.426		
3#Male	0.162	0.007	0.000	0.148	0.175	2.180	0.140
3#Female	0.141	0.015	0.000	0.111	0.171		
4#Male	0.236	0.006	0.000	0.224	0.247	2.600	0.107
4#Female	0.197	0.024	0.000	0.149	0.245		

(i) Sexual

	Margin	S.E	P Value	95% Conf. Interval		Difference χ^2	P Value
1#Male	0.570	0.006	0.000	0.558	0.581	37.360***	0.000
1#Female	0.650	0.012	0.000	0.627	0.673		
2#Male	0.368	0.006	0.000	0.356	0.380	34.950***	0.000
2#Female	0.308	0.010	0.000	0.289	0.327		
3#Male	0.062	0.003	0.000	0.056	0.068	41.730***	0.000
3#Female	0.042	0.003	0.000	0.036	0.048		

(j) Theft and Fraud

B.4 Interations

Table 52: Interaction Terms for Arson and Criminal Damage

	Coef	P-value
Premeditation#Gender	-0.311	0.543
Damage of High Value#Gender	-0.458	0.293
Alcohol#Gender	0.072	0.819
Revenge#Gender	0.392	0.432
Fear Caused#Gender	-0.296	0.635
Age#Gender	0.134	0.684
Remorse#Gender	0.173	0.604
Addressing Addiction#Gender	-0.297	0.381
Illness#Gender	0.196	0.521
Difficult Background#Gender	-0.199	0.577
Out of Character#Gender	0.610*	0.060
Cooperation#Gender	-0.486	0.211

C THE SEVERITY OF ROAD TRAFFIC ACCIDENTS

C.1 Variable Descriptions

Table 53: Variable Descriptions

Variable	Frequency	Percentage	Variable	Frequency	Percentage
Vehicles			Casualties		
1	19,240	25.88	1	53,621	72.12
2	38,232	51.42	2	13,468	18.12
3	12,284	16.52	3	4,552	6.12
4	3,387	4.56	4	1,651	2.22
5	745	1.00	5	600	0.81
6	281	0.38	6	298	0.40
7	122	0.16	7	85	0.11
8	35	0.05	8	39	0.05
9	10	0.01	9	3	0.00
10	11	0.01	10	1	0.00
Driver Sex			11	9	0.01
Male	49,328	67.66	12	4	0.01
Female	23,582	32.34	16	7	0.01
Driver Age			17	4	0.01
17-20	8,691	13.45	22	2	0.00
21-29	13,344	20.64	27	1	0.00
30-39	12,358	19.12	42	2	0.00
40-49	12,312	19.05	Pedestrians		
50-59	8,372	12.95	0	66,723	89.75
60-69	5,431	8.40	1	7,339	9.87
70+	4,132	6.39	2	245	0.33
Breath Test			3	20	0.03
Negative	49,937	96.99	4	3	0.00
Positive	1,550	3.01	5	16	0.02
Hit and run			26	1	0.00
No	70,701	95.47	Cycles		
Yes	3,358	4.53	0	66,863	89.93
Seatbelt			1	7,276	9.79
Yes	11,204	96.85	2	152	0.20
No	364	3.15	3	5	0.01
UK Licence			6	51	0.07
Full	18,913	80.32	OAPs		
Provisional	917	3.89	0	61,841	83.18
Unlicenced	3,716	15.78	1	10,513	14.14
Ethnicity			2	1,646	2.21
White	20,017	97.20	3	242	0.33

Asian	206	1.00
Black	210	1.02
Mixed Background	90	0.44
Oriental	70	0.34
Severity		
Slight	62,135	83.57
Serious	10,695	14.39
Fatal	1,517	2.04
Road Condition		
Dry	48,276	64.93
Wet/Damp	23,192	31.19
Snow	730	0.98
Frost/Ice	2,040	2.74
Flood	109	0.15
Visibility		
Daylight	45,092	72.51
Darkness - lights lit	8,662	13.93
Darkness - no lighting	8,434	13.56

4	82	0.11
5	10	0.01
6	4	0.01
9	2	0.00
12	5	0.01
42	2	0.00
Weekday		
Sunday	7,891	10.61
Monday	10,927	14.70
Tuesday	10,913	14.68
Wednesday	11,183	15.04
Thursday	10,904	14.67
Friday	12,495	16.81
Saturday	10,034	13.50
Road Type		
Roundabout	4,778	6.43
One Way Street	1,811	2.44
Dual Carriageway	6,259	8.43
Single Carriageway	60,896	81.98
Slip Road	534	0.72
Speed Limit		
20	768	1.03
30	35,332	47.52
40	6,381	8.58
50	3,119	4.20
60	23,853	32.08
70	4,894	6.58

C.2 χ^2 Significance Test for Simulated Poisson distribution

Table 54: Observed and Expected Accident Occurrences

No. of Accidents (X_i)	Observed (O_i)	$P(X = x)$	Expected (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
1	0	0.00	0.17	-0.17	0.03	0.17
2	4	0.00	0.83	3.17	10.06	12.14
3	2	0.01	2.76	-0.76	0.58	0.21
4	9	0.02	6.90	2.10	4.39	0.64
5	17	0.04	13.81	3.19	10.18	0.74
6	30	0.06	23.02	6.98	48.79	2.12
7	27	0.09	32.88	-5.88	34.56	1.05
8	44	0.11	41.10	2.90	8.42	0.20
9	46	0.13	45.67	0.33	0.11	0.00
10	36	0.13	45.67	-9.67	93.42	2.05
11	43	0.11	41.51	1.49	2.21	0.05
12	33	0.09	34.59	-1.59	2.54	0.07
13	22	0.07	26.61	-4.61	21.27	0.80
14	11	0.05	19.01	-8.01	64.13	3.37
15	12	0.03	12.67	-0.67	0.45	0.04
16	10	0.02	7.92	2.08	4.33	0.55
17	4	0.01	4.66	-0.66	0.43	0.09
18	5	0.01	2.59	2.41	5.82	2.25
19	4	0.00	1.36	2.64	6.96	5.11
20	1	0.00	0.68	0.32	0.10	0.15
21	0	0.00	0.32	-0.32	0.11	0.32
22	1	0.00	0.15	0.85	0.73	4.93
23	4	0.00	0.06	3.94	15.49	241.68
	Total	1	365			278.72

- H_0 : Number of daily accidents ˜ Poisson
- H_1 : Number of daily accidents is not ˜ Poisson
- Significance level: $\alpha = 0.05$
- Degrees of freedom: $23 - 2 = 21$
- Critical value: 32.671
- $\chi^2 = \frac{(O_i-E_i)^2}{E_i} = 278.72 > 32.671$

The null hypothesis is rejected. However, this result seems to be driven by slightly larger then expected deviations in the extremes. When $X_i = 1, 2, 23$ are excluded, the null hypothesis is no longer rejected.

- Degrees of freedom: $20 - 2 = 18$
- Critical value: 28.869
- $\chi^2 = \frac{(O_i - E_i)^2}{E_i} = 24.74 < 28.869$

Therefore, the distribution of daily accidents can be approximated reasonably well by the Poisson distribution.

C.3 Ordered Logistic Regression Analysis

Table 55: Ologit Results

	Coef.	Std. Err.	P>|z|	[95% Conf.	Interval]
Vehicles	-0.012	0.040	0.760	-0.092	0.067
Driver Sex = Female	-0.396***	0.093	0.000	-0.579	-0.213
Driver Age	0.003	0.003	0.225	-0.002	0.008
Breath Test = Positive	1.031***	0.178	0.000	0.682	1.379
Hit and Run	-0.623	0.559	0.265	-1.717	0.472
Seatbelt = No Seatbelt	1.477***	0.198	0.000	1.089	1.865
UK Licence = Provisional	0.012	0.216	0.957	-0.412	0.436
UK Licence = Unlicensed	-0.633**	0.253	0.012	-1.129	-0.137
Ethnicity = Asian	0.530	0.361	0.141	-0.177	1.237
Ethnicity = Black	0.270	0.347	0.437	-0.410	0.950
Ethnicity = Mixed Background	1.275**	0.543	0.019	0.211	2.340
Ethnicity = Oriental	-0.608	0.778	0.435	-2.133	0.917
Road Condition = Wet/Damp	-0.025	0.106	0.815	-0.232	0.182
Road Condition = Snow	-0.158	0.576	0.784	-1.287	0.972
Road Condition = Frost/Ice	-0.619**	0.289	0.032	-1.185	-0.052
Road Condition = Flood	0.595	0.540	0.271	-0.464	1.653
Visibility = Darkness - Lights lit	0.299*	0.167	0.073	-0.028	0.626
Visibility = Darkness - no lighting	0.383***	0.103	0.000	0.181	0.584
Casualties	0.402***	0.034	0.000	0.335	0.469
Road Type = One Way Street	-0.150	0.676	0.824	-1.475	1.174
Road Type = Dual Carriageway	-0.045	0.375	0.905	-0.780	0.690
Road Type = Single Carriageway	0.549	0.361	0.129	-0.159	1.257
Road Type = Slip Road	-2.053**	0.839	0.014	-3.697	-0.409
Speed Limit	0.021***	0.004	0.000	0.013	0.029
Weather = Raining without high winds	-0.163	0.147	0.266	-0.451	0.124
Weather = Snowing without high winds	0.277	0.503	0.581	-0.709	1.263
Weather = Fine with high winds	0.445	0.272	0.102	-0.088	0.977
Weather = Raining with high winds	-0.958**	0.399	0.016	-1.740	-0.176
Weather = Snowing with high winds	0.012	0.896	0.990	-1.744	1.767
Weather = Fog or Mist	-0.324	0.337	0.337	-0.984	0.337
Weekday	0.303**	0.136	0.026	0.036	0.570
OAPs	0.447***	0.069	0.000	0.312	0.582
Pedestrians	1.246***	0.314	0.000	0.630	1.861
Cycles	-0.352	0.282	0.212	-0.904	0.201
Constant					
Observations	5201				
Pseudo R squared	0.0959				

Table 56: Likelihood Ratio Test

	Coef.	Std. Err.	P-value
Vehicles	-0.028	0.040	0.488
Driver Sex	-0.402***	0.092	0.000
Driver Age	0.003	0.003	0.228
Breath Test	1.032***	0.175	0.000
Hit and Run	-0.645	0.548	0.240
Seat Belt	1.416***	0.194	0.000
UK Licence	-0.233**	0.107	0.030
Ethnicity	0.149	0.095	0.118
Road Condition	-0.085	0.064	0.183
Visibility	0.174***	0.050	0.001
Casualties	0.397***	0.034	0.000
Road Type	0.208***	0.079	0.009
Speed Limit	0.014***	0.003	0.000
Weather	-0.051	0.041	0.217
Weekday	0.272**	0.135	0.044
OAPs	0.453***	0.068	0.000
Pedestrians	0.993***	0.299	0.001
Cycles	-0.258	0.249	0.300
Observations	5201.000		

χ^2: 34.71

p-value: 0.0151

Table 57: Brant Test of Parallel Regression Assumption

Variable	χ^2	p-value	df
All	30.77**	0.014	16
Vehicles	3.36*	0.067	1
Driver Sex	2.84*	0.092	1
Driver Age	2.18	0.140	1
Breath Test	5.42**	0.020	1
Hit and Run	1.04	0.307	1
Seatbelt	2.44	0.118	1
UK Licence	1.42	0.234	1
Ethnicity	2.38	0.123	1
Road Condition	2.05	0.152	1
Visibility	8.71***	0.003	1
Casualties	0.11	0.740	1
Road Class	0.04	0.842	1
Road Type	1.85	0.173	1
Speed Limit	0.75	0.387	1
Weather	0.24	0.628	1
Weekday	0.14	0.709	1

The variables OAPs, pedestrians and cycles were dropped from this test since the low frequency of non-zero values prevented the test from running.

Table 58: HCM Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Severity	Severity	Lnsigma	Lnsigma	cut1	cut1	cut2	cut2
Severity								
Vehicles	-0.0372	0.572						
Driver Sex = Female	-0.637***	5.90e-05						
Driver Age	0.00249	0.726	0.00598***	0.00899				
Breath test = Positive	1.452***	2.54e-07						
Hit and Run = 1	-0.867	0.284						
Seatbelt = No Seatbelt	2.112***	4.35e-09						
Uk licence = Provisional	0.0655	0.831						
Uk licence = Unlicensed	-0.746*	0.0605						
Ethnicity = Asian	0.616	0.260						
Ethnicity = Black	0.325	0.521						
Ethnicity = Mixed Background	2.152***	0.00677						
Ethnicity = Oriental	-0.908	0.433						
Road Condition = Wet/Damp	-0.0184	0.907						
Road Condition = Snow	-0.438	0.608						
Road Condition = Frost/Ice	-0.836**	0.0416						
Road Condition = Flood	0.820	0.315						
Visibility = Darkness - Lights lit	0.435*	0.0870						
Visibility = Darkness - no lighting	0.452***	0.00372						
Casualties	0.568***	0	0.0824***	0.00231				
Road Type = One Way Street	-0.220	0.819						
Road Type = Dual Carriageway	-0.139	0.800						
Road Type = Single Carriageway	0.731	0.167						
Road Type = Slip Road	-3.009**	0.0292						
Speed Limit	0.0314***	6.84e-06						
Weather = Raining without high winds	-0.276	0.218						
Weather = Snowing without high winds	0.493	0.507						
Weather = Fine with high winds	0.713*	0.0781						
Weather = Raining with high winds	-1.350**	0.0278						
Weather = Snowing with high winds	0.103	0.936						
Weather = Fog or Mist	-0.368	0.462						
Weekday = 1	0.368*	0.0768						
lnsigma								
cut1								
cut2								
Constant					6.401***	0	10.09***	0
Observations	5,201		5,201		5,201		5,201	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 59: Constrained Gologit Results

	Slight	P-value	Serious	P-value
Vehicles	-0.012	0.760	-0.012	0.760
Driver Sex = Female	-0.396***	0.000	-0.396***	0.000
Driver Age	0.003	0.225	0.003	0.225
Breath Test = Positive	1.031***	0.000	1.031***	0.000
Hit and Run	-0.623	0.265	-0.623	0.265
Seatbelt = No Seatbelt	1.477***	0.000	1.477***	0.000
UK Licence = Provisional	0.012	0.957	0.012	0.957
UK Licence = Unlicensed	-0.633**	0.012	-0.633**	0.012
Ethnicity = Asian	0.530	0.141	0.530	0.141
Ethnicity = Black	0.270	0.437	0.270	0.437
Ethnicity = Mixed Background	1.275**	0.019	1.275**	0.019
Ethnicity = Oriental	-0.608	0.435	-0.608	0.435
Road Condition = Wet/Damp	-0.025	0.815	-0.025	0.815
Road Condition = Snow	-0.158	0.784	-0.158	0.784
Road Condition = Frost/Ice	-0.619**	0.032	-0.619**	0.032
Road Condition = Flood	0.595	0.271	0.595	0.271
Visibility = Darkness - Lights lit	0.299*	0.073	0.299**	0.073
Visibility = Darkness - no lighting	0.383***	0.000	0.383***	0.000
Casualties	0.402***	0.000	0.402***	0.000
Road Type = One Way Street	-0.150	0.824	-0.150	0.824
Road Type = Dual Carriageway	-0.045	0.905	-0.045	0.905
Road Type = Single Carriageway	0.549	0.129	0.549	0.129
Road Type = Slip Road	-2.053**	0.014	-2.053**	0.014
Speed Limit	0.021***	0.000	0.021***	0.000
Weather = Raining without high winds	-0.163	0.266	-0.163	0.266
Weather = Snowing without high winds	0.277	0.581	0.277	0.581
Weather = Fine with high winds	0.445	0.102	0.445	0.102
Weather = Raining with high winds	-0.958**	0.016	-0.958**	0.016
Weather = Snowing with high winds	0.012	0.990	0.012	0.990
Weather = Fog or Mist	-0.324	0.337	-0.324	0.337
Weekday	0.303**	0.026	0.303**	0.026
OAPs	0.447***	0.000	0.447***	0.000
Pedestrians	1.246***	0.000	1.246***	0.000
Cycles	-0.352	0.212	-0.352	0.212
Constant	-4.750***	0.000	-7.074***	0.000
Observations	5201.000			
Pseudo R squared	0.096			

Table 60: Unconstrained Gologit Results

	Slight	P-value	Serious	P-value
Vehicles	-0.026	0.530	0.075	0.327
Driver Sex = Female	-0.411***	0.000	0.020	0.933
Driver Age	0.003	0.204	0.003	0.625
Breath Test = Positive	1.100***	0.000	-0.127	0.848
Hit and Run	-0.643	0.250	1.278	0.248
Seatbelt = No Seatbelt	1.381***	0.000	2.003***	0.000
UK Licence = Provisional	-0.020	0.928	0.184	0.711
UK Licence = Unlicensed	-0.627**	0.013	-0.429	0.467
Ethnicity = Asian	0.440	0.227	1.040**	0.099
Ethnicity = Black	0.242	0.485	0.997	0.142
Ethnicity = Mixed Background	1.102**	0.044	2.044**	0.014
Ethnicity = Oriental	-0.557	0.474	-15.585	0.996
Road Condition = Wet/Damp	-0.019	0.858	-0.269	0.332
Road Condition = Snow	-0.199	0.730	-0.174	0.914
Road Condition = Frost/Ice	-0.616**	0.033	-15.595	0.989
Road Condition = Flood	0.556	0.303	2.597**	0.023
Visibility = Darkness - Lights lit	0.316*	0.059	-1.594	0.125
Visibility = Darkness - no lighting	0.328***	0.002	1.042***	0.000
Casualties	0.391***	0.000	0.442***	0.000
Road Type = One Way Street	-0.108	0.874	-2.963	0.999
Road Type = Dual Carriageway	-0.038	0.919	13.032	0.993
Road Type = Single Carriageway	0.549	0.129	13.641	0.992
Road Type = Slip Road	-2.038**	0.016	-3.598	0.999
Speed Limit	0.021***	0.000	0.016	0.114
Weather = Raining without high winds	-0.166	0.259	-0.186	0.676
Weather = Snowing without high winds	0.264	0.601	1.572	0.323
Weather = Fine with high winds	0.369	0.176	1.213**	0.015
Weather = Raining with high winds	-0.909**	0.023	-16.462	0.990
Weather = Snowing with high winds	0.081	0.928	-14.638	0.997
Weather = Fog or Mist	-0.304	0.367	-0.087	0.935
Weekday	0.311**	0.024	0.085	0.810
OAPs	0.428***	0.000	0.642***	0.000
Pedestrians	1.295***	0.000	1.719***	0.003
Cycles	-0.255	0.366	-0.928	0.108
Constant	-4.689***	0.000	-20.453	0.988
Observations		5201		
Pseudo R squared		0.1083		

Table 61: Partially Constrained Gologit Results

	Slight	p-value	Fatal	p-value
Vehicles	-0.025	0.539	0.114*	0.082
Driver Sex = Female	-0.397***	0.000	-0.397***	0.000
Driver Age	0.003	0.208	0.003	0.208
Breath Test = Positive	1.048***	0.000	1.048***	0.000
Hit and Run	-0.644	0.248	1.059	0.269
Seatbelt = No Seatbelt	1.437***	0.000	1.437***	0.000
UK Licence = Provisional	0.007	0.976	0.007	0.976
UK Licence = Unlicensed	-0.609**	0.016	-0.609**	0.016
Ethnicity = Asian	0.523	0.146	0.523	0.146
Ethnicity = Black	0.277	0.424	0.277	0.424
Ethnicity = Mixed Background	1.285**	0.018	1.285**	0.018
Ethnicity = Oriental	-0.603	0.436	-0.603	0.436
Road Condition = Wet/Damp	-0.031	0.770	-0.031	0.770
Road Condition = Snow	-0.174	0.763	-0.174	0.763
Road Condition = Frost/Ice	-0.632**	0.029	-0.632**	0.029
Road Condition = Flood	0.580	0.281	0.580	0.281
Visibility = Darkness - Lights lit	0.324*	0.053	-1.792*	0.078
Visibility = Darkness - no lighting	0.342***	0.001	0.878***	0.000
Casualties	0.399***	0.000	0.399***	0.000
Road Type = One Way Street	-0.133	0.845	-0.133	0.845
Road Type = Dual Carriageway	-0.027	0.943	-0.027	0.943
Road Type = Single Carriageway	0.565	0.118	0.565	0.118
Road Type = Slip Road	-2.024**	0.016	-2.024**	0.016
Speed Limit	0.021***	0.000	0.021***	0.000
Weather = Raining without high winds	-0.160	0.275	-0.160	0.275
Weather = Snowing without high winds	0.295	0.557	0.295	0.557
Weather = Fine with high winds	0.441	0.103	0.441	0.103
Weather = Raining with high winds	-0.944**	0.018	-0.944**	0.018
Weather = Snowing with high winds	0.010	0.991	0.010	0.991
Weather = Fog or Mist	-0.312	0.354	-0.312	0.354
Weekday	0.298**	0.028	0.298**	0.028
OAPs	0.424***	0.000	0.648***	0.000
Pedestrians	1.253***	0.000	1.253***	0.000
Cycles	-0.363	0.205	-0.363	0.205
Constant	-4.711***	0.000	-7.631***	0.000
Observations	5201			
Pseudo R squared	0.1017			

Table 62: Estimated coefficients for Ologit, HCM and PC-Gologit

Variable	Ologit	HCM	PC-Gologit	
			Slight	Serious
Vehicles	-0.012	-0.0372	-0.025	0.114*
Driver Sex = Female	-0.396	-0.637***	-0.397***	-0.397***
Driver Age	0.003	0.00249	0.003	0.003
Breath Test = Positive	1.031	1.452***	1.048***	1.048***
Hit and Run	-0.623	-0.867	-0.644	1.059
Seatbelt = No Seatbelt	1.477	2.112***	1.437***	1.437***
UK Licence = Provisional	0.012	0.0655	0.007	0.007
UK Licence = Unlicensed	-0.633	-0.746*	-0.609**	-0.609**
Ethnicity = Asian	0.530	0.616	0.523	0.523
Ethnicity = Black	0.270	0.325	0.277	0.277
Ethnicity = Mixed Background	1.275	2.152***	1.285**	1.285**
Ethnicity = Oriental	-0.608	-0.908	-0.603	-0.603
Road Condition = Wet/Damp	-0.025	-0.0184	-0.031	-0.031
Road Condition = Snow	-0.158	-0.438	-0.174	-0.174
Road Condition = Frost/Ice	-0.619	-0.836**	-0.632**	-0.632**
Road Condition = Flood	0.595	0.820	0.580	0.580
Visibility = Darkness - Lights lit	0.299	0.435*	0.324*	-1.792*
Visibility = Darkness - no lighting	0.383	0.452***	0.342***	0.878***
Casualties	0.402	0.568***	0.399***	0.399***
Road Type = One Way Street	-0.150	-0.220	-0.133	-0.133
Road Type = Dual Carriageway	-0.045	-0.139	-0.027	-0.027
Road Type = Single Carriageway	0.549	0.731	0.565	0.565
Road Type = Slip Road	-2.053	-3.009***	-2.024**	-2.024**
Speed Limit	0.021	0.0314***	0.021***	0.021***
Weather = Raining without high winds	-0.163	-0.276	-0.160	-0.160
Weather = Snowing without high winds	0.277	0.493	0.295	0.295
Weather = Fine with high winds	0.445	0.713*	0.441	0.441
Weather = Raining with high winds	-0.958	-1.350**	-0.944**	-0.944**
Weather = Snowing with high winds	0.012	0.103	0.010	0.010
Weather = Fog or Mist	-0.324	-0.368	-0.312	-0.312
Weekday	0.303	0.368*	0.298**	0.298**
OAPs	0.447		0.424***	0.648***
Pedestrians	1.246		1.253***	1.253***
Cycles	-0.352		-0.363	-0.363

Table 63: Log Odds and Odds

	Slight		Fatal	
Variable	Log Odds	Odds	Log Odds	Odds
Vehicles	-0.025	0.975	0.114	1.121
Driver Sex = Female	-0.397	0.673	-0.397	0.673
Driver Age	0.003	1.003	0.003	1.003
Breath Test = Positive	1.048	2.851	1.048	2.851
Hit and Run	-0.644	0.525	1.059	2.884
Seatbelt = No Seatbelt	1.437	4.207	1.437	4.207
UK Licence = Provisional	0.007	1.007	0.007	1.007
UK Licence = Unlicensed	-0.609	0.544	-0.609	0.544
Ethnicity = Asian	0.523	1.687	0.523	1.687
Ethnicity = Black	0.277	1.319	0.277	1.319
Ethnicity = Mixed Background	1.285	3.614	1.285	3.614
Ethnicity = Oriental	-0.603	0.547	-0.603	0.547
Road Condition = Wet/Damp	-0.031	0.970	-0.031	0.970
Road Condition = Snow	-0.174	0.841	-0.174	0.841
Road Condition = Frost/Ice	-0.632	0.532	-0.632	0.532
Road Condition = Flood	0.580	1.786	0.580	1.786
Visibility = Darkness - Lights lit	0.324	1.382	-1.792	0.167
Visibility = Darkness - no lighting	0.342	1.407	0.878	2.406
Casualties	0.399	1.490	0.399	1.490
Road Type = One Way Street	-0.133	0.876	-0.133	0.876
Road Type = Dual Carriageway	-0.027	0.974	-0.027	0.974
Road Type = Single Carriageway	0.565	1.759	0.565	1.759
Road Type = Slip Road	-2.024	0.132	-2.024	0.132
Speed Limit	0.021	1.021	0.021	1.021
Weather = Raining without high winds	-0.160	0.852	-0.160	0.852
Weather = Snowing without high winds	0.295	1.343	0.295	1.343
Weather = Fine with high winds	0.441	1.554	0.441	1.554
Weather = Raining with high winds	-0.944	0.389	-0.944	0.389

Weather = Snowing with high winds	0.010	1.010	0.010	1.010
Weather = Fog or Mist	-0.312	0.732	-0.312	0.732
Weekday	0.298	1.348	0.298	0.424
OAPs	0.424	1.528	0.648	1.911
Pedestrians	1.253	3.502	1.253	3.502
Cycles	-0.363	0.696	-0.363	0.696

Table 64: Marginal Effects

Variable	Value	Pr(Slight)	Pr(Severe)	Pr(Fatal)
Vehicles	1	0.849	0.135	0.016
	2	0.852	0.130	0.018
	3	0.855	0.125	0.021
	4	0.857	0.119	0.023
	5	0.860	0.113	0.027
	6	0.863	0.107	0.030
	7	0.865	0.101	0.034
	8	0.868	0.094	0.038
	9	0.870	0.087	0.043
	10	0.873	0.079	0.048
Driver Sex	Male	0.838	0.140	0.022
	Female	0.882	0.103	0.015
Driver Age	20	0.878	0.106	0.015
	30	0.866	0.117	0.017
	40	0.853	0.128	0.019
	50	0.839	0.140	0.022
	60	0.823	0.152	0.024
	70	0.807	0.166	0.027
	80	0.790	0.180	0.030
Breath Test	Positive	0.858	0.123	0.018
	Negative	0.709	0.245	0.046
Hit and Run	Not Hit and Run	0.852	0.128	0.020
	Hit and Run	0.917	0.073	0.010
Seatbelt	Seatbelt	0.859	0.123	0.018
	No Seatbelt	0.628	0.307	0.065
UK Licence	Full	0.851	0.129	0.020
UK Licence	Provisional	0.846	0.133	0.021
	Unlicensed	0.895	0.092	0.013
Ethnicity	White	0.854	0.127	0.019
	Asian	0.785	0.184	0.031
	Black	0.831	0.146	0.023
	Mixed Background	0.632	0.303	0.065
	Oriental	0.914	0.076	0.010
Road Condition	Dry	0.848	0.132	0.020
	Wet/Damp	0.854	0.127	0.019
	Snow	0.880	0.105	0.015
	Frost/Ice	0.912	0.077	0.011
	Flood	0.774	0.193	0.033
Visibility	Daylight	0.865	0.119	0.016
	Darkness - Lights lit	0.827	0.170	0.003
	Darkness - no lighting	0.826	0.140	0.034

Casualties	1	0.902	0.087	0.011
	2	0.859	0.124	0.017
	3	0.802	0.173	0.025
	4	0.730	0.232	0.038
	5	0.643	0.300	0.056
	6	0.547	0.370	0.083
	7	0.447	0.434	0.119
	8	0.351	0.482	0.167
	9	0.265	0.506	0.229
	10	0.194	0.503	0.304
Road Type	Roundabout	0.893	0.094	0.014
	One Way Street	0.899	0.088	0.013
	Dual Carriageway	0.897	0.090	0.013
	Single Carriageway	0.835	0.143	0.023
	Slip Road	0.977	0.020	0.003
Speed Limit	30	0.900	0.088	0.012
	40	0.881	0.104	0.014
	50	0.860	0.122	0.018
	60	0.836	0.143	0.021
	70	0.809	0.166	0.026
Weather	Fine without high winds	0.850	0.130	0.020
	Raining without high winds	0.867	0.116	0.017
	Snowing without high winds	0.805	0.167	0.027
	Fine with high winds	0.789	0.181	0.030
	Raining with high winds	0.932	0.060	0.008
	Snowing with high winds	0.846	0.133	0.020
	Fog or Mist	0.879	0.106	0.015
Weekday	Weekend	0.880	0.105	0.015
	Weekday	0.849	0.131	0.020
OAPs	1	0.819	0.155	0.026
	2	0.755	0.198	0.047
	3	0.678	0.240	0.081
	4	0.590	0.273	0.136
	5	0.497	0.286	0.217
	6	0.403	0.273	0.324
	7	0.315	0.235	0.450
	8	0.239	0.182	0.580
	9	0.175	0.127	0.698
	10	0.125	0.080	0.795
Pedestrians	1	0.658	0.284	0.058

	2	0.391	0.456	0.153
	3	0.172	0.493	0.335
	4	0.061	0.364	0.575
	5	0.019	0.199	0.782
	6	0.006	0.088	0.907
	7	0.002	0.032	0.966
	8	0.000	0.010	0.989
	9	0.000	0.003	0.997
	10	0.000	0.001	0.999
Cycles	1	0.889	0.097	0.014
	2	0.918	0.072	0.010
	3	0.940	0.053	0.007
	4	0.956	0.039	0.005
	5	0.969	0.028	0.004
	6	0.978	0.020	0.003
	7	0.984	0.014	0.002
	8	0.989	0.010	0.001
	9	0.992	0.007	0.001
	10	0.994	0.005	0.001

Table 65: Partially Constrained Gologit Results with Interactions

	Slight	P-value	Serious	P-value
Vehicles	-0.022	0.594	0.112*	0.090
Driver Sex = Female	-0.400***	0.000	-0.400***	0.000
Driver Age	0.002	0.432	0.002	0.432
Breath Test = Positive	1.075***	0.000	1.075***	0.000
Hit and Run	-0.676	0.228	1.046	0.275
Seatbelt = No Seatbelt	1.440***	0.000	1.440***	0.000
UK Licence = Provisional	0.007	0.972	0.007	0.972
UK Licence = Unlicensed	-0.559**	0.025	-0.559**	0.025
Ethnicity = Asian	0.522	0.150	0.522	0.150
Ethnicity = Black	0.277	0.425	0.277	0.425
Ethnicity = Mixed Background	1.290**	0.018	1.290**	0.018
Ethnicity = Oriental	-0.621	0.426	-0.621	0.426
Road Condition = Wet/Damp	-0.013	0.899	-0.013	0.899
Road Condition = Snow	-0.143	0.803	-0.143	0.803
Road Condition = Frost/Ice	-0.624***	0.032	-0.624**	0.032
Road Condition = Flood	0.605	0.262	0.605	0.262
Visibility = Darkness - Lights lit	0.325*	0.052	-1.814*	0.075
Visibility = Darkness - no lighting	0.343***	0.001	0.867***	0.000
Casualties	0.431***	0.000	0.431***	0.000
Road Type = One Way Street	-0.190	0.782	-0.190	0.782
Road Type = Dual Carriageway	-0.033	0.930	-0.033	0.930
Road Type = Single Carriageway	0.554	0.125	0.554	0.125
Road Type = Slip Road	-2.337**	0.011	-2.337**	0.011
Speed Limit	0.022***	0.000	0.022***	0.000
Weather = Raining without high winds	-0.180	0.222	-0.180	0.222
Weather = Snowing without high winds	0.277	0.581	0.277	0.581
Weather = Fine with high winds	0.431	0.113	0.431	0.113
Weather = Raining with high winds	-0.938**	0.018	-0.938**	0.018
Weather = Snowing with high winds	0.028	0.975	0.028	0.975
Weather = Fog or Mist	-0.314	0.351	-0.314	0.351
Weekday	0.288**	0.034	0.288**	0.034
OAPs	0.748***	0.000	1.003***	0.000
Pedestrians	1.786***	0.000	1.786***	0.000
Cycles	-0.167	0.473	-0.167	0.473
Constant	-0.101**	0.017	-0.101**	0.017
OAPs#Casualties	-0.251	0.104	-0.251	0.104
OAPs#Casualties#Pedestrians	-4.778***	0.000	-7.698***	0.000
Observations	5201			
Pseudo R squared	0.103			

Table 66: PC-Gologit Results for 2005 - 2014, 2005 - 2009 and 2010 - 2014

Year	**2005-2014**				**2005-2009**				**2010-2014**			
Variable	Slight		Fatal		Slight		Fatal		Slight		Fatal	
	Coef	p-value	Coef	p-value	Coef	p-value	Coef	p-value	Coef	p-value	Coef	p-value
Vehicles	-0.024	0.560	0.132**	0.050	-0.020	0.796	0.244*	0.094	-0.017	0.728	-0.017	0.728
Driver Sex = Female	-0.404***	0.000	-0.404***	0.000	-0.449***	0.010	-0.449***	0.010	-0.399***	0.000	-0.399***	0.000
Driver Age	0.012***	0.000	0.012***	0.000	0.005	0.266	0.005	0.266	0.014***	0.000	0.030***	0.000
Breath Test = Positive	1.009***	0.000	1.009***	0.000	1.118***	0.001	1.118***	0.001	1.008***	0.000	1.008***	0.000
Hit and Run	-0.703	0.206	-0.703	0.206	-17.797	0.998	-17.797	0.998	-0.548	0.340	-0.548	0.340
Seatbelt = No Seatbelt	1.410***	0.000	1.410***	0.000	1.683***	0.000	1.683***	0.000	1.291***	0.000	1.291***	0.000
UK Licence = Provisional	0.043	0.840	0.043	0.840	0.551	0.123	0.551	0.123	-0.256	0.348	0.967*	0.052
UK Licence = Unlicensed	-0.433*	0.074	-0.433*	0.074	-0.813	0.294	-0.813	0.294	-0.397	0.125	-0.397	0.125
Ethnicity = Asian	0.519	0.140	0.519	0.140	-0.172	0.830	-0.172	0.830	0.763*	0.057	0.763*	0.057
Ethnicity = Black	0.195	0.573	0.195	0.573	-18.813	0.998	5870000000.000	.	0.581	0.123	0.581	0.123
Ethnicity = Mixed Background	1.368***	0.010	1.368***	0.010	-0.005	0.997	-0.005	0.997	1.573***	0.007	3.190***	0.000
Ethnicity = Oriental	-0.643	0.407	-0.643	0.407	-18.370	0.999	-18.370	0.999	-0.365	0.650	-0.365	0.650
Road Condition = Wet/Damp	-0.056	0.596	-0.056	0.596	0.062	0.737	0.062	0.737	-0.104	0.418	-0.104	0.418
Road Condition = Snow	-0.301	0.605	-0.301	0.605	0.108	0.947	-310.942	.	-0.459	0.474	-0.459	0.474
Road Condition = Frost/Ice	-0.679	0.018	-0.679	0.018	-1.415	0.056	-1.415	0.056	-0.500	0.121	-0.500	0.121
Road Condition = Flood	0.535	0.317	0.535	0.317	1.548	0.057	1.548	0.057	-0.102	0.897	-0.102	0.897
Visibility = Darkness - Lights lit	0.318*	0.055	-1.809*	0.075	0.468	0.144	0.468	0.144	0.282	0.148	0.282	0.148
Visibility = Darkness - no lighting	0.327***	0.001	0.782***	0.000	0.304	0.107	0.304	0.107	0.358***	0.004	0.358***	0.004
Casualties	0.434***	0.000	0.434***	0.000	0.419***	0.000	0.419***	0.000	0.450***	0.000	0.628***	0.000
Road Type = One Way Street	-0.079	0.904	-0.079	0.904	-18.153	0.997	-18.153	0.997	0.569	0.432	0.569	0.432
Road Type = Dual Carriageway	-0.050	0.893	-0.050	0.893	-0.461	0.470	-0.461	0.470	0.174	0.708	0.174	0.708
Road Type = Single Carriageway	0.547	0.125	0.547	0.125	0.351	0.563	0.351	0.563	0.751	0.096	0.751	0.096
Road Type = Slip Road	-1.727**	0.032	-1.727**	0.032	-18.109	0.997	-18.109	0.997	-1.209	0.157	-1.209	0.157
Speed Limit	0.021***	0.000	0.021***	0.000	0.027***	0.001	0.093***	0.000	0.018***	0.000	0.018***	0.000
Weather = Raining without high winds	-0.150	0.305	-0.150	0.305	-0.597**	0.039	-0.597**	0.039	0.020	0.910	0.020	0.910
Weather = Snowing without high winds	0.346	0.496	0.346	0.496	-1.031	0.513	-310.942	.	0.678	0.217	0.678	0.217
Weather = Fine with high winds	0.458*	0.085	0.458*	0.085	0.533	0.238	0.533	0.238	0.319	0.346	1.695***	0.002
Weather = Raining with high winds	-0.965**	0.015	-0.965**	0.015	-1.532*	0.065	-1.532*	0.065	-0.716	0.120	-0.716	0.120
Weather = Snowing with high winds	0.032	0.972	0.032	0.972	-17.673	0.999	-17.673	0.999	0.458	0.635	0.458	0.635
Weather = Fog or Mist	-0.270	0.415	-0.270	0.415	-1.184	0.126	-1.184	0.126	0.035	0.926	0.035	0.926
Weekday	0.293**	0.031	0.293**	0.031	0.635**	0.018	-0.708	0.151	0.146	0.363	0.146	0.363
Constant	-4.919***	0.000	-7.672***	0.000	-5.093***	0.000	-10.523***	0.000	-4.962***	0.000	-8.720***	0.000

C.4 Robustness Checks

C.5 χ^2 Tests

Figure 22: Driver Age 17 - 20

(a) All drivers aged 17-20

(b) Male and female drivers ages 17 - 20

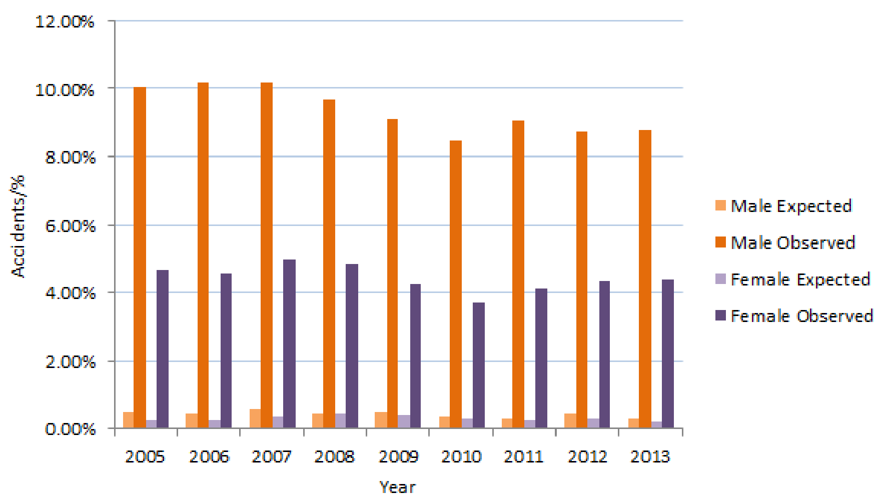


Table 67: Driver Age 17 - 20

	All				Male				Female			
Year	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value
2005	0.007	0.145	17747***	0.000	0.005	0.100	12378***	0.000	0.003	0.046	4516***	0.000
2006	0.007	0.146	16549***	0.000	0.004	0.102	14905***	0.000	0.003	0.046	3594***	0.000
2007	0.009	0.149	13044***	0.000	0.006	0.102	9600***	0.000	0.004	0.050	3444***	0.000
2008	0.009	0.144	10924***	0.000	0.004	0.097	12241***	0.000	0.004	0.048	2921***	0.000
2009	0.009	0.131	8621***	0.000	0.005	0.091	7989***	0.000	0.004	0.043	2119***	0.000
2010	0.006	0.122	10556***	0.000	0.003	0.085	10968***	0.000	0.003	0.037	1980***	0.000
2011	0.006	0.131	16583***	0.000	0.003	0.091	16901***	0.000	0.003	0.041	3380***	0.000
2012	0.007	0.129	13334***	0.000	0.004	0.087	11469***	0.000	0.003	0.043	3735***	0.000
2013	0.005	0.131	19297***	0.000	0.003	0.088	15184***	0.000	0.002	0.044	5839***	0.000

Figure 23: Driver Age 21 - 29

(a) All drivers aged 21 - 29

(b) Male and female drivers aged 21 - 29

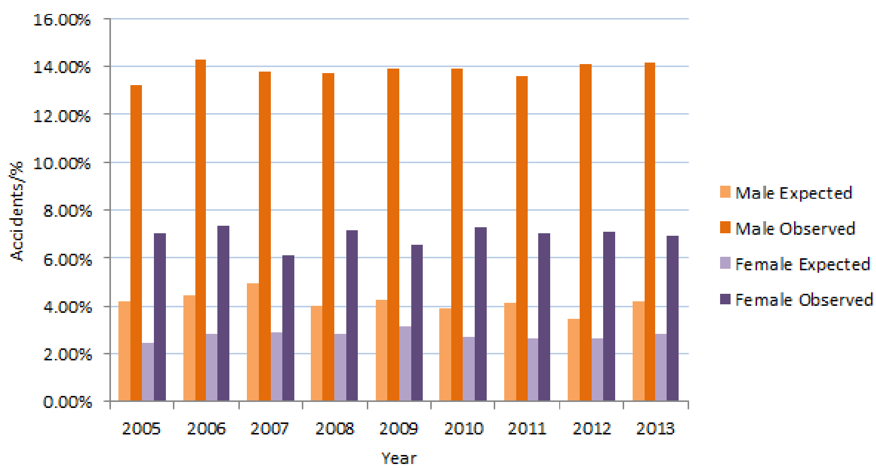
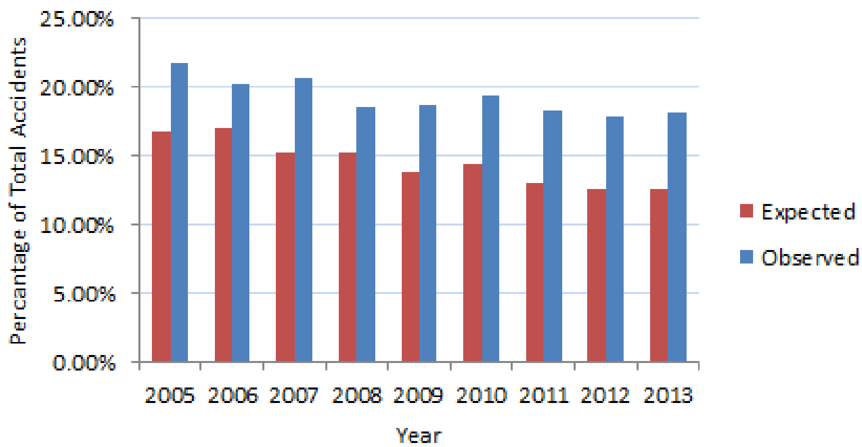


Table 68: Driver Age 21 - 29

	All				Male				Female			
Year	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value
2005	0.066	0.202	1521***	0.000	0.042	0.132	1185***	0.000	0.025	0.070	552***	0.000
2006	0.072	0.215	1356***	0.000	0.044	0.143	1222***	0.000	0.028	0.074	452***	0.000
2007	0.078	0.197	872***	0.000	0.050	0.138	847***	0.000	0.029	0.061	219***	0.000
2008	0.068	0.207	1264***	0.000	0.040	0.137	1202***	0.000	0.029	0.072	343***	0.000
2009	0.073	0.204	980***	0.000	0.042	0.139	1074***	0.000	0.031	0.066	197***	0.000
2010	0.065	0.209	1245***	0.000	0.039	0.139	1133***	0.000	0.027	0.073	362***	0.000
2011	0.067	0.203	1444***	0.000	0.041	0.136	1293***	0.000	0.027	0.070	443***	0.000
2012	0.061	0.211	1945***	0.000	0.035	0.141	1897***	0.000	0.026	0.071	492***	0.000
2013	0.070	0.210	1381***	0.000	0.042	0.142	7215***	0.000	0.029	0.069	336***	0.000

Figure 24: Driver Age 30 - 39

(a) All drivers aged 30 - 39



(b) Male and female drivers aged 30 - 39

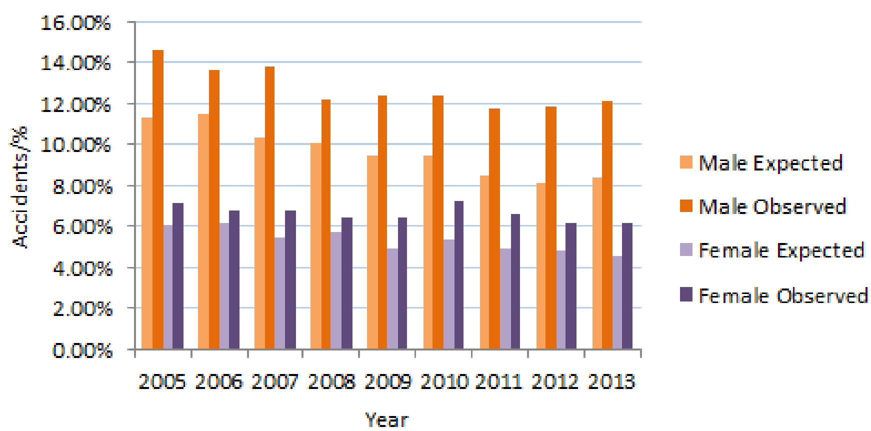


Table 69: Driver Age 30 - 39

	All				Male				Female			
Year	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value
2005	0.172	0.216	2.271	0.132	0.113	0.146	15.164***	0.000	0.061	0.072	4.673**	0.031
2006	0.174	0.202	2.080	0.149	0.115	0.136	1.747	0.186	0.062	0.068	0.365	0.546
2007	0.156	0.205	13.222***	0.000	0.103	0.139	29.891***	0.000	0.055	0.068	10.218***	0.003
2008	0.156	0.185	0.000	0.996	0.101	0.122	4.778**	0.029	0.057	0.064	1.207	0.272
2009	0.141	0.187	15.822***	0.000	0.094	0.124	20.063***	0.000	0.050	0.064	13.688***	0.000
2010	0.146	0.194	14.138***	0.000	0.095	0.124	16.071***	0.000	0.054	0.072	13.288***	0.000
2011	0.132	0.182	35.960***	0.000	0.085	0.118	43.478***	0.000	0.049	0.066	28.168***	0.000
2012	0.128	0.178	39.506***	0.000	0.081	0.118	65.408***	0.000	0.049	0.062	13.748***	0.000
2013	0.127	0.181	50.713***	0.000	0.083	0.121	62.734***	0.000	0.045	0.061	27.756***	0.000

Figure 25: Driver Age 40 - 49

(a) All drivers aged 40 - 49

(b) Male and female drivers aged 40 - 49

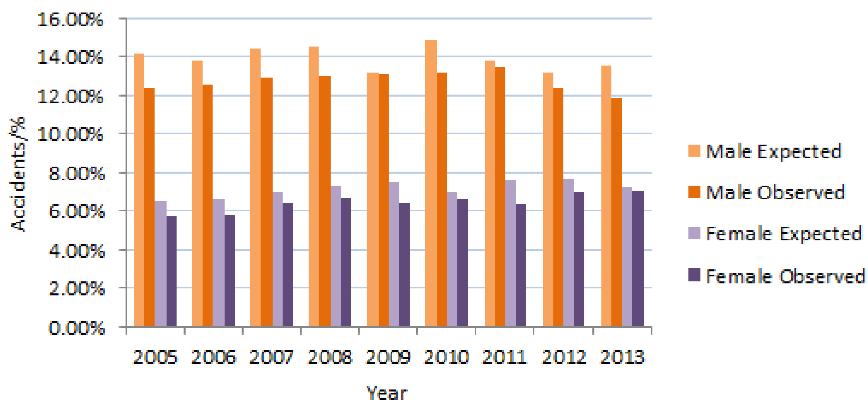


Table 70: Driver Age 40 - 49

	All				Male				Female			
Year	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value
2005	0.202	0.181	132.560***	0.000	0.141	0.124	65.968***	0.000	0.066	0.057	17.566***	0.000
2006	0.201	0.183	108.710***	0.000	0.138	0.125	45.862***	0.000	0.066	0.058	14.400***	0.000
2007	0.209	0.192	116.170***	0.000	0.144	0.129	56.330***	0.000	0.069	0.064	9.268***	0.002
2008	0.215	0.196	116.710***	0.000	0.145	0.131	49.681***	0.000	0.073	0.067	10.781***	0.001
2009	0.204	0.193	77.022***	0.000	0.132	0.131	14.957***	0.000	0.075	0.064	19.605***	0.000
2010	0.214	0.195	97.502***	0.000	0.148	0.132	46.949***	0.000	0.070	0.066	6.087**	0.014
2011	0.210	0.197	106.400***	0.000	0.138	0.135	25.562***	0.000	0.076	0.064	28.847***	0.000
2012	0.207	0.190	116.370***	0.000	0.132	0.124	33.338***	0.000	0.077	0.069	15.845***	0.000
2013	0.205	0.187	113.910***	0.000	0.135	0.118	56.984***	0.000	0.073	0.071	5.036**	0.025

Figure 26: Driver Age 50 - 59

(a) All drivers aged 50 - 59

(b) Male and female drivers aged 50 - 59

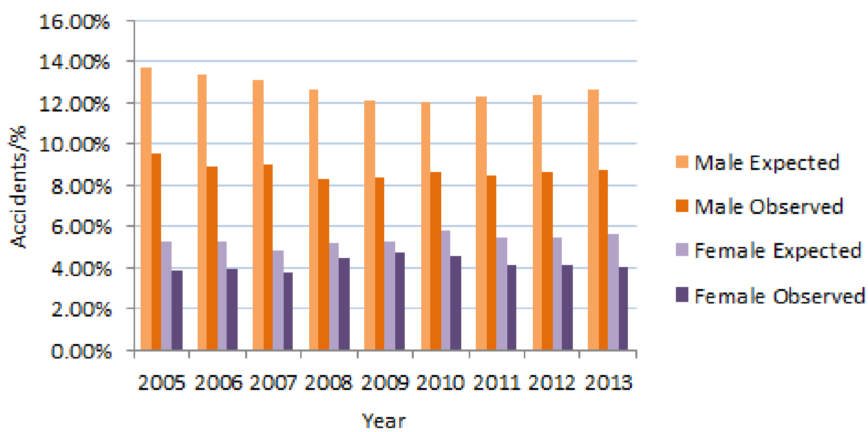
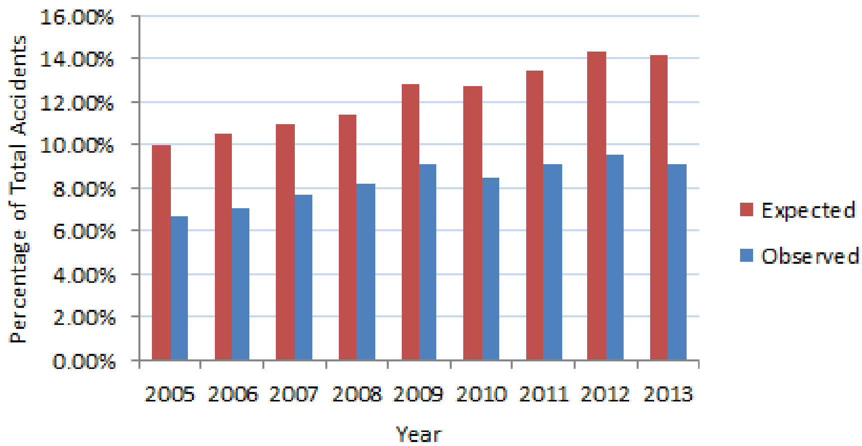


Table 71: Driver Age 50 - 59

	All				Male				Female			
Year	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value
2005	0.185	0.133	258.200***	0.000	0.138	0.095	180.460***	0.000	0.053	0.039	39.005***	0.000
2006	0.181	0.128	239.590***	0.000	0.134	0.089	174.830***	0.000	0.053	0.039	32.309***	0.000
2007	0.173	0.126	204.580***	0.000	0.131	0.090	155.730***	0.000	0.048	0.037	22.363***	0.000
2008	0.173	0.126	183.880***	0.000	0.126	0.083	147.580***	0.000	0.052	0.044	11.866***	0.001
2009	0.171	0.130	147.600***	0.000	0.122	0.084	117.480***	0.000	0.053	0.048	7.023***	0.008
2010	0.176	0.130	153.830***	0.000	0.120	0.087	88.713***	0.000	0.058	0.045	22.535***	0.000
2011	0.174	0.124	226.160***	0.000	0.123	0.085	145.100***	0.000	0.055	0.041	33.799***	0.000
2012	0.174	0.126	217.180***	0.000	0.124	0.086	143.110***	0.000	0.055	0.041	34.423***	0.000
2013	0.179	0.128	123.220***	0.000	0.126	0.087	144.300***	0.000	0.056	0.040	41.023***	0.000

Figure 27: Driver Age 60 - 69

(a) All drivers aged 60 - 69



(b) Male and female drivers aged 60 - 69

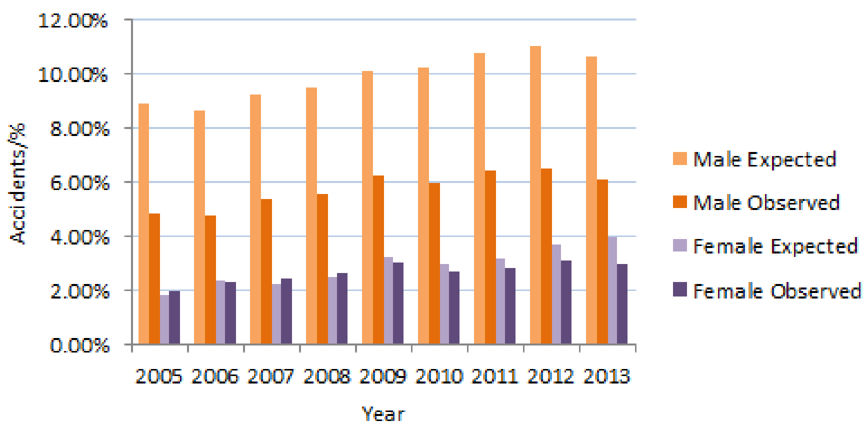
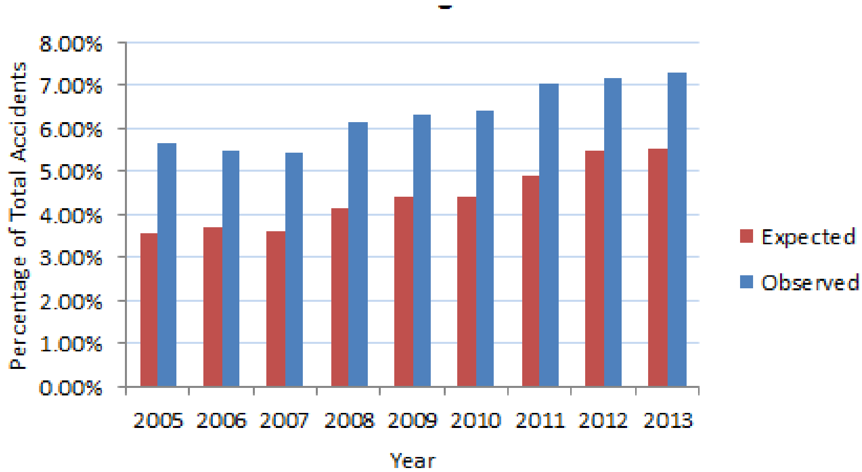


Table 72: Driver Age 60 - 69

	All				Male				Female			
Year	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value
2005	0.097	0.067	108.010***	0.000	0.089	0.048	187.440***	0.000	0.018	0.019	0.337	0.562
2006	0.101	0.071	98.283***	0.000	0.086	0.048	149.400***	0.000	0.024	0.023	0.905	0.342
2007	0.106	0.076	97.256***	0.000	0.092	0.053	150.860***	0.000	0.022	0.024	0.658	0.417
2008	0.111	0.082	85.735***	0.000	0.095	0.056	137.000***	0.000	0.025	0.026	0.157	0.692
2009	0.125	0.091	102.030***	0.000	0.101	0.062	125.020***	0.000	0.032	0.030	1.741	0.187
2010	0.124	0.085	110.970***	0.000	0.102	0.059	130.820***	0.000	0.030	0.027	2.506	0.113
2011	0.131	0.091	154.910***	0.000	0.107	0.065	172.870***	0.000	0.031	0.028	3.696*	0.055
2012	0.141	0.095	191.290***	0.000	0.110	0.065	189.700***	0.000	0.037	0.031	10.335***	0.001
2013	0.140	0.091	202.360***	0.000	0.106	0.061	183.840***	0.000	0.039	0.030	19.676***	0.000

Figure 28: Driver Age 70+

(a) All drivers aged 70+



(b) Male and female drivers aged 70+

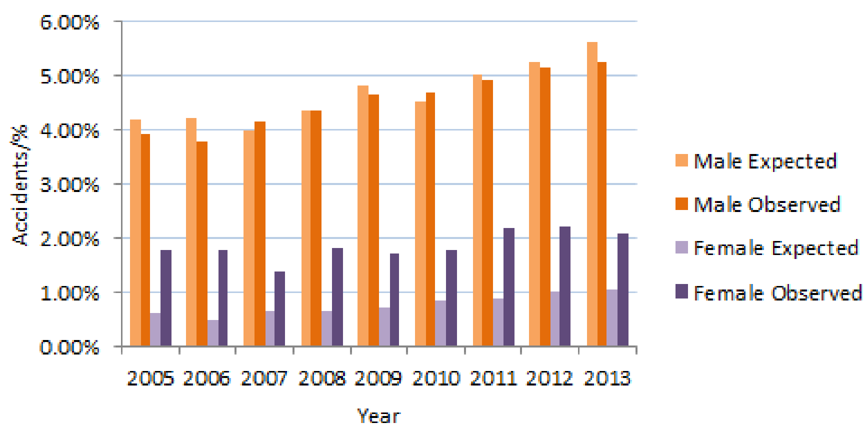


Table 73: Driver Age 70+

	All				Male				Female			
Year	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value	Expected	Observed	χ^2	p-value
2005	0.038	0.056	51.587***	0.000	0.042	0.039	3.849**	0.050	0.006	0.018	173.080***	0.000
2006	0.035	0.055	63.044***	0.000	0.042	0.038	5.689**	0.017	0.005	0.018	216.450***	0.000
2007	0.036	0.054	51.465***	0.000	0.040	0.041	0.006	0.941	0.006	0.014	70.803***	0.000
2008	0.040	0.061	54.960***	0.000	0.043	0.043	0.359	0.549	0.007	0.018	106.460***	0.000
2009	0.044	0.063	36.246***	0.000	0.048	0.046	1.773	0.183	0.007	0.017	82.727***	0.000
2010	0.044	0.064	36.228***	0.000	0.045	0.047	0.021	0.887	0.008	0.018	63.422***	0.000
2011	0.049	0.070	45.589***	0.000	0.050	0.049	1.617	0.204	0.009	0.022	125.280***	0.000
2012	0.053	0.072	30.450***	0.000	0.053	0.051	2.481	0.115	0.010	0.022	102.830***	0.000
2013	0.056	0.073	20.646***	0.000	0.056	0.052	9.935***	0.002	0.011	0.021	55.056***	0.000