An Empirical Analysis of the Determinants of Guilty Plea Discount

Jose Pina-Sánchez

PhD student in Social Statistics at the University of Manchester

Executive Summary

In this report, I assess the application of the 2007 Sentencing Guidelines Council guideline, *Reductions for a Guilty Plea,* empirically using data collected on the Sentencing Council's Crown Court Sentencing Survey in the year 2011. I begin by using an exploratory analysis to observe the relationship between the level of discount applied and the stage at which the guilty plea was entered. I then consider the possible impact of other factors taken into account when sentencing on the reduction applied for a guilty plea. For this, I specify different models for discrete data to regress the level of discount on a broad set of explanatory variables. Results point towards a substantial degree of agreement between the recommendations provided in the 2007 guideline and the actual level of discount received by offenders who plead guilty. In particular, the stage based approach recommended in the 2007 guideline was found to be the major factor determining the level of discount applied. However, the results also show there to be a number of departures from the guideline, such as: a) a high proportion of cases where the reduction given was higher than the maximum recommended level of 33%, with these anomalies concentrated in specific Courts; b) the presence of particular aggravating factors, on average, leading to lower levels of discount after controlling for the stage when the guilty plea was entered; and c) the presence of the mitigating factor remorse, on average, having a positive significant effect on the level of discount.

1. Introduction

In 2007 the Sentencing Guidelines Council, issued a revised guideline for judges to determine the appropriate reduction to apply in cases where a guilty plea was entered. This built on the then current guideline which was issued in 2004. The new 2007 guideline maintained the consideration that the level of reduction applied to an offender's sentence when they plead guilty should continue to be based on pragmatic justification. In particular, it describes guilty plea discounts as recognition that, by avoiding the full court process, a guilty plea saves courts from the cost of conducting further administrative work and sometimes, the cost of a trial, but also spares victims and witnesses from the stress of attending or giving evidence at the trial. The main novelty of the 2007 revision was the "hardening" of the process used in the 2004 guideline to determine the appropriate level of guilty plea discount.

By law, when determining the reduction to apply to the sentence of an offender who has pleaded guilty to an offence, the court must take into account: 1) the stage in the proceedings for the offence at which the offender indicated his intention to plead guilty; and 2) the circumstances in which this indication was given. The system proposed in the revised 2007 guideline establishes a clearer process for taking account of the legislation, which, in practice, should reduce inconsistency in the level of discount applied across different courts and judges. To address point 1), the guideline recommends that the level of reduction should be gauged on a sliding scale ranging from a recommended one third (where the guilty plea is entered at the first reasonable opportunity in relation to the offence being sentenced), reducing to a recommended one quarter (where the plea is entered after a trial date has been sent), down to a recommended one tenth (where the guilty plea is entered at the 'door of the court' or after the trial has begun) (Roberts and Bradford, 2013).

Figure 1: The sliding scale of discounts defined in the 2007 guideline

In each category, there is a presumption that the recommended reduction will be given unless there are good reasons for a lower amount.

First reasonable opportunity	After a trial date is set	Door of the court/ after trial has begun
recommended 1/3	**recommended 1/4**	**recommended 1/10**

Source: *Reduction in Sentence for a guilty plea*, p. 6.

To address point 2) which specifies that the circumstances under which the indication was given should be taken into account, the 2007 guideline identifies a list of cases in which the discount indicated by the sliding scale could be modified (see Section E of the 2007 guideline). This includes a recommendation that in cases where there is overwhelming evidence pointing at the culpability of the offender, the maximum amount of discount should be limited to a maximum of 20%, whilst later pleas which might have otherwise attracted a reduction of 25% or 10% should receive a lower reduction.

Another central topic covered by the 2007 guideline is the clarification that the reduction principle is derived from the need for the effective administration of justice and therefore, should not be seen as an aspect of mitigation. In particular, the revised guideline emphasises that remorse and assistance provided to prosecuting authorities are separate issues from whether a guilty plea was entered and makes clear that the approach to calculating the reduction should not take account of the latter circumstances. That is, when deciding on the most appropriate length of sentence, sentencers should address separately the issue of remorse, and any other mitigating features, before calculating the reduction for a guilty plea. This process was established to prevent sentencers from double counting for any mitigating factors when deriving the final sentence.

In this paper I analyse empirically the data from the Crown Court Sentencing Survey (CCSS) on for sentences passed in 2011 where the offender entered a guilty plea. The aim is twofold: to test whether discounts for a guilty plea are consistent with the recommendations provided in the 2007 guideline; and to identify whether other factors taken into account when sentencing have a significant effect on the level of discount received.

In the following section I provide an overview of the data used, covering its strengths and weaknesses. The results of my analysis are then presented in Section 3, which is divided into three subsections: the first showing results from an initial exploratory analysis; and the second and third showing the use of an ordered and two binary logit model, respectively. Section 4 concludes with a summary of the main findings.

2. Data

The Crown Court Sentencing Survey (CCSS) is a census survey of all cases sentenced at the Crown Court in England and Wales. It records the key factual elements of each case that were taken into account by the judge to determine an appropriate sentence for the offender. For example, information is recorded on the seriousness of the offence, the aggravating and mitigating factors present, and the number of recent, relevant previous convictions of the offender. In cases where the offender entered a guilty plea, the survey also collects information on the stage of the proceedings at which the plea was entered, whether it was entered at first reasonable opportunity, and the discount given.

Before this survey, there was no comprehensive source of data on the determinants of sentences that would allow this kind of exploration of sentencing decisions. This dataset allows us to explore guilty plea reduction in much more detail than was previously possible. As Roberts and Bradford (2013) note, before the inception of the CCSS, the only previous insight into the magnitude of pleas discounts has come from the sentencing statistics published by the Ministry of Justice. However, these statistics are very limited as they do not allow differences to be identified between the set of offenders who enter a guilty plea to those who do not. For example, no information on aggravating or mitigating factors is collected, and information is only available on whether a guilty plea was entered or not, not the level of discount received or stage of the plea.

On the other hand, like most other survey data, the CCSS is prone to data quality issues that should be kept in mind. The key concerns that affect the analysis are pointed out below. First of all, rather than capturing the exact discount level received for a guilty plea, the survey uses a rather coarse unit of measure to capture the discount level (no discount given, 1% to 10%, 11%,to 20%, 21% to 32%, a third, or more than third). The use of this discrete metric represents a loss of information since the variable derived from this question is not continuous but ordinal.

In addition, and in spite of the improvement in the amount of information available, there are still some relevant areas in the study of guilty plea reductions which are not covered by the CCSS. For example, there are no questions capturing the presence of overwhelming evidence or indicating whether the case included a Newton hearing[1], both of which might legitimately influence the discount level without going against the recommendations of the guideline.

Although the survey is a census, it is not a mandatory requirement and therefore suffers from a significant problem of missing data. In 2011, the average response rate of the CCSS was 61%. This is particularly problematic as there is no other source of data that will allow us to rule out the possibility that this missing data is non-ignorable[2]. In particular, it could be argued that judges who think positively about the sentencing guidelines might be the ones with a higher response rate. In the analysis I also discarded cases for which the section on guilty plea discounts on the form had not been fully completed. Where any of the questions on stage, discount, or first opportunity were unanswered, the record was removed. These restrictions reduced the sample size to a total of 40,783 sentences imposed at the Crown Court in 2011.

3. Results

3.1. Exploratory Analysis

Table 1 shows the distribution of reductions received according to the discrete categories used on the survey forms. The majority of cases, 64.3% received a discount of exactly 33%. A small proportion of sentences received discounts outside of the minimum 10% and maximum 33% specified by the sliding scale approach of the 2007 guideline. In total, 1.6% of received a discount of 0% and 6.8% were given a discount of more than 33%. Cases where no discount was received might be explained where there was overwhelming evidence pointing to the guilt of the offender and the plea was entered at the last possible opportunity. However, the reviewed guidelines do not consider situations where the offender might get more than the maximum 33% discount.

Table 1. Percentage of Cases by Category of Discount

0%	1-10%	11-20%	21-32%	33%	>33%
1.6%	8.2%	7.6%	11.6%	64.3%	6.8%

Looking at the stage of the proceedings at which the guilty plea entered, I observe a more uniform distribution of cases. Table 2 shows that all categories have 10% cases or more, with the main concentration of cases, 45%, indicated to have entered their plea at the "plea and case management hearing" (PCMH).

[1] These are a small number of cases when the defendant may, on the day of trial, enter a plea to a lesser charge, and when this occurs, he or she may still benefit from the maximum recommended reduction as this point is considered the first opportunity to enter a plea to this specific allegation.

[2] See Rubin (1987) for a classification of the implications and possible adjustments for the different missing data mechanisms.

Table 2. Percentage of Cases by Stage of Plea

At Magistrates	Prior to PCMH	At PCMH	After PCMH	At Trial
15%	11.1%	45%	1.10%	18.8%

In order to see how these different stages match the discounts that we have observed, in Table 3, I provide a contingency table with these two variables. Although the categories available on the survey form for both of these variables differ to the ones defined in the 2007 guideline, the variables collected still provide a good proxy for the stages and discounts specifically referenced in the sliding scale shown in Figure 1. As expected, Table 3 shows a strong association between earlier pleas and higher levels of discount. When pleading guilty at the trial, the majority of sentencers (45.1%) applied a discount of 1-10%. For pleas entered after the PCMH but before trial, discounts are mainly concentrated in the three categories 11-20%, 21%-32% and 33%, with the biggest proportion of cases (36.8%) receiving reductions in the range 21-32%. The other three stages, which represent the earliest points of the court proceedings, have very similar compositions. In each of these categories, the vast majority of cases (about 80%) obtain reductions of 33%. However, differences can be seen between these stages when looking at the proportion that receive discounts of 21-32% or >33%. The former discount category is inversely related with earlier stages, while the latter is directly related. In particular, we can now see that the probability of obtaining anomalous cases of >33% is higher when the guilty plea is entered at the magistrates court than in any other stage.

Table 3. Contingency Table of Discounts vs Stages

	At Magistrates	Prior to PCMH	At PCMH	After PCMH	At Trial
0%	.5%	0%	.4%	1.1%	.4%
1-10%	.7%	0%	.3%	4.7%	45.1%
11-20%	1.7%	.4%	1.7%	22.1%	27.2%
21-32%	3.7%	6.3%	13.1%	36.8%	9.8%
33%	77.2%	81.8%	81.6%	33.5%	12.4%
>33%	16.2%	11.5%	3.0%	1.8%	.1%

In order to examine whether these patterns are general across the full sample or whether they are influenced by particular courts, I now look at the distributions of these two key variables (level of discount and stage) by court. For reasons of space, this analysis is shown graphically. Appendix I includes histograms of the number of cases falling into each of the discount and stage categories for each of the 76 Crown Court centres covered by the CCSS (Figure A1 and A2, respectively). A third chart showing the proportion of all guilty plea sentences that received a discount of >33% for each court centre is also provided (Figure A3).

From the two first histograms we can observe a strong variability in terms of the number of volume of sentences by court centres. Court centres such as Dorchester, Taunton, or Salisbury, sentenced very few cases, 75, 71, and 32 cases respectively, whilst court centres such as Liverpool returned forms for 2,249 sentences. We can also see substantial variability amongst court centres in the proportion of cases falling into each of the discount and stage categories. For example, Mold Crown Court only recorded a few cases (26 in total, 4% of the guilty pleas entered in this court) of guilty pleas entered on the day of the trial, whilst Sheffield Crown Court shows a similar proportion for pleas entered on the day of the trial to pleas entered at the early stage of the PCMH, which as we saw in Table 2, is the stage where most of the pleas are entered.

In order to visualize differences between court centres in relative terms, we can observe the proportion of discounts received that were larger than 33% for each court centre. This is shown in Figure A3. Here we can see that the proportion of anomalous >33% reductions is higher in Nottingham, Derby, or Truro (19.4%, 17.3%, and 14.1%, respectively) than in other court centres. In contrast, for court centres such as Swindon, Stoke-on-Trent, or Winchester, these cases are exceptional (2.2%, 2.1%, and 1.5%, respectively).

In summary, in this exploratory analysis I have detected both expected and unexpected patterns. On one hand, I have found that the recommended sliding scale linking reductions to the stage at which the plea was entered is largely complied to, with pleas entered at the earlier stages being more likely to receive a higher level of reduction. On the other hand, there still exist a fair number of anomalies that fall outside of the guideline approach, including >33% discounts which are not contemplated by the guidelines. There is also evidence to suggest that the approach taken differs amongst court centres[3]. I now proceed to extend the analysis using inferential statistics. In particular, I use an ordered and a multinomial logit model to regress the category of discount on a set of explanatory variables drawn from the CCSS dataset. These two models will be used to test whether guilty plea discounts are uniquely determined by the guideline's sliding scale or whether there are other variables that also have a significant effect.

The aggravating and mitigating factors available on the survey form differ according the type of offence being sentenced. To ensure that every record in the sample contains information on the same set of explanatory variables, I have restricted the analysis to sentences passed for offences covered by the offence type of assault and other public order offences. This restriction gives us a sample size of 9,187 forms for sentences where the offender pleaded guilty, and the following list of explanatory variables: stage of guilty plea (at magistrates, prior to PCMH, PCMH, after PCMH, and day of trial used as the reference case), type of assault (GBH, GBH with intent, common assault, affray, and ABH used as the reference case), sentence length, gender, previous convictions, a group of aggravating and mitigating factors (remorse, vulnerable, public worker, sustained, and drugs), and the different Crown Court centre at which the sentence was passed (with Aylesbury as the reference case).

3.2. The ordered logit model

The first regression model I present is an ordered logit model where the set of explanatory variables is regressed on the variable capturing levels of discount. This model assumes that the response variable can be interpreted as an ordinal variable[4]. That is, the different levels of discounts can be ranked from low to high, without necessarily knowing what the distances between adjacent categories are.

[3] Similar findings on lack of consistency in the application of guilty pleas across Crown Court centres were found in Robertshaw and Milne (1992). Here, the authors looked at changes in whether a custodial sentence is passed or not in 60 Crown Courts in 1987 and 1988. They found that on average those who pleaded guilty had lower custody rates, but that this was not the case in every Court.
[4] See Appendix II

Results from this model are included in Table 4 below, with the exception of the coefficients for the different Crown Courts, which for reasons of space have been relegated to Appendix III. I find that, as expected, the strongest predictors of sentence discounts are the different stages where the GP was introduced, and whether this was done at the first opportunity.

Furthermore, most of the additional variables included in the model were not found statistically significant. In particular, it is interesting to note the lack of significance of sentence length in logs and previous convictions. These two variables could be understood as proxies for severity of the offence and dangerousness of the offender. So, the fact that they are not associated with levels of discount shows evidence on lack of double mitigation, but it also disproves the focal hypothesis. Ball (2006) used this hypothesis to explain how disparities in plea bargaining might be due to judges having incomplete information about defendants and their cases and, thus, rely on a perceptual shorthand to which they apply their own biases and interject stereotypes regarding the dangerousness of a particular offender.

However, other factors have an effect on GP reductions, which contradicts the guidelines' recommendations, and shows evidence of both double aggravation and double mitigation. The aggravating factors of perpetrating an assault sustained in time and being under the influence of drugs reduce the probability of obtaining a higher discount; while the mitigating factor of remorse, and in particular the interaction of remorse and entering the GP at the magistrates' court increase the probability of obtaining a higher discount. In addition cases of affray and GBH with intent also have a positive effect on levels of discount.

It is possible that the significant negative effect of the aggravating factors of a sustained assault and being under the influence might be related to the problem of omitted relevant variables. The model does not include a variable to capture the effect of reduced discounts for reasons of overwhelming evidence. Cases where there is overwhelming evidence often occur when the prosecution is able to provide DNA or CCTV evidence against the offender. It could be argued that sustained assaults or those perpetrated under the influence of drugs or alcohol might be more susceptible to being captured on CCTV. If so, the legitimate negative influence of the presence of overwhelming evidence would be wrongly picked up by variables which indicate the presence of these two aggravating factors, therefore negatively biasing their estimates in the model.

The positive effect of remorse represents a general departure from the approach recommended by the guideline. Regardless of the stage at which the plea was entered, results indicate that the presence of remorse, on average, has the effect of inflating the discount level received. This departure from the guidelines might represent an element of double counting for this mitigating factor in cases when the judge has already accounted for remorse at an earlier stage.

In addition, by including interaction effects between remorse and the different stages of plea, I find that this effect is especially pronounced for cases where the plea was entered at the magistrates court.

The positive effects for GBH with intent and affray are more difficult to understand. It might be argued that offences that are generally considered to be less serious compared to other offences are likely to show a better agreement with the guidelines, while the most serious cases are subject to deeper and more subjective deliberations from the sentencer[5] so may be more likely to depart from the guideline. This could explain the departure that we see for cases of GBH with intent, since these are more serious offences than the reference case of ABH, however, this hypothesis does not explain the positive effect that we see for cases of affray.

Table 4. Ordered Logit Model*

Variable	Coefficient	Standard Error
Magistrates	**4.8**	.14
Prior to PCMH	**5.03**	.17
PCMH	**3.9**	.08
After PCMH	**1.85**	.09
First op.	**1.37**	.06
Log-length	-.004	.01
Female	.07	.08
GBH	.05	.1
Intent	**.26**	.06
Common	-.002	.08
Affray	**.31**	.07
Previous conv.	-.04	.03
Remorse	**.27**	.09
Rem*Magist.	**.64**	.18
Rem*PCMH	.05	.12
Rem*PriorPCMH	.1	.23
Rem*AftPCMH	.08	.14
Carer	-.11	.1
Gang	.01	.06
Vulnerable	.09	.08
Pub. Worker	-.09	.1
Sustained	**-.12**	.06
Drugs	**-.21**	.05
Thresholds		
0% / 1-10%	**-2.25**	.29
1-10% / 11-20%	.49	.28
11-20% / 21-32%	**1.81**	.28
21-32% / 33%	**3.21**	.28
33% / >33%	**8.54**	.30

*In bold results which are statistically significant for a 95% confidence level.

We have seen that there are a number of factors that have an effect on the discount level received for a guilty plea outside of the stage of plea and whether the plea was entered at the first reasonable opportunity. This could show evidence that in some ways, current sentencing practice does not fully comply with the 2007 guideline. However, to assess the magnitude of this level of departure, it important to look at the effect size of the variables found problematic, and compare this to the

[5] This is known as the liberation hypothesis, see Ball (2006).

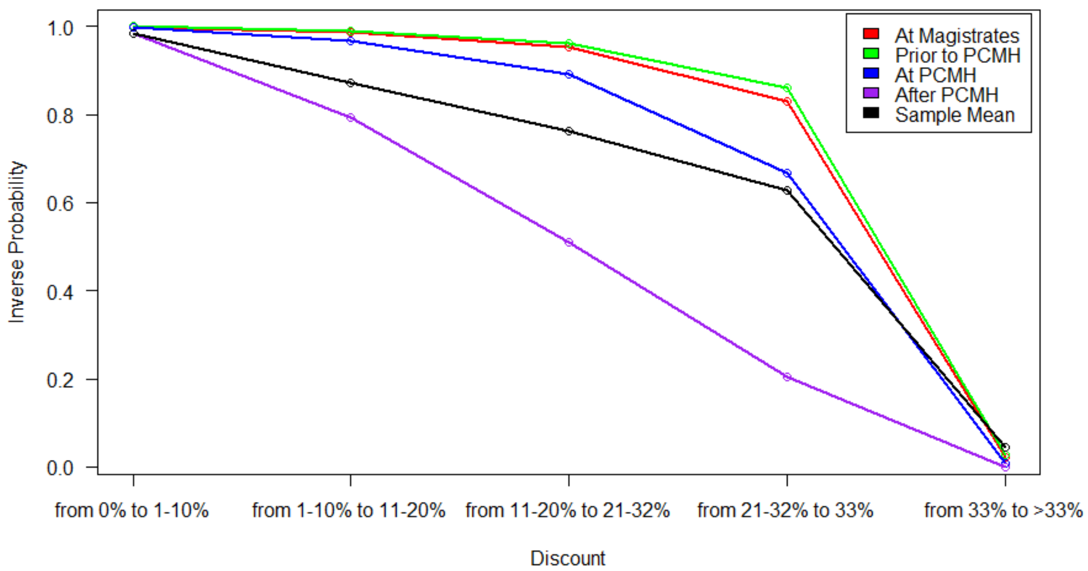
effect size of those variables that we expect should determine the different levels of discount received. I illustrate this comparison graphically in Figures 3 and 4. In these figures I plot the inverse probability of receiving a particular level of discount broken down by the stage at which the plea was entered (Figure 3). and when the effect of remorse and its interactions with the stage are included (Figure 4).[6]

Both in Figures 3 and 4, the black line represents the probability of making a transition across adjacent levels of discount for an average offender. For example, we can see that the probability of going from a 0% discount to at least a 1-10% is.98. As the level of discount increases, the probability of moving to the next adjacent level of discount steadily descends, with the probability of moving from 21-32% to 33% being .63. After this point, the probability of obtaining an even higher discount (>33%) descends drastically to .04.

In Figure 3, the other lines show these probabilities according to the stage at which the plea was entered. This shows that entering a plea at the magistrates court, prior to the PCMH, or at the PCMH, increases the probability of obtaining a higher than average discount, whilst pleading guilty after the PCMH[7] substantially reduces the probability of obtaining a higher level of discount.

Figure 4 shows the same probabilities after including the estimated effect of the mitigating factor remorse. Comparing this to Figure 3 gives us a sense of how the probability of receiving each of the levels of discount changes when remorse is present as mitigation. We can see that, compared to Figure 3, the different probabilities have been shifted up. In particular, we see that the red line, which depicts cases where the plea was entered at the magistrates court, shows more of an upwards shift than the other lines.

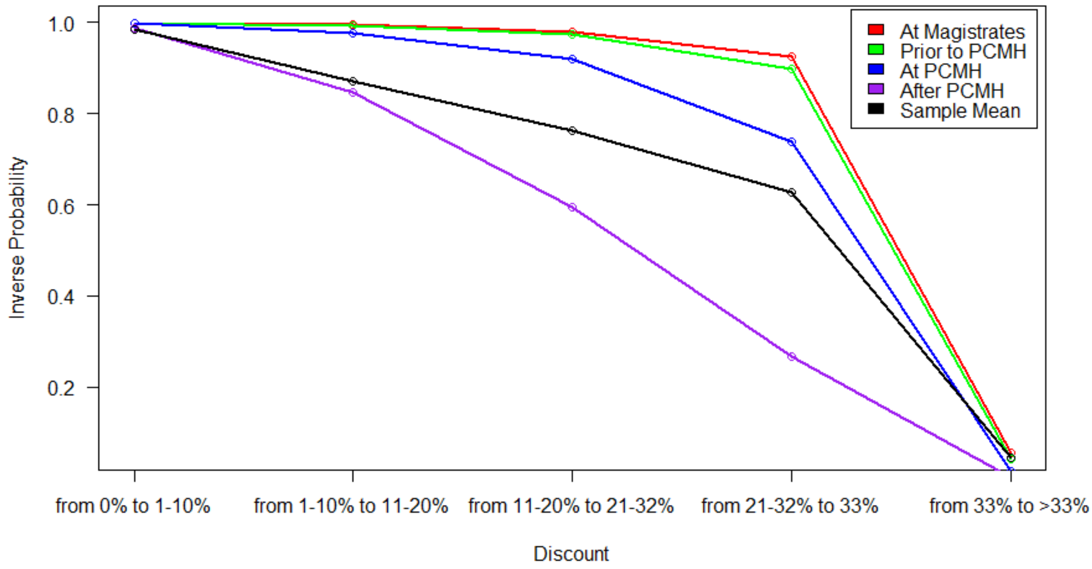
Figure 3. Inverse Cumulative Probabilities of Discounts by Stage



[6] Since the coefficients from Table 4 use a logit scale, we need to transform them to calculate the probabilities used in Figures 3 and 4. I do that by taking the exponential of the coefficients for the different thresholds, stage of plea, and remorse, over one plus that same exponential.

[7] Probabilities for guilty pleas entered the day of the trial have not been included because, for reasons of multicollinearity, they had to be left out of the model.

Figure 4. Inverse Cumulative Probabilities of Discounts by Stage and Remorse



In summary, although there is an observed effect for remorse, it is important to see this effect in relative terms. When comparing how much the curves shift when remorse is added to the model with the shift associated with adding the different stages of plea o the model, we can clearly see the effect of the former is minimal. Moreover, this conclusion could be generalised to the rest of the variables which I have found have an unexpected effect on the level of guilty plea discount (GBH with intent, affray, sustained attack, and under the influence), since none of their coefficients in the model is larger than the one found for the effect of remorse.

3.3. Binary logit models

One of the key assumptions underlying ordinal logit regression is that the relationship between each pair of outcome groups is the same. In other words, ordinal logit regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. In other words, the probability of moving up a discount category is the same, regardless of which discount category you start in. This is called the proportional odds assumption or the parallel regression assumption. Because the relationship between all pairs of groups is the same, there is only one set of coefficients. If this was not the case, we would need different sets of coefficients in the model to describe the relationship between each pair of outcome groups.

I specify two binary logit model, the first analysing the probability of obtaining 0% discount against all the other levels of discount, and the second looking at the probability of >33% discounts against all the other levels. These two models allow us looking into anomalous cases at the extreme of the guilty plea discount levels in more details, and exploring (partially) the possibility of unparallel effects.

Results from the first model are shown in Table 5 below. None of the previous interaction effects between remorse and stages of guilty plea were found significant so they were removed from both models. Almost any stage where guilty plea was entered granted a lower probability of obtaining a 0% discount than entering the plea at the day of the trial. The only exception is prior to PCMH, which is estimated with a much greater standard error than any other coefficient, hence, rendering the estimate statistically non-significant. This problem is probably derived from a reduced number of cases pleading guilty at that stage and getting a 0% discount. In addition first opportunity is not significant either. Therefore the timing of the guilty plea is not as important to determine whether offenders get 0% discounts as it was before when I was considering all the levels of discount. On the other hand, remorse has a significant negative effect, stronger than before. This represents that the double mitigation from remorse is more likely to occur in transitions from 0% to other levels of discount, than from one particular level to the next. Other variables which are now found significant are common assault, GBH, and previous convictions, with a positive effect, and sentence length with a negative effect.

Table 5. Binary Logit Model (0% vs Other Discounts)*

Variable	Coefficient	Standard Error
Constant	**-3.39**	.31
Magistrates	**-1.80**	.41
Prior to PCMH	-16.26	414
PCMH	**-2.26**	.27
After PCMH	**-1.27**	.29
First op.	-.09	.24
Log-length	**-.19**	.04
Female	.43	.25
GBH	**1.42**	.39
Intent	.45	.27
Common	**.72**	.25
Affray	.42	.24
Previous conv.	**.31**	.12
Remorse	**-.91**	.21
Carer	-.04	.33
Gang	-.33	.26
Vulnerable	-.46	.40
Pub. Worker	-.13	.47
Sustained	.10	.25
Drugs	.17	.20

*Results which are statistically significant at the 95% confidence level are shown in bold.

Results from the binary logit model looking at reductions of more than 33% are shown in Table 6 below. Here, the sliding scale is perfectly represented, as the probability of getting more than a 33% discount increases for every earlier stage. The double mitigation effect from remorse is also present, and the negative effect of drugs which we saw in the ordered logit model is now twice as big, which indicates that this erratic result is specially concentrated in the highest levels of discount. In addition, just like in the previous binary logit model, sentence length has a negative and GBH a positive effect. These two effects might be indicating that exceptional discounts greater than 33% are given in situations where the offence was categorised too severely. In such situations judges might be using guilt plea reductions as a tool to rebalance the appropriate sentence length without having to change the type of offence (from GBH to ABG, for example.

Table 6. Binary Logit Model (>33% vs Other Discounts)*

Variable	Coefficient	Standard Error
Constant	**-4.72**	.27
Magistrates	**2.81**	.24
Prior to PCMH	**2.45**	.25
PCMH	**.99**	.24
After PCMH	**.45**	.32
First op.	**.38**	.11
Log-length	**-.11**	.02
Female	-.15	.19
GBH	**1.06**	.22
Intent	.08	.15
Common	-.02	.18
Affray	-.01	.15
Previous conv.	.07	.07
Remorse	**.54**	.11
Carer	.15	.20
Gang	-.08	.16
Vulnerable	-.16	.20
Pub. Worker	-.21	.26
Sustained	-.07	.14
Drugs	**-.40**	.12

*Results which are statistically significant at the 95% confidence level are shown in bold.

4. Discussion

In this report I have empirically assessed the application of the 2007 guilty plea guideline. I have run an exploratory analysis using guilty pleas recorded by the CCSS in 2011, and focusing on the relation between levels of discount and the stages when the guilty pleas were entered. In a second part, I have used different models for discrete data to regress levels of discount on a broad set of explanatory variables offences of assault. Results point at a substantial degree of agreement of guilty plea sentences with the recommendations included in the 2007 guideline. In particular, the sliding scale recommended in the 2007 guilty plea guidelines was found to be the major factor determining levels of discount. However, I have also obtained results that show departures from the guidelines.

In the exploratory analysis I found that 6.8% of the guilty pleas were resolved with reductions higher than 33%. This is problematic since the guideline does not recommend levels of reduction beyond the 33%. In addition, I have found that the origin of such anomalous discounts varies substantially across Courts, and it is especially concentrated in the Courts of Nottingham and Derby.

From the models that I used I obtained additional evidence on departures from the guidelines. In the ordered logit regression model I found that after controlling for the stages when the guilty plea was entered, types of offences such as GBH with intent or affray were associated with higher discounts, while aggravating factors such as sustained assault or under the influence of drugs were associated with lower discounts. I also found that the mitigating factor of remorse has a positive and significant effect, showing evidence on double mitigation. Furthermore, this effect is twice as strong when the guilty plea was entered at the Magistrates. Finally, from the binary logit models that was used, I showed that the sliding scale is not being applied in the lowest level of discount (0%), and that the reduced discounts from drugs seem to apply only to cases that got more than a third discount.

References

Ball, J. (2006). Is It a Prosecutor's World? Determinants of Count Bargaining Decisions. Journal of Contemporary Criminal Justice, Vol. 22, pp. 241-260.

Roberts, J. (2013), Complying with Sentencing Guidelines: Latest Findings from the Crown Court Sentencing Survey. In Ashworth, A. and Roberts, J. (eds.) – *Structured Sentencing in England and Wales: Perspectives on the Definitive Guidelines*. Oxford: Oxford University Press.

Robertshaw, P. and Milne, A. (1992) The Guilty Plea Discount: Rule of Law or Role of Chance? *The Howard Journal*, Vol. 31, No 1, pp. 53-75.

Scott Long, J. (1997), *Regression Models for Categorical and Limited Dependent Variables*. SAGE: London.

Sentencing Council (2012) *Crown Court Sentencing Survey*. London: Sentencing Council of England and Wales.

Sentencing Guidelines Council (2004) *Reduction in Sentence for a Guilty Plea. Guideline.* Sentencing Guidelines Secretariat.

Sentencing Guidelines Council (2007) *Reduction in Sentence for a Guilty Plea. Definitive Guideline.* Sentencing Guidelines Secretariat.

Appendix I. Exploratory Analysis by Courts

Figure A1. Histogram of Guilty Plea Discounts by Court

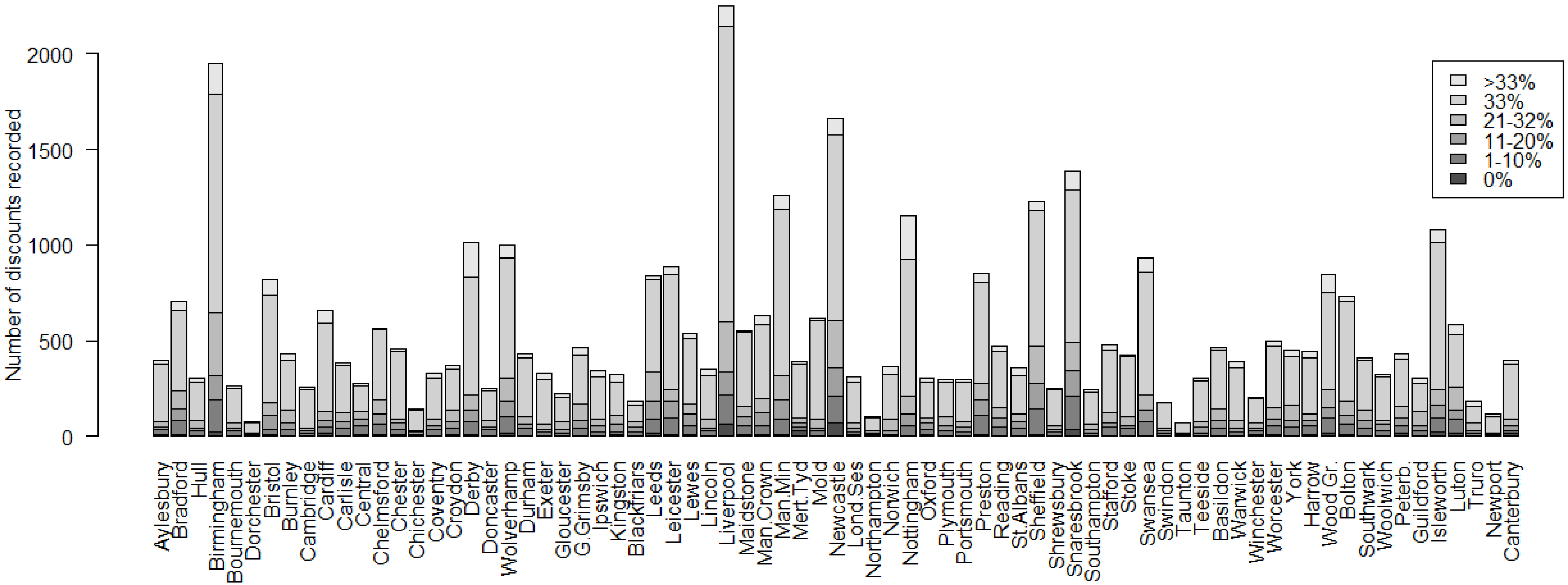


Figure A2. Histogram of Guilty Plea Stages by Court

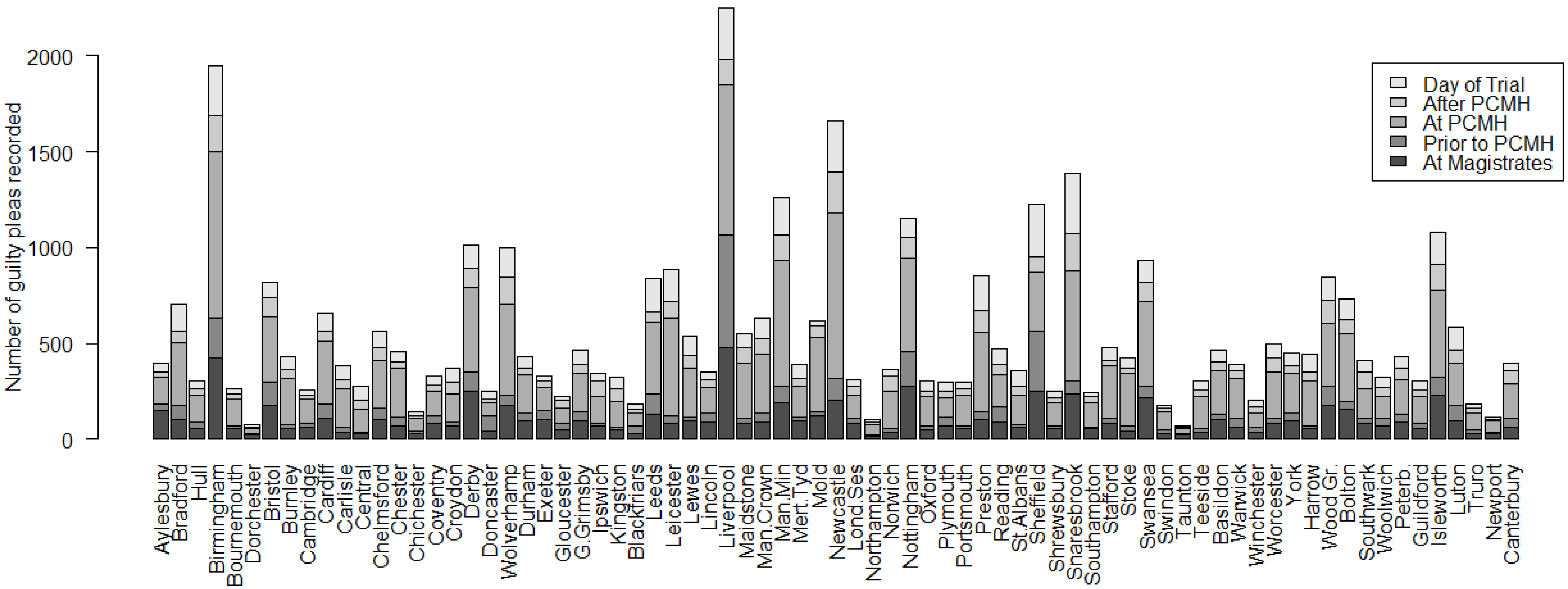
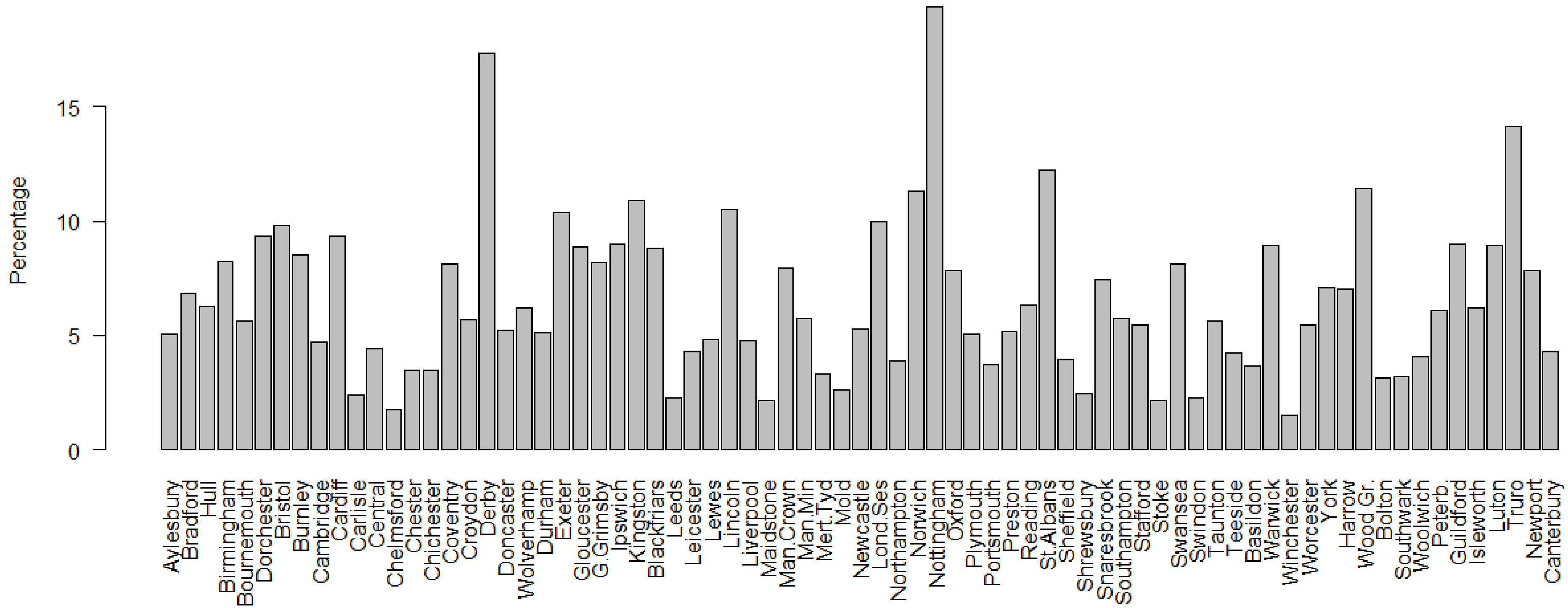


Figure A3. Histogram of the Percentage of >33% Discounts over Total Discounts by Court



Appendix II. Specification of the Ordered and Multinomial Logit Models

First I present the process used to specify the ordered logit model. For that I draw from Scott Long (1997)[8]. The ordinal regression model can be derived from a measurement model in which a latent variable Y^* ranging from -∞ to ∞ is mapped to an observed variable Y. The variable Y is thought of as providing incomplete information about an underlying Y^* according to the measurement equation:

$$y_i = m \quad \text{if} \quad \tau_{m-1} \le y_i^* < \tau_m \qquad\qquad \text{for } m = 1 \text{ to } J$$

The τ 's are known as thresholds. The extreme categories 1 and J are defined by open-ended intervals with $\tau_0 = -\infty$ and $\tau_j = \infty$. So, in my case, the guilty pleas reductions captured by the CCSS form, that is the observed y is related to y* according to the measurement model:

$$y_i = \begin{cases} 1 \Rightarrow 0\%.................if\ -\infty \le y_i^* < \tau_1. \\ 2 \Rightarrow 1-11\%..........if\ ..\tau_1 \le y_i^* < \tau_2.. \\ 3 \Rightarrow 11-20\%.........if\ ..\tau_2 \le y_i^* < \tau_3. \\ 4 \Rightarrow 21-32\%.........if\ ..\tau_3 \le y_i^* < \tau_4. \\ 5 \Rightarrow 33.3\%............if\ ..\tau_4 \le y_i^* < \tau_5. \\ 6 \Rightarrow > 33.3\%..........if\ ..\tau_5 \le y_i^* < \infty.. \end{cases} \tag{1}$$

The structural model is specified as follows,

$$y_i^* = x_i\beta + \varepsilon_i \tag{2}$$

where x_i and β are vectors representing the explanatory variables and their respective coefficients, and ε_i is a vector of residuals. However, since I do not observe y_i^* I cannot estimate equation 2 directly. The estimation process uses maximum likelihood under the assumption that the error term follows a logistic distribution with a mean of 0 and variance $\pi^2/3$. So, under this assumption, the cumulative distribution function is expressed as,

$$\Lambda(\varepsilon) = \frac{\exp(\varepsilon)}{1 + \exp(\varepsilon)} \tag{3}$$

Once the distribution of the errors is specified the probabilities of observing values of y given x can be computed as the area of the distribution of residuals between two thresholds (equation 1). I illustrate this process using the example of discounts of 1-11%, that is, when $y = 2$. Following the measurement model defined in equation 1, I observe $y = 2$ when y^* falls between τ_1 and τ_2. This can be more formally expressed as,

$$\Pr(y_i = 2 \mid x_i) = \Pr(\tau_1 \le y_i^* < \tau_2 \mid x_i)$$

substituting y^* from equation 1 and rearranging I obtain,

$$\Pr(y_i = 2 \mid x_i) = \Pr(\varepsilon_i < \tau_2 - x_i\beta \mid x_i) - \Pr(\varepsilon_i < \tau_1 - x_i\beta \mid x_i)$$

---

[8] Chapter 5 in Scott-Long (1997) covers ordered logit and probit models. Multinomial logit and probit models are included in Chapter 6.

this is equivalent to the conditional cumulative distribution function of residuals (equation 3) delimited by the two thresholds,

$$\Pr(y_i = 2 \mid x_i) = \Lambda(\tau_2 - x_i\beta_i) - \Lambda(\tau_1 - x_i\beta_i)$$

This process can be generalised for the rest of the levels of discount. So, for my model with 6 observed outcomes the final ordered model could be specified by the following system of equations,

$$\begin{cases} \Pr(y_i = 1 \mid x_i) = \Lambda(\tau_1 - x_i\beta_i) \\ \Pr(y_i = 2 \mid x_i) = \Lambda(\tau_2 - x_i\beta_i) - \Lambda(\tau_1 - x_i\beta_i) \\ \Pr(y_i = 3 \mid x_i) = \Lambda(\tau_3 - x_i\beta_i) - \Lambda(\tau_2 - x_i\beta_i) \\ \Pr(y_i = 4 \mid x_i) = \Lambda(\tau_4 - x_i\beta_i) - \Lambda(\tau_3 - x_i\beta_i) \\ \Pr(y_i = 5 \mid x_i) = \Lambda(\tau_5 - x_i\beta_i) - \Lambda(\tau_4 - x_i\beta_i) \\ \Pr(y_i = 6 \mid x_i) = 1 - \Lambda(\tau_5 - x_i\beta_i) \end{cases} \quad (4)$$

<u>The binary logit model</u>

The binary logit model can be understood as a simplified verion of the ordered logit where only two levels (and one threshold) are considered.

Like in the ordinal logit model, the binary logit models use a latent variable to map the observed values (0,1). So, we have the following simplified version of equation 1 for the two logit models that I run in this paper,

$$y_i = \begin{cases} 1 \Rightarrow 0\% \ldots\ldots\ldots\ldots if\ .y_i^* > \tau \\ 0 \Rightarrow other \ldots\ldots\ldots if\ .y_i^* \le \tau \end{cases}$$

$$y_i = \begin{cases} 1 \Rightarrow > 33\% \ldots\ldots\ldots if\ .y_i^* > \tau \\ 0 \Rightarrow other \ldots\ldots\ldots if\ .y_i^* \le \tau \end{cases}$$

Appendix III. Court Effects from the Ordered Logit Model*

Variable	Coefficient	Standard Error
Bradford	.41	0.31
Hull	**.91**	0.37
Birmingham	.17	0.29
Bournemouth	-.04	0.39
Dorchester	.93	0.55
Bristol	**1.21**	0.32
Burnley	.5	0.33
Cambridge	**.91**	0.43
Cardiff	**1.15**	0.31
Carlisle	.65	0.34
Central	-.19	0.45
Chelmsford	.01	0.33
Chester	**.93**	0.36
Chichester	.78	0.4
Coventry	.64	0.36
Croydon	.62	0.43
Derby	**1.33**	0.3
Doncaster	-.06	0.39
Wolverhampton	.45	0.3
Durham	**.73**	0.33
Exeter	.56	0.37
Gloucester	.77	0.4
Great Grimsby	.36	0.32
Ipswich	**1.47**	0.37
Kingston	.62	0.39
Blackfriars	-.39	0.46
Leeds	.19	0.3
Leicester	**.98**	0.3
Lewes	**.83**	0.33
Lincoln	**.88**	0.35
Liverpool	.12	0.28
Maidstone	-.005	0.34
Manchester Cr.	.5	0.32
Manchester Mi.	**.92**	0.29
Merthyr Tydfil	**.67**	0.32
Mold	.43	0.31
Newcastle	.45	0.28
Inner London	**1.52**	0.43
Northampton	.89	0.63
Norwich	**.86**	0.34
Nottingham	**1.17**	0.3
Oxford	.46	0.38
Plymouth	.01	0.34

Portsmouth	**1.1**	0.39
Preston	.58	0.3
Reading	.39	0.35
St. Albans	**.88**	0.38
Sheffield	.37	0.29
Shrewsbury	.08	0.4
Snaresbrook	.5	0.3
Southampton	**1.19**	0.41
Stafford	.25	0.32
Stoke-on-Trent	.43	0.33
Swansea	**.71**	0.3
Swindon	.53	0.48
Taunton	.51	0.55
Teesside	**1.21**	0.37
Basildon	-.26	0.34
Warwick	.57	0.35
Winchester	.08	0.41
Worcester	.18	0.34
York	.008	0.32
Harrow	**.77**	0.37
Wood Green	**.97**	0.33
Bolton	**.99**	0.31
Southwark	**1.25**	0.42
Woolwich	**1.21**	0.41
Peterborough	.09	0.35
Guildford	.35	0.38
Isleworth	**.79**	0.32
Luton	.54	0.34
Truro	-.1	0.4
Newport	.31	0.5
Canterbury	.35	0.38
Salisbury	1.07	1.32

*In bold results which are statistically significant for a 95% confidence level.