doi:10.1093/bjc/azaa030 BRIT. J. CRIMINOL

RE-ASSESSING THE CONSISTENCY OF SENTENCING DECISIONS IN CASES OF ASSAULT: ALLOWING FOR WITHIN-COURT INCONSISTENCIES

Ian Brunton-Smith, Jose Pina-Sánchez and Guangquan Li*

Empirical research has repeatedly focused on the potential existence of sentencing disparities. In particular, a growing number of studies have used multilevel models to quantify the extent that 'similar' offences are treated alike in different courts. This reliance on multilevel models has resulted in a natural focus on differences in the mean sentence awarded between courts, with the amount of within-group variability generally assumed to be the same in each court. In this paper, we show how multilevel models can be extended by allowing the magnitude of within-court differences to be different in each court. This provides a natural framework to connect between-court disparities with the sentencing differences that are thought to originate between judges operating within the same court, particularly in the absence of more fine-grained sentencing data about the judge residing in each case. Focusing specifically on cases of assault sentenced in 2011, we show that there are substantial differences in the range of sentences awarded in different courts, with the range almost twice as large in some courts. We also find that it is those courts that appear to show the traits of more homogeneous sentencing that sentence more harshly and that offences involving the presence of a weapon or evidence of good character and/or exemplary conduct were associated with higher levels of internal consistency.

Key Words: sentencing, consistency, location scale, multilevel model, within-court variability, Crown Court

Introduction

Applying judicial sentences in a consistent manner lies at the heart of the rule of law. Consistency in sentencing fosters predictability, transparency and legal certainty, which, in turn, enhances public trust and promotes the legitimacy of the Criminal Justice System (Roberts and Plesnicar 2015). Achieving consistency is, however, not an easy task and jurisdictions across the world are taking an increasingly active role. In England, the Sentencing Guidelines Council developed offence-specific guidelines designed to remind sentencers of their need to take into account relevant factors, such as the seriousness of the offence, the presence of aggravating and mitigating circumstances or whether a guilty plea was entered (see Sentencing Guidelines Council 2005). Consideration of these guidelines became mandatory following the Coroners and Justice Act (2009), demonstrating a clear commitment to promoting consistent practice.

*Ian Brunton-Smith, Department of Sociology, University of Surrey, Guildford GU2 7XH, UK; i.r.brunton-smith@surrey.ac.uk;Jose Pina-Sánchez, School of Law, University of Leeds, Leeds LS2 9JT, UK; Guangquan Li, Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Newcastle-upon-Tyne NE1 8ST, UK.

Academic studies examining sentencing consistency have generally concentrated on the degree of variability in sentence outcomes across courts (Britt 2000; Pina-Sánchez *et al.* 2017; Pina-Sánchez and Grech 2017) with studies highlighting the importance of the amount of resources allocated to a case, the number of cases processed in a court (Dixon 1995; Kautt 2002) and the importance of the 'localized sentencing culture' (Church 1982) that emerges within particular courts. Other studies have considered broader geographic units, including counties (Fearn 2005; Haynes 2011; Johnson 2005), districts (Feldmeyer and Ulmer 2011; Johnson *et al.* 2008) and police forces (Mason *et al.* 2007), arguing that the socio-economic and demographic characteristics of the area where the court is located can also shape sentencing behaviour. To quantify the extent of disparities in sentencing for 'similar' offences (e.g. adjusting for known differences between cases), studies have increasingly used multilevel models, with the standard random intercepts model providing a 'built-in' coefficient quantifying the degree and significance of illegitimate between-cluster disparities—the intra-class correlation (Kautt 2002). For example, looking at the sentences imposed for offences of common assault from the Crown Court (England and Wales) and controlling for the main legal factors defining these cases, Pina-Sánchez and Linacre (2013) found an intra-class correlation of 0.018, suggesting that 1.8 per cent of the variation in sentence outcomes was the result of court differences. Multilevel models are so well suited to this task that the UK sentencing council recommended in August 2019 that these models are adopted as standard to correctly study sentencing consistency.

Other studies have concentrated more closely on differences *within* courts, arguing that this is where there is the largest opportunity for disparities to manifest (Anderson *et al.* 1999; Anderson and Spohn 2010; Johnson 2006; 2014). This recognizes that, as a highly complex cognitive activity, sentencing is subject to potential biases that can influence the judicial decision-making process. Judges working in some courts may be more willing to deviate from sentencing guidelines and tailor their sentence to the features of a specific offence (Johnson 2005), leading to higher than average sentencing variability in those courts. Other courts may be characterized by a clear set of informal rules and behaviours, including a shared sense of what constitutes an acceptable sentence for particular offences (Eisenstein *et al.* 1988). But moving the focus of attention to differences *within* courts typically necessitates more fine-grained sentencing data that includes the ability to uniquely identify each sentencing judge, something that the judiciary has been reluctant to accommodate because of the risks it may pose to judicial independence (Gertner 2012). As a result, whilst clearly important to a complete understanding of sentencing practice, most empirical studies have been unable to effectively account for it when using a multilevel approach.

In this study, we outline an extension to the standard multilevel modelling approach that enables researchers to directly incorporate differences *within* courts. This is made possible by relaxing the assumption of a common within-cluster variance and, instead, allowing it to vary across courts. With this extended model, we are able to simultaneously quantify differences in the average sentence awarded by each court and the amount of unique sentencing variation in each court. Importantly, whilst this approach recognizes that the presence of a different composition of judges within different courts will lead some courts to exhibit more variation in sentencing outcomes than others, it does not require the ability to uniquely identify individual judges. It also allows us to assess whether legitimate sentencing factors—harm and culpability

measures, aggravating and mitigating circumstances, prior convictions and guilty plea reductions—are more (or less) reliably incorporated into sentencing decisions than others. We demonstrate our approach using data on cases of assault that received a prison sentence in England and Wales in 2011, restricting the analysis to cases dealt with in the first 6 months following the release of a comprehensive set of guidelines for sentencing assault cases (Sentencing Council 2011).

Understanding Sentencing Disparities Within and Between Courts

The principles of fairness, proportionality and individualization necessitate a certain level of sentencing variation, with features of the offence and the offender leading to more or less lenient sentencing outcomes. However, the extent that these legal factors are considered in a uniform fashion by different sentencers is not always clear, with studies highlighting ways that judges' own attitudes and beliefs may lead to more or less emphasis being placed on particular case features (Anderson and Spohn 2010). And beyond these known legal features that may differ between cases, scholars have identified substantial variation in sentence outcomes both *within* and *between* courts. This includes evidence of differences in practice towards offenders of different races, gender and social class (Albonetti 1997; Steffensmeier and Demuth 2000; Mustard 2001; Everett and Wojtkiewicz 2002; Stacey and Spohn 2006; Doerner and Demuth 2010; King and Johnson 2016; Lightowlers 2019), the number of cases processed in a court (Dixon 1995; Johnson 2006) and the socio-economic and demographic characteristics of the area where the court is located (Ulmer and Johnson 2004; Fearn 2005; Johnson 2006; Johnson *et al.* 2008).

Differences between courts have been attributed to the existence of 'court cultures' (Church 1982), with a number of studies pointing to the relationships between the variety of court actors (prosecutors, judges, defence lawyers, etc.) that can lead to the emergence of distinctive local legal cultures (Eisenstein *et al.* 1988). These differences in the culture of each court may include a collective understanding of what constitutes acceptable and unacceptable behaviour, specific rituals in relation to practice within the courtroom and the presence of a 'them and us' relationship between legal professionals and court users (Kirby 2017). Local court cultures are thought to influence the sentencing decisions of judges working within those courts. For example, local understandings of the 'going rate' for particular offences may become embedded within the working practices of particular courts, with all sentencing judges tending to align their own sentencing practice with the broader activities of other court actors. And, in addition to having a direct role in shaping the sentencing outcomes of the court, the set of informal norms, expectations and working practices that make up the local court culture are also thought to be the primary conduit through which broader structural features of courts—including their size and number of cases dealt with, as well as the availability of sufficient resources to deal with cases—may influence sentences (Johnson *et al.* 2008; Ulmer and Johnson 2004). For example, larger courts may have a stronger court culture because they are less influenced by external pressures and more able to develop their own set of distinctive sentencing practices. Conversely smaller courts may feel more pressure to 'tow the line' and align their sentencing practice more closely with existing guidelines.

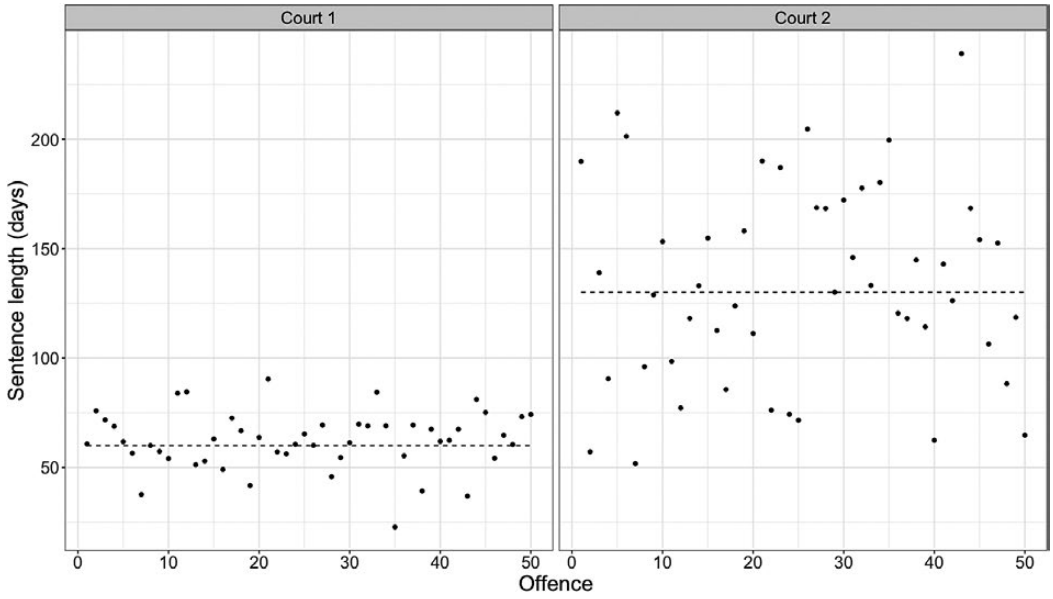


FIG. 1. Graphical representation of location-scale model for sentence outcomes across courts.

The recent emphasis on disparities in the typical sentence awarded between different courts has obscured another source of inconsistencies that is often unacknowledged, those arising *within* courts. This is demonstrated graphically in Figure 1, which plots hypothetical standardized sentence outcomes (e.g. custodial sentence lengths) for 100 cases dealt with by two different courts. The horizontal lines denote the average sentence outcome for each court, with a higher average sentence in Court 2. But, in addition to the higher average sentence awarded in court 2, there is also a greater degree of *variability* in sentence outcomes in this court when compared to Court 1. These differences in the degree of variability (or consistency) between courts is another measure of overall sentencing consistency that previous assessments have failed to capture.

Considering disparities that manifest *within* courts, the unique combination of attributes and actors within each sentencing location is likely to result in diverse sentence outcomes. Studies rooted in symbolic interactionism (Blumer 1969) recognize that sentencing is a combination of close adherence to established sentencing guidelines, the specific features of a particular case and the differential actions of individual sentencers. In particular, scholars have identified a set of 'focal concerns' (Steffensmeier *et al.* 1998)—offender culpability, the need to protect the community and resource constraints limiting the provision of adequate punishment—that shape sentencing decisions. The relative weight given to each of these focal concerns is influenced by each sentencer's own unique set of attitudes, beliefs and background experiences, which, in turn, results in substantial differences in sentence outcome between different sentencers.

In the absence of sufficient background information about a case, or when faced with pressures due to a lack of time and resources, sentencers are thought to make use of a 'perceptual shorthand' (Hawkins 1981) when assessing cases. Here, they may draw on non-legal contextual factors when making their decision in addition to legitimate

offence characteristics and offending history. This can lead to the existence of sentencing biases. Extra-legal factors have also been identified as important modifiers of consistent sentencing practice within courts. Work has also drawn attention to characteristics of the sentencer, with Johnson (2006) finding that minority judges were generally less punitive than their white counterparts (see also Steffensmeier and Britt 2001; Welch *et al.* 1988) and younger judges typically more punitive than older judges. The gender (Kritzer and Uhlman 1977; Gruhl *et al.* 1981; Steffensmeier and Hebert 1999), seniority (Spohn 1990; Pina-Sánchez *et al.* 2019), religion (Myers and Talarico 1987) and political orientation (Tiede *et al.* 2010; Lim *et al.* 2016) of the judge may also be important.

There is also good reason to anticipate that the localized sentencing practices within particular courts will also result in differences in the extent of sentence variation within different courts. This was first highlighted by Brantingham (1985) who, exploring disparities at the judge level, warned that even relatively small differences in the average sentence between clusters of analysis (what she referred to as 'first-order disparities') may be masking more substantial variations in the range of sentences utilized within particular clusters for similar offences called 'second-order disparities'. There is clear scope within existing theoretical work on court cultures to expect that the amount of variation in sentences being awarded will be different in different courts, with some courts adopting a narrower range of sentences and other courts making use of a much wider range of sentence outcomes for similar offences. In particular, courts that adopt less variable practice may hold stronger normative understandings of what constitutes an 'appropriate' sentence, with the sentencing decisions of different judges likely to closely match one another. Conversely, courts with a weaker sentencing culture may be expected to exhibit more sentencing variability, with individual judges giving more emphasis to their own interpretations of the focal concerns, manifesting in more notable differences in the final sentences awarded by different judges. Court cultures may also play a role in determining the extent that individual sentencing guidelines are adhered to, with courts holding differing understandings of the extent that deviations from sentencing guidelines are appropriate (Johnson 2005). And some courts may have developed a lower threshold of acceptability for departing from sentencing guidelines, making it less problematic for actors in some courts to move to the extremes of the guideline.

Second-order disparities may also manifest as differences in the weight that particular case features are given by judges operating in some courts. Here, we might anticipate that sentencing factors that are well understood by the majority of judges will be more consistently accounted for in sentencing decisions. Other legitimate case factors may be incorporated less consistently, perhaps, because they are less closely aligned with judges' 'focal concerns' or represent more ambiguous features of the case. Case features that have only recently been explicitly incorporated into sentencing guidelines and are not yet embedded into judges' decision-making may also be associated with less consistent sentencing outcomes.

Extending Multilevel Models to Incorporate Differences in the Amount of Within-Court Consistency

To measure differences in the range of sentences that are awarded for similar offences in different courts, we make use of the location-scale extension to standard multilevel models (Hedeker *et al.* 2008; Leckie *et al.* 2014; Brunton-Smith *et al.* 2018). This approach relaxes the common residual variance assumption (homoscedasticity), instead, allocating each court its own residual variance by re-specifying the Level-1 residual variance as a function of covariates and an additional random effect.

Let y_{ij} denote a continuous sentence outcome, the custodial sentence length for offence i $(i = 1, \ldots, n_j)$, tried within court j $(j = 1, \ldots, \mathcal{J})$. The standard two-level random-intercept model for examining sentencing consistency across courts can then be written as

$$y_{ij} = \mathbf{x}_{ij}^{'} \beta + u_j + e_{ij} \tag{1}$$

where $\mathbf{x}_{ij}$ is the intercept and a vector of offence-specific covariates that may be related to sentence outcomes, including those identified in sentencing guidelines. Characteristics of offenders and courts can also be included, enabling more direct estimation of the causes of any illegitimate variations in sentencing practice (Pina-Sánchez 2015). The random effect u_j represents differences in sentence outcome in court j, and e_{ij} is the residual, capturing any remaining differences in sentence outcomes. The random effect and residual are assumed mutually independent, independent of the covariates, and normally distributed with zero means and constant variances, $u_j \sim \mathcal{N}\left(0, \sigma_u^2\right)$ and $e_{ij} \sim \mathcal{N}\left(0, \sigma_e^2\right)$. The between-court random effects variance σ_u^2 captures the degree of sentence inconsistency between courts, adjusted for any known determinants of differential sentences at the offence level. The residual variance σ_e^2 measures the variability in sentence outcomes unexplained by the model.

Equation (1) assumes constant residual variance, constraining σ_e^2 to be equal across all courts. This assumption is relaxed by specifying an auxiliary log-linear equation for the within-court variance as a function of covariates and an additional court random effect.[1] This equation can be written as

$$\ln\left(\sigma_{eij}^2\right) = b + \mathbf{w}_{ij}^{'} \gamma + u_j^{[2]} \tag{2}$$

where $\ln\left(\sigma_{eij}^2\right)$ denotes the log of the now heterogeneous within-court variance and $\mathbf{w}_{ij}$ is a vector of (optional) offence-level covariates. An additional court random effect, $u_j^{[2]}$, is also included. The '[2]' superscript distinguishes this random effect from the standard court-level random effect, which is now denoted $u_j^{[1]}$. The two sets of court random effects are assumed bivariate normally distributed.

$$\begin{pmatrix} u_j^{[1]} \\ u_j^{[2]} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u[1]}^2 & \\ \sigma_{u[1]u[2]} & \sigma_{u[2]}^2 \end{pmatrix}\right) \tag{3}$$

[1]The log link function ensures that the within-court variance takes only positive values.

The variance–covariance matrix summarizes how courts differ not just in average sentences (summarized by $\sigma^2_{u[2]}$). In other words, we allow each court to have a different degree of sentencing consistency in addition to our measure of differences in average sentence between courts. Information is also available about the relationship between the mean and variance in each court ($\sigma_{u[1]u[2]}$). Court-specific estimates of the within-court variance can then be estimated as

$$\sigma^2_{eij} = \exp\left(b + \mathbf{w}_{ij}{}' \gamma + u_j^{[2]}\right) \tag{4}$$

Finally, having generated a court-specific estimate of the within-court variance, σ^2_{eij}, we can combine this with the court-specific differences in the mean sentence awarded, $\hat{u}_j^{[1]}$, to provide an overall estimate of the degree of sentencing consistency for each court. Here, we treat $\hat{u}_j^{[1]}$ as an estimate sentencing bias in each court and use the mean squared error (MSE) formula to combine this with the within-court variance:

$$\text{MSE}_j = \sqrt{\left(\hat{u}_j^{[1]}\right)^2 + \sigma^2_{eij}} \tag{5}$$

Data

To assess the extent that courts differ in their levels of between- and within-court variation, we use data from Crown Courts Sentencing Survey (CCSS). Fielded between October 2010 and March 2015 by the Sentencing Council for England and Wales to assist them in the design and monitoring of their guidelines, the CCSS required judges to complete offence-specific questionnaires immediately after each sentence was passed in the Crown Courts (Ashworth and Roberts 2013). In addition to the specific offence, judges also provided information on the offender, the circumstances surrounding the offence and factors informing the final sentence outcome. This included offender criminal history, the range of aggravating and mitigating circumstances associated with the case and whether a guilty plea reduction was made. Collecting this information directly from judges at the time of sentence resulted in a unique data set characterized by a combination of high coverage and depth (Roberts and Hough 2015). We use data collected in the second half of 2011 covering all cases of assault, including completed and attempted cases of actual and grievous bodily harm (hereafter, denoted as ABH and GBH), common assault and affray sentenced in courts in England and Wales. This period immediately followed the introduction of new sentencing guidelines for cases of assault, allowing us to leverage information on the full range of aggravating and mitigating circumstances surrounding each case, as well as initial sentence ratings of harm and culpability. The 2011 CCSS achieved an overall response rate of 61 per cent (Sentencing Council 2012) and there was little evidence of any systematic relationship between court-specific response rates and offence type or sentence severity.

Sentence length

We restrict our focus to the subset of 46 per cent of violent cases that resulted in a custodial sentence (31 per cent of the remaining cases received a suspended sentence, 19 per cent a community order and 3 per cent a conditional discharge or fine), a final analytic sample of 1,663 cases sentenced within 74 courts. Cases received a mean sentence length of 766 days, with a range from 4 days to more than 30 years. Due to the extremely skewed distribution of the original sentence length, this is log-transformed to more closely approximate a normal distribution.

Legal and extra-legal sentencing factors

Even restricting our analysis to cases of assault, it is still likely that many of the observed differences in sentence outcome within courts, and by association many of the observed differences in the variability of sentence outcomes between courts, are the result of differences in the characteristics of the specific cases sentenced. We use statistical controls to adjust for the confounding influence of these legitimate sentencing variations across cases. We include a categorical variable to distinguish between the main forms of assault in our model, with 396 cases of GBH, 859 of ABH, 110 common assaults and 298 cases of affray. From the sentencing guidelines, we also include details of the full range of harm and culpability factors that judges accounted for, as well as the presence of any aggravating and mitigating circumstances, and the size of any sentence adjustment for a guilty plea. Finally, we control for the number of prior convictions that the offender may have.[2] Of course, the possibility remains that other relevant legal factors are unobserved. If that is the case, any apparent differences in variability between and within courts picked up by the model should be interpreted as indicative of potential inconsistencies rather than providing conclusive proof.

To explore the extent that sentencing decisions are also influenced by extra-legal features, we also include controls for offender gender and age (the only offender characteristics collected in the CCSS), as well as the volume of cases that were processed in each court in 2011, and a measure of the pressure on available court resources (the number of days Crown Court rooms were in use as a percentage of the total number of working days in the year).[3]

Analytic strategy

To undertake a comprehensive assessment of disparities within and between courts, we estimate three models. Model 1 does not include any covariates and acts as an initial baseline estimate of disparities within and between courts. Model 2 includes the full

[2]Judges also record an initial estimate of offence seriousness on the CCSS; however, this information was incomplete for a large number of sentenced cases. Whilst this may result in further legitimate differences in sentence outcome, our models include the full set of harm and culpability factors that judges use when determining the level of seriousness, so it is unlikely to lead to substantial biases.

[3]Data on the volume of cases processed was obtained from the following government website: https://www.gov.uk/government/publications/data-pack-tables-for-crime-tender-2015. Data on the relative use of court rooms was obtained from a National Office Audit report (https://www.nao.org.uk/report/administration-of-the-crown-court/). This last variable was only available for 39 of the Crown Court locations; the missing 35 were mean imputed.

range of statistical controls for known case characteristics in the response equation (Equation 1 above). This ensures that our estimates of the within- and between-court consistency are adjusted as far as possible for legitimate sources of variation in sentence outcomes. No covariates are included in the within-court variance equation. Finally, in Model 3, we include the same set of case characteristics in the within-court variance equation (Equation 2 above), providing an indication of whether some sentencing factors are associated with more (or less) consistent sentencing practice than others. For example, a negative effect of a guilty plea reduction would imply that offences where a guilty plea was entered are being dealt with in a more consistent fashion (there is less variability in sentence outcome) than offences where a guilty plea was not entered. Conversely, a positive effect would indicate that guilty pleas resulted in less consistent outcomes (or more variability in sentence outcome). This model also adds basic offender details and information about each court into the response and the scale equation to explore other potential sentencing biases that may be influencing sentencing decisions.

The restriction of our analysis to those offenders sentenced to custody leaves open the possibility that our results will be affected by selection bias if the decision whether or not to incarcerate also informs the likely sentence length. The included set of covariates represents the most comprehensive account of the sentencing process for assault available, reducing the potential impact of selection bias. Nevertheless, it remains possible that other, unobserved characteristics are influencing the decision whether or not to incarcerate and also what sentence length to award (Bushway *et al.* 2007). To further mitigate the potential impacts of selection bias, our models also include the Inverse Mills Ratio (Heckman 1976).[4]

Estimation

Models are estimated using Markov Chain Monte Carlo (MCMC) methods as implemented in the WinBUGS statistics package (Lunn *et al.* 2012). All models are run using two MCMC chains with dispersed starting values and uninformative prior distributions for all parameters. Chains are estimated with a burn-in period of 3,000 iterations and a monitoring period of 20,000 iterations. Visual assessments of the parameter chains and standard MCMC convergence diagnostics suggest that these periods are sufficiently long to generate robust parameter summaries (Spiegelhalter *et al.* 2002). Reported results present the means and 95 per cent uncertainty intervals of the 40,000 monitoring iterations pooled across the two chains. These are generally equivalent to effect sizes and confidence intervals in standard frequentist models.

[4]Specifically, we include the hazard rate derived from a probit model predicting the probability of receiving a custodial sentence. Bushway *et al.* (2007) advocate the use of exclusion criteria to reduce the problem of inflated standard errors. The assault guidelines represent a useful guide for appropriate variables to exclude from our sentence length models, with judges only supposed to use initial harm and culpability factors to determine the starting point for sentencing decisions and not the final sentence to be awarded. However, Pina-Sánchez *et al.* (2018) recently demonstrated that these factors were also influencing the final sentencing decision. As a result, we choose to include the full set of controls in the selection model and model of interest. Initial assessments of the correlations between the Inverse Mills Ratio and included covariates suggest that our models are unlikely to be inefficient. Furthermore, our substantive conclusions remain consistent whether the Inverse Mills Ratio is included or excluded.

Table 1. *Multilevel location-scale model results for cases of assault (July–December 2011)—no covariates in the within-court variance equation*

	Model 1: unconditional model			Model 2: legitimate sentencing factors		
	95% interval			95% interval		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%
Within-court variance (intercept)	0.00	−0.10	0.09	−1.40	−1.55	−1.26
Between-court variance (intercept)	6.14	6.07	6.21	5.83	5.72	5.94
GBH				1.64	1.56	1.73
Common assault				−1.25	−1.36	−1.15
Affray (other)				−0.02	−0.09	0.06
Deliberately causes more harm than is necessary for the commission of the offence				0.14	0.04	0.24
Intention to commit more serious harm than actually resulted from the offence				0.05	−0.05	0.15
Leading role in group or gang				−0.02	−0.11	0.07
Offence motivated by hostility towards the victim based on the victim's age, sex or gender identity				0.13	−0.11	0.36
A significant degree of pre-meditation				0.12	0.05	0.19
Offence racially or religiously aggravated				0.27	0.07	0.48
Deliberate targeting of vulnerable victim				0.02	−0.06	0.10
Use of weapon or weapon equivalent				0.11	0.06	0.16
Injury that is serious in the context of the offence				0.16	0.10	0.22
Sustained or repeated assault on the same victim				0.11	0.05	0.17
Victim is particularly vulnerable because of personal circumstances				0.16	0.09	0.23
Lack of pre-meditation				−0.11	−0.19	−0.04
Mental disorder or learning disability, where linked to the commission of the offence				0.05	−0.19	0.30
Greater degree of provocation				−0.25	−0.37	−0.12
Excessive self-defence				−0.16	−0.32	0.00
Subordinate role in group/gang				−0.15	−0.28	−0.02
Injury that is less serious in the context of the offence				−0.16	−0.22	−0.09
Abuse of a position of trust or power				−0.07	−0.23	0.08
Offence against public sector worker or those providing a service to the public				−0.01	−0.11	0.09
Offence committed on bail				0.01	−0.10	0.13
An attempt to conceal or dispose of evidence (new assault)				0.18	−0.03	0.39
Victim forced to leave home (DV cases)				0.04	−0.14	0.22

Table 1. *Continued*

	Model 1: unconditional model			Model 2: legitimate sentencing factors		
		95% interval			95% interval	
	Mean	2.5%	97.5%	Mean	2.5%	97.5%
Established evidence of community impact (new assault)				0.00	−0.25	0.26
Failure to respond to warnings or concerns expressed by others about the offender's behaviour (new assault)				−0.07	−0.24	0.10
Failure to comply with current court orders				−0.01	−0.10	0.08
Gratuitous degradation of victim				0.09	−0.04	0.23
Location of the offence				0.05	−0.01	0.11
Offence committed whilst on licence				0.03	−0.11	0.17
Ongoing effect upon the victim				0.05	−0.02	0.11
Presence of others, including relatives, especially children or partner of the victim				−0.03	−0.10	0.03
Previous violence or threats to the same victim				0.04	−0.04	0.12
Timing of the offence				−0.05	−0.13	0.03
Commission of offence whilst under the influence of alcohol or drugs				−0.03	−0.08	0.03
Steps taken to address addiction				−0.04	−0.17	0.08
Age and/or lack of maturity				−0.14	−0.23	−0.04
Good character and/or exemplary conduct				−0.05	−0.16	0.05
Isolated incident				0.04	−0.07	0.15
Lapse of time since the offence				0.05	−0.18	0.27
Serious medical condition				0.09	−0.17	0.34
Mental disorder				−0.10	−0.28	0.08
No previous convictions or no recent/relevant convictions				−0.09	−0.16	−0.01
Sole or primary carer for dependant relatives				−0.04	−0.22	0.14
Genuine remorse				−0.03	−0.10	0.03
Single blow				−0.15	−0.24	−0.07
No guilty plea reduction				−0.02	−0.20	0.17
1–10% reduction				0.06	−0.04	0.16
11–20% reduction				−0.05	−0.15	0.05
21–32% reduction				−0.14	−0.24	−0.05
33% or more reduction				−0.12	−0.18	−0.05
PCs taken into account (1–3)				0.04	−0.03	0.11
PCs taken into account (4–9)				0.18	0.10	0.26
Inverse Mills Ratio				−0.09	−0.17	−0.02
Random effects						
Between-court variance	0.04	0.01	0.08	0.01	0.00	0.02
Covariance	0.02	0.00	0.06	−0.03	−0.06	−0.01
Within-court variance	0.05	0.01	0.12	0.22	0.12	0.37
Intra-court correlation	0.04	0.01	0.07	0.03	0.01	0.06

Results

It is first instructive to examine the unconditional model results, where no correction has been applied for the various case features embedded in the sentencing guidelines (Table 1, Model 1). Here, we find evidence of a small but significant amount of variability between courts in the average sentence awarded for cases of assault, with the intra-court correlation indicating that this accounts for around 4 per cent of the total variation in sentence outcomes. Of course, a substantial proportion of this variation is the result of legitimate differences between cases. When the full set of sentencing features is included in Model 2, it accounts for more than two-thirds of the differences observed between courts (with the between-court variance falling from 0.04 to 0.01). However, some differences between courts remain, with the intra-cluster correlation showing that courts still account for around 3 per cent of the total variability. This points to the existence of small but significant differences in the likely sentence received, which depend on where an offender is sentenced. The included predictors generally operate in the expected direction in Model 2. Cases of GBH receive substantially longer sentences on average, whilst common assault cases are usually awarded a shorter sentence. Features of the case indicating higher levels of culpability are also associated with longer sentence lengths—with the largest increases for racially and religiously motivated offences (approximately 107 days on average[5]), followed by cases where there is a serious injury (59 days) or the victim was particularly vulnerable (58 days). Longer sentences were also awarded in cases where prior convictions had also been considered (a 66 days premium when 4 or more prior convictions were accounted for). Conversely, shorter sentences were awarded for cases where the offender was judged less culpable—in particular, when a high degree of provocation was involved (75 days less) or the offender was a subordinate part of a group of offenders (47 days less). Shorter sentences were also awarded when less serious injury was incurred (50 days less) and where guilty pleas have been factored into sentencing decisions. There is also some evidence that mitigating factors are incorporated into sentencing decisions, but there are no significant aggravating case factors. This is likely due to the relatively small sample size available and the low prevalence of some of these factors.

Importantly, courts also differ in the amount of within-court variation, with a significant within-court variance of 0.05 in Model 1, increasing to 0.20 in Model 2. The increase suggests that once known differences between cases are accounted for, the differences between courts in the *remaining* within-court variation become clearer. This can be clearly seen when looking at the model-estimated court residuals (Figure 2), which we have transformed into the metric of sentence days for ease of comparison.[6] The two plots on the right show the size of the within-court standard deviation in sentence length for each court, distinguishing between the unconditional estimate (top) and the adjusted within-court standard deviation when account is taken of legitimate sentencing differences (bottom). In both cases, there is evidence that courts exhibit different amounts of within-court variation. In the unconditional model, we observe

[5]Calculated as $e^{(\beta_0+\beta_x)} - e^{(\beta_0)}$. This provides us with an approximate value for the change in sentence outcome in days for each included covariate when compared against the intercept (a case of ABH with no mitigating or aggravating circumstances, no prior convictions and no details of any guilty plea being factored into the sentencing decision).

[6]Calculated using the formula $(e^{\sigma^2_{ij}} - 1) \times e^{2\mu_j + \sigma^2_{ij}}$, where μ_j is the court-specific mean of the logged sentence length and σ^2_{ij} is the within-court variance of the logged sentence length.

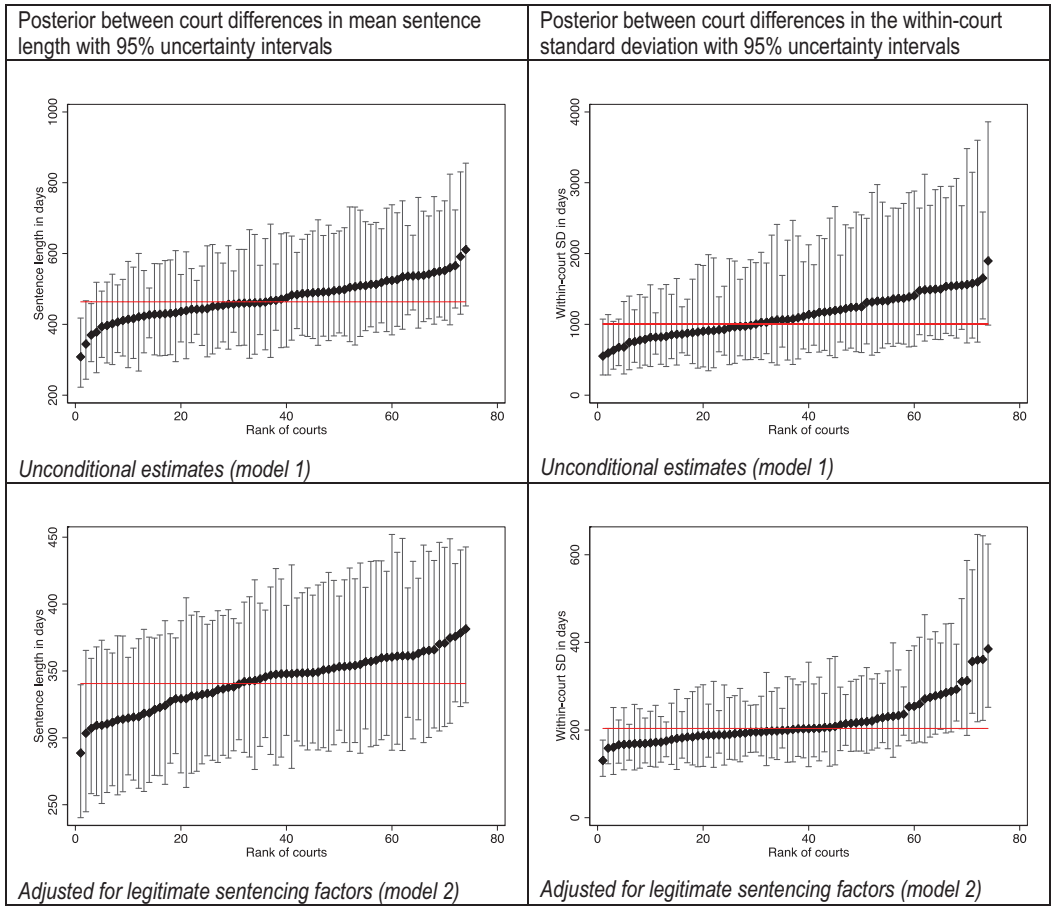
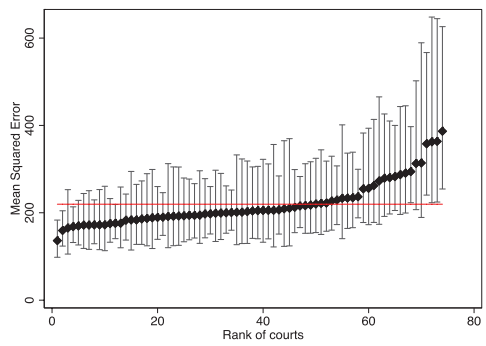


FIG. 2. Predicted court-specific mean sentence length and standard deviation from the empty location-scale model (Model 1) and accounting for legitimacy sentencing factors (Model 2).

an overall standard deviation of roughly 1,000 days but with an estimated within-court standard deviation that is lower in some courts (551 days in the court with the lowest standard deviation to the left of the plot) and around 50 per cent larger in other courts (reaching 1,895 days for the court with the highest standard deviation to the right of the plot). Whilst the overall size of the within-court standard deviation has reduced substantially when known sentence characteristics are taken account of (from approximately 1,000 days to around 200), the relative differences between courts have become more apparent. In particular, there is a clear cluster of courts to the right of the graph with standard deviations approximately twice as large as the average, with sentences varying by approximately 400 days on average. The two plots on the left give the equivalent estimates of the mean sentence length[7] for each court. We also note a significant negative covariance between the location and scale residuals (−0.03, or a correlation of −0.63), suggesting that it is those courts that tend to impose longer sentences, on average, that are also the least variable in their sentencing practice.

[7]Calculated as $e^{\beta_0+u_j^{[1]}}$.



Rank	Court name	Estimated MSE
1	Merthyr Tydfil	136.2
2	Leeds	159.4
3	Great Grimsby	165.1
4	Mold	168.3
5	Stafford	169.8
70	Teesside	314
71	St. Albans	357.9
72	Hull	362.4
73	Derby	363.4
74	Durham	386.7

FIG. 3. Estimated court-specific MSE and 95% uncertainty intervals, along with details of top and bottom ranking courts (based on MSE).

Combining both sources of variation into a single metric, the estimated MSE for each court, provides a straightforward method to rank the sampled courts based on their overall degree of sentencing consistency (Figure 3). Here, a higher score indicates more inconsistency, with courts penalized both for awarding a mean sentence that is notably higher (or lower) than the expected average sentence based on the composition of cases and for less consistent (e.g. more variable) sentencing practice. Here, we highlight the five courts with the highest and lowest MSE, with the scores for Derby and Durham (the highest scoring courts) almost twice as large as the scores for Merthyr Tydfil and Leeds (the lowest scoring courts).

Finally, Model 3 (Table 2) adds the full list of control variables into the variance equation (Model 3: sentence variability). Here, we find clear evidence that cases of GBH are sentenced more consistently, on average, than cases of ABH, common assault and affray (as evidenced by the negative effect of −0.77 and uncertainty intervals that do not cross 0). We also find moderate evidence that some legitimate sentencing factors are associated with more consistent sentencing practice (as evidenced by negative coefficients with uncertainty intervals that do not cross 0). Judges are generally more consistent in the application of a sentence uplift when a weapon was involved, and reduce sentences more consistently when the defendant exhibited good character and/or exemplary conduct. More consistency is also observed when the defendant was believed to play a leading role in a gang or failed to comply with a current court order. However, the substantive implication of these effects is less clear, with the equivalent effects on the typical sentence awarded (Model 3: mean sentence) being close to zero. This suggests that, whilst there may be more uncertainty amongst judges in how to apply sentence adjustments for these factors, they do not lead to substantial alterations to the overall profile of sentences (as many judges increase the sentence as reduce the sentence). Similarly, whilst judges seem to be more consistent in applying adjustments for offences against public sector workers, this also does not translate into different sentence outcomes.

Model 3 also includes details of defendant gender and age, as well as the volume of cases dealt with by the court and the level of pressure on court resources. Sentences

TABLE 2. *Full multilevel location-scale model results for cases of assault (July–December 2011)*

	Model 3: mean sentence			Model 3: Sentence variability		
	95% interval			95% interval		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%
Within-court variance (intercept)				−0.96	−1.28	−0.62
Between-court variance (intercept)	5.86	5.76	5.96			
GBH	1.61	1.55	1.68	−0.77	−1.02	−0.52
Common assault	−1.28	−1.39	−1.17	−0.10	−0.42	0.24
Affray (other)	−0.02	−0.09	0.06	−0.02	−0.26	0.21
Deliberately causes more harm than is necessary for the commission of the offence	0.10	0.01	0.19	0.20	−0.12	0.54
Intention to commit more serious harm than actually resulted from the offence	0.06	−0.02	0.13	-0.15	−0.47	0.19
Leading role in group or gang	−0.01	−0.07	0.06	−0.47	−0.78	−0.16
Offence motivated by hostility towards the victim based on the victim's age, sex or gender identity	0.05	−0.16	0.27	−0.26	−1.06	0.64
A significant degree of pre-meditation	0.13	0.06	0.19	0.04	-0.21	0.29
Offence racially or religiously aggravated	0.28	0.11	0.45	−0.20	−0.88	0.54
Deliberate targeting of vulnerable victim	0.01	−0.06	0.08	−0.24	−0.51	0.03
Use of weapon or weapon equivalent	0.09	0.04	0.14	−0.32	−0.50	−0.14
Injury that is serious in the context of the offence	0.18	0.12	0.23	0.03	−0.18	0.23
Sustained or repeated assault on the same victim	0.10	0.05	0.15	−0.11	−0.30	0.08
Victim is particularly vulnerable because of personal circumstances	0.15	0.08	0.22	0.09	−0.15	0.34
Lack of pre-meditation	−0.10	−0.17	−0.03	0.10	−0.14	0.35
Mental disorder or learning disability, where linked to the commission of the offence	0.03	−0.26	0.30	0.30	−0.53	1.19
Greater degree of provocation	−0.24	−0.36	−0.13	0.08	−0.32	0.48
Excessive self-defence	−0.18	−0.32	−0.04	−0.23	−0.74	0.31
Subordinate role in group/gang	−0.12	−0.24	0.00	0.08	−0.33	0.51
Injury that is less serious in context of the offence	−0.11	−0.17	−0.06	−0.17	−0.39	0.03
Abuse of a position of trust or power	−0.08	−0.22	0.05	−0.25	−0.74	0.27
Offence against public sector workers or those providing a service to the public	−0.05	−0.15	0.06	0.32	0.01	0.65
Offence committed on bail	0.06	−0.03	0.15	−0.36	−0.74	0.03
An attempt to conceal or dispose of evidence (new assault)	0.16	0.00	0.31	−0.08	−0.82	0.71
Victim forced to leave home (DV cases)	0.08	−0.06	0.22	−0.35	−0.91	0.28
Established evidence of community impact (new assault)	0.00	−0.17	0.18	−0.74	−1.55	0.15
Failure to respond to warnings or concerns expressed by others about the offender's behaviour (new assault)	0.00	−0.16	0.15	−0.05	−0.60	0.54
Failure to comply with current court orders	0.02	−0.06	0.09	−0.42	−0.70	−0.12

Table 2. *Continued*

	Model 3: mean sentence			Model 3: Sentence variability		
	95% interval			95% interval		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%
Gratuitous degradation of victim	0.11	0.00	0.22	0.10	−0.32	0.57
Location of the offence	0.04	−0.01	0.10	−0.11	−0.32	0.10
Offence committed whilst on licence	0.03	−0.08	0.14	−0.41	−0.89	0.10
Ongoing effect upon the victim	0.06	0.00	0.12	0.03	−0.19	0.25
Presence of others, including relatives, especially children or partner of the victim	−0.03	−0.09	0.03	0.07	−0.14	0.28
Previous violence or threats to the same victim	0.02	−0.05	0.10	0.10	−0.18	0.38
Timing of the offence	−0.05	−0.12	0.02	0.08	−0.20	0.36
Commission of offence whilst under the influence of alcohol or drugs	−0.03	−0.07	0.02	−0.23	−0.42	−0.04
Steps taken to address addiction	−0.06	−0.17	0.06	−0.07	−0.49	0.38
Age and/or lack of maturity	−0.10	−0.17	−0.02	−0.14	−0.46	0.20
Good character and/or exemplary conduct	−0.09	−0.17	−0.01	−0.47	−0.81	−0.10
Isolated incident	0.09	−0.01	0.18	0.28	−0.07	0.65
Lapse of time since the offence	0.13	−0.11	0.37	0.24	−0.51	1.05
Serious medical condition	0.07	−0.11	0.26	−0.58	−1.49	0.40
Mental disorder	−0.09	−0.28	0.08	0.40	−0.18	1.04
No previous convictions or no recent/relevant convictions	−0.07	−0.13	−0.01	−0.13	−0.37	0.12
Sole or primary carer for dependant relatives	−0.01	−0.14	0.12	−0.27	−0.93	0.43
Genuine remorse	−0.05	−0.11	0.01	0.16	−0.05	0.39
Single blow	−0.11	−0.19	−0.03	0.11	−0.15	0.39
No guilty plea reduction	−0.02	−0.17	0.13	−0.31	−0.87	0.30
1–10% reduction	0.06	−0.03	0.15	−0.14	−0.46	0.19
11–20% reduction	−0.04	−0.13	0.05	0.10	−0.21	0.43
21–32% reduction	−0.16	−0.25	−0.07	0.07	−0.23	0.38
33% or more reduction	−0.13	−0.19	−0.07	−0.15	−0.35	0.04
PCs taken into account (1–3)	0.04	−0.03	0.11	0.02	−0.21	0.28
PCs taken into account (4–9)	0.11	0.04	0.19	0.00	−0.27	0.28
Inverse Mills Ratio	−0.12	−0.20	−0.05	0.14	−0.07	0.35
Female	−0.10	−0.20	−0.01	−0.02	−0.41	0.38
Age (centred)	0.00	0.00	0.01	0.00	−0.01	0.01
Volume of offences	0.00	−0.04	0.03	0.00	−0.13	0.14
Pressure on court resources	0.00	−0.03	0.03	−0.10	−0.23	0.03
Random effects						
Between-court variance	0.01	0.00	0.01			
Covariance	−0.02	−0.04	0.00			
Within-court variance	0.13	0.05	0.26			

tend to be shorter, on average, for women than men, and there is moderate evidence that older offenders receive longer sentences. However, we find no evidence that sentences are more (or less) consistently applied to women than men or that the amount of sentence variability differs as a function of defendant age. And the differences between courts in both the mean and variance of sentence length are not well accounted for by the volume of cases dealt with by each court or the level of pressure on court resources.

Discussion

We have outlined a new approach to more accurately measure the different dimensions of sentencing variation that are present within and across courts using recent extensions to multilevel modelling that allow a different amount of within-group variation in each cluster. Our results present a comprehensive picture of the multiple sources of variation in sentencing practice both within and across courts. We find that considerable sentencing variation exists when considering cases of assault, although this is mostly the result of legitimate differences between the offences being sentenced—including the type of assault, harm and culpability factors and prior convictions. But even when these legitimate case features have been accounted for, substantial variation remains both between and within courts.

The magnitude of these differences is considerable. We find that the typical range of sentences awarded may be almost twice as large in some courts. Consequently, the failure to incorporate within-court disparities in existing research on sentencing consistency means that we are seriously underestimating the true extent of the problem of inconsistencies in practice across courts. Importantly, this result takes account of a wide range of legitimate deviations in sentence outcomes. Therefore, we have good reason to anticipate that these results are reflective of real inconsistencies in practice across courts. This is an important advance over existing studies, enabling novel insights about the degree of consistency with which similar cases are treated in the same court.

Existing studies have highlighted the potential effect that court culture can have in sentencing decisions (Church 1985; Hucklesby 1997), and our results are consistent with the contention that different courts have developed their own unique set of practices in determining a suitable range of sentence outcomes to be awarded for particular types of offence. Where the magnitude of the within-court variance is large, this may be indicative of court cultures where deviations from sentencing guidelines are deemed more acceptable (Johnson 2005), with judges encouraged to place greater emphasis on their own interpretations of the relative importance of case characteristics. The large disparities in some courts may also be reflective of a more heterogeneous workforce in these courts, signalling important differences in the practice of different judges. To understand this effect further, and the impact that it might have in sentencing consistency, information on the sentencing judge should be made routinely available alongside data on courts. Our models also included general measures of the volume of cases dealt with by each court and the pressures on sentencing resources, as well as the gender and age of defendants. However, these were unrelated to sentence outcomes (in the mean or variance equation). The failure to connect within-court sentencing disparities directly to known features of cases leaves open the possibility that other features of courts and cases may be influencing the decisions of the courts.

We find that consistency is correlated with severity. It is those courts that tend to award a longer sentence that also exhibit less variable practice. This is consistent with Allen's (2016) suggestion that an unquestioning focus on increased consistency via sentencing guidelines may be promoting substantial sentence inflation, where the drive for more consistency is associated with a greater number of upward adjustments to sentences than downwards adjustments. Our results suggest this may be a court-specific phenomenon, with some courts making use of a narrow band of more severe penalties. Specifically, it is those courts that appear to show the traits of more homogeneous sentencing (a stronger court culture) that sentence more harshly, whereas those where judges behave more freely (a less clearly defined culture) appear to be more lenient.

This research also sheds light on the ways that courts adapt their sentences based on assessments of culpability and harm, as well as known mitigating and aggravating circumstances. In particular, we found that the presence of a weapon and evidence of good character and/or exemplary conduct were associated with higher levels of internal consistency. Cases of GBH, which attracted sentences nearly 3 years longer, on average, than cases of ABH, were also sentenced more consistently. However, we found little evidence that other factors are associated with more or less variable sentencing decisions. Whilst this may, in part, reflect the comparatively small sample of cases used in the current analysis, it also mirrors the recent findings of Pina-Sánchez (2015) and Pina-Sánchez and Linacre (2013) who showed that there were minimal variations in the mean impact of mitigating and aggravating circumstances across courts (estimated using a random slopes extension to standard multilevel models). This suggests that courts may be generally consistent in their use of these factors in the Crown Court.

Whilst we believe that this approach provides an important new direction for sentencing scholars, there are some important limitations that should be addressed in future studies. First, in order to mitigate the effects of legitimate variations in sentence outcome, we restricted our focus to a specific group of offences: assault offences. This increases the confidence that we have in the estimates of within- and between-court consistency but leaves open the possibility that our results are reflective of the unique features of cases of violent crime cases rather than pointing to the existence of broader inconsistencies in practice between courts. Future studies should examine the extent that within-court disparities are evident across other sentence types to better understand the generality of our findings. Second, despite generating robust estimates of the magnitude of the within-court variance parameter, our ability to accurately pinpoint differences between particular courts was more limited. Repeating this analysis on a larger number of cases within each court would enable more precise estimates of the relative consistency in different courts. Third, although we were able to incorporate a number of aggravating and mitigating factors to capture legitimate variations in sentencing, it remains possible that other important case features are missing. This may mean that our estimates of the degree of inconsistency between and within courts are biased upwards.

Conclusion

Multilevel modelling has rightly become the dominant research framework for the empirical identification of inconsistencies in sentencing practice. There is no doubt that

the capacity to model simultaneously cluster-level and offence-level variations whilst controlling for legitimate differences between cases makes multilevel models an ideal tool for exploring consistency. In particular, the standard random intercepts model provides a 'built-in' coefficient quantifying the degree and significance of illegitimate between-cluster disparities. But the ready availability of such a useful measure, coupled with the fact that most data on sentencing only allows for the identification of the court where sentences occur and not the judge involved in each particular sentencing decision, runs the risk of placing the emphasis squarely on descriptions of consistency in sentencing based on between-court differences in the typical sentence awarded. This misses important disparities taking place within courts, presenting an incomplete picture of the overall degree of consistency.

In the absence of routinely available data on judges, we believe that the location-scale model represents a viable way forward that will enable researchers to provide insight into the effects of differences in the unique combination of attributes and actors within each sentencing location. The MSE can then be used as an effective summary measure capturing both forms of inconsistency simultaneously, providing a useful tool to monitor sentencing practice. This may be particularly useful to Sentencing Commissions and Sentencing Councils in charge of monitoring the degree of compliance with their guidelines. For example, the England and Wales Sentencing Council dedicates resources to keep an open channel of communications with courts to learn about their views on the guidelines and clarify their application. Given the need to rationalize resources, we would recommend the Council to prioritize contacts not only on those courts systematically showing harsher or more lenient sentencing but also those where the internal disparities are also high, i.e. those with a higher MSE.

Funding

This work was supported by the National Centre for Research Methods (RIS14241/06).

References

Albonetti, C. A. (1997), 'Sentencing Under the Federal Guidelines: Effects of Defendant Characteristics, Guilty Pleas, and Departures on Sentence Outcomes for Drug Offenses, 1991–1992', *Law Society Review,* 31:789–22.

Allen, R. (2016), 'The Sentencing Council for England and Wales: Brake or Accelerator on the Use of Prison?', available online at http://www.transformjustice.org.uk/wp-content/uploads/2016/12/TJ-DEC-8.12.16-1.pdf. Accessed 09 April 2020.

Anderson, A. and Spohn, C. (2010), 'Lawlessness in the Federal Sentencing Process: A Test for Uniformity and Consistency in Sentence Outcomes', *Justice Quarterly,* 27: 362–93.

Anderson, J., Kling, J. and Stith, K. (1999), 'Measuring Inter-Judge Sentencing Disparity: Before and After the Federal Sentencing Guidelines', *Journal of Law and Economics,* 42: 271–307.

Ashworth, A. and Roberts, J. V. (2013), *Sentencing Guidelines: Exploring the English Model*. Oxford University Press.

Blumer, H. (1969), *Symbolic Interactionism: Perspective and Method*. University of California Press.

Brantingham, P. (1985), 'Sentencing Disparity: An Analysis of Judicial Consistency', *Journal of Quantitative Criminology,* 1: 281–305.

Britt, C. L., (2000), 'Social Context and Racial Disparities in Punishment Decisions', *Justice Quarterly,* 17: 707–32.

Brunton-Smith, I., Sturgis, P. and Leckie, G. (2018), 'How Collective is Collective Efficacy? The Importance of Consensus in Judgments About Community Cohesion and Willingness to Intervene', *Criminology,* 56: 608–37.

Bushway, S. D., Johnson, B. D. and Slocum, L. A. (2007), 'Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology', *Journal of Quantitative Criminology,* 23: 151–78.

Church Jr, T. (1982), 'The "Old and the New" Conventional Wisdom of Court Delay', *The Justice System Journal,* 7: 395–412.

Church, T. W. (1985), 'Examining Local Legal Culture', *American Bar Foundation Research Journal,* 1, 449–518.

Coroners and Justice Act. (2009), available online at http://www.legislation.gov.uk/ukpga/2009/25/contents. Accessed 09 April 2020.

Dixon, J. (1995), 'The Organizational Context of Criminal Sentencing', *American Journal of Sociology,* 100: 1157–98.

Doerner, J. K. and Demuth, S. (2010), 'The Independent and Joint Effects of Race/Ethnicity, Gender, and Age on Sentencing Outcomes in US Federal Courts', *Justice Quarterly,* 27: 1–27.

Eisenstein, J., Flemming, R. and Nardulli, P. (1988), *The Contours of Justice: Communities and Their Courts.* Little, Brown and Co.

Everett, R. S. and Wojtkiewicz, A. (2002), 'Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing', *Journal of Quantitative Criminology,* 18: 189–211.

Fearn, N. E. (2005), 'A Multilevel Analysis of Community Effects on Criminal Sentencing', *Justice Quarterly,* 22: 452–87.

Feldmeyer, B., and Ulmer, J. T. (2011), 'Racial/Ethnic Threat and Federal Sentencing', *Journal of Research in Crime and Delinquency,* 48: 238–70.

Gertner, N. (2012), 'TRAC Data on Federal Sentencing: Judge Identifiers, TRAC, and a Perfect World', *Federal Sentencing Reporter,* 25: 46–9.

Gruhl, J., Spohn, C. and Welch, S. (1981), 'Women as Policymakers: The Case of Trial Judges', *American Journal of Political Science,* 25: 308–22.

Hawkins, D. (1981), 'Causal Attribution and Punishment for Crime', *Deviant Behavior,* 1: 207–30.

Haynes, S. H. (2011), 'The Effects of Victim-Related Contextual Factors on the Criminal Justice System', *Crime and Delinquency,* 57: 298–328.

Heckman, J. J. (1976), 'The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables', *Annals of Economic and Social Measurement,* 5: 475–92

Hedeker, D., Mermelstein, R. J. and Demirtas, H. (2008), 'An Application of a Mixed-Effects Location Scale Model for Analysis of Ecological Momentary Assessment (EMA) Data', *Biometrics,* 64, 627–34.

Hucklesby, A. (1997), 'Court Culture: An Explanation of Variations in the Use of Bail by Magistrates' Courts', *Howard Journal,* 36: 129–45.

Johnson, B. D. (2005), 'Contextual Disparities in Guidelines Departures: Courtroom Social Contexts, Guidelines Compliance, and Extralegal Disparities in Criminal Sentencing', *Criminology,* 43: 761–96.

—— (2006), 'The Multilevel Context of Criminal Sentencing: Integrating Judge-and County-Level Influences', *Criminology,* 44: 259–98.

—— (2014), 'Judges on Trial: A Reexamination of Judicial Race and Gender Effects Across Modes of Conviction', *Criminal Justice Policy Review,* 25, 159–84.

Johnson, B. D., Ulmer, J. T. and Kramer, J. H. (2008), 'The Social Context of Guidelines Circumvention: The Case of Federal District Courts', *Criminology,* 46: 737–83.

Kautt, P. M. (2002), 'Location, Location, Location: Interdistrict and Intercircuit Variation in Sentencing Outcomes for Federal Drug-Trafficking Offenses', *Justice Quarterly,* 19: 633–71.

King, D. K. and Johnson, B. D. (2016), 'A Punishing Look: Skin Tone and Afrocentric Features in the Halls of Justice', *American Journal of Sociology,* 122: 90–124.

Kirby, A. (2017), 'Effectively Engaging Victims, Witnesses and Defendants in the Criminal Courts: A Question of "Court Culture"?', *Criminal Law Review,* 12: 949–68.

Kritzer, H. M. and Uhlman, T. (1977), 'Sisterhood in the Courtroom: Sex of Judge and Defendant in Criminal Case Disposition', *Social Science Journal,* 14: 77–88.

Leckie, G., French, R., Charlton, C. and Browne, W. (2014), 'Modeling Heterogeneous Variance-Covariance Components in Two-Level Models', *Journal of Educational and Behavioral Statistics,* 39: 307–32.

Lightowlers, C. (2019), 'Drunk and Doubly Deviant? The Role of Gender and Intoxication in Sentencing Assault Offences', *British Journal of Criminology.* 59: 693–717.

Lim, C. S. H., Silveira, B. S. and Snyder, J. M. (2016), 'Do Judges' Characteristics Matter? Ethnicity, Gender, and Partisanship in Texas State Trial Courts', *American Law and Economics Review,* 18: 302–57.

Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2012), *The BUGS book: A Practical Introduction to Bayesian Analysis.* Chapman and Hall–CRC.

Mason, T., de Silva, N., Sharma, N., Brown, D. and Harper, G., (2007), *Local Variation in Sentencing in England and Wales.* Ministry of Justice.

Mustard, D. B. (2001), 'Racial Ethnic and Gender Disparities in Sentencing: Evidence from the US Federal Courts', *Journal of Law Economics,* 44: 285–314.

Myers, M. A. and Talarico, S. M. (1987), *The Social Contexts of Criminal Sentencing.* Springer.

Pina-Sánchez J. (2015), 'Defining and Measuring Consistency in Sentencing', in Roberts J. V., eds., *Exploring Sentencing Practice in England and Wales*, 76–92. Palgrave Macmillan.

Pina-Sánchez, J., Brunton-Smith, I., Grech, D. and Sferopoulos, D. (2019), 'Exploring the Origin of Sentencing Disparities in the Crown Court: Using Text Mining Techniques to Differentiate between Court and Judge Disparities', *Social Science Research,* 84: 1023–43.

Pina-Sánchez, J., Brunton-Smith, I. and Li, G. (2018), 'Mind the Step: A More Insightful and Robust Analysis of the Sentencing Process in England and Wales Under the New Sentencing Guidelines', *Criminology & Criminal Justice,* Online First. doi:1748895818811891.

Pina-Sánchez, J. and Grech, D. C. (2017), 'Location and Sentencing: To What Extent Do Contextual Factors Explain between Court Disparities?', *British Journal of Criminology,* 58: 529–49.

Pina-Sánchez, J., Lightowlers, C. and Roberts, J. (2017), 'Exploring the Punitive Surge: Crown Court Sentencing Practices Before and After the 2011 English Riots', *Criminology & Criminal Justice,* 17: 319–39.

Pina-Sánchez, J. and Linacre, R. (2013), 'Sentence Consistency in England and Wales Evidence From the Crown Court Sentencing Survey', *British Journal of Criminology,* 53: 1118–38.

Roberts, J. V. and Hough M. (2015), 'Empirical Sentencing Research: Options and Opportunities', in Roberts J. V., eds., *Exploring Sentencing Practice in England and Wales*, 1–17. Palgrave Macmillan.

Roberts, J. and Plesnicar, M. (2015), 'Sentencing, Legitimacy, and Public Opinion', in G. Meszko and J. Tankebe, eds. *Legitimacy and Criminal Justice in Europe*, 33–51. Springer.

Sentencing Council (2011), 'Assault Definitive Guideline', available online at https://www.sentencingcouncil.org.uk/offences/crown-court/item/assault-occasioning-actual-bodily-harm-racially-religiously-aggravated-abh/. Accessed 09 April 2020.

—— (2012), 'Crown Court Sentencing Survey 2011 Annual Publication', available online at https://www.sentencingcouncil.org.uk/publications/item/crown-court-sentencing-survey-annual-publication-2011-full-report/. Accessed 09 April 2020.

Sentencing Guidelines Council (2005), 'Manslaughter by Reason of Provocation: Guideline', available online at http://www.sentencingcouncil.org.uk/wp-content/uploads/Manslaughter_by_Reason_of_Provocation.pdf. Accessed 09 April 2020.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002), 'Bayesian Measures of Model Complexity and Fit (With Discussion)', *Journal of the Royal Statistical Society Series B,* 64: 583–639.

Spohn, C. (1990), 'Decision Making in Sexual Assault Cases: Do Black and Female Judges Make a Difference?', *Women and Criminal Justice,* 2: 83–105.

Stacey, A. M. and Spohn, C. (2006), 'Gender and the Social Costs of Sentencing: An Analysis of Sentences Imposed on Male and Female Offenders in Three US District Courts', *Berkeley Journal of Criminal Law,* 11: 43–76.

Steffensmeier, D. and Britt, Ch. (2001), 'Judge's Race and Judicial Decision Making: Do Black Judges Sentence Differently?', *Social Science Quarterly,* 82: 749–65.

Steffensmeier, D. and Demuth, S. (2000), 'Ethnicity and Sentencing Outcomes in US Federal Courts: Who Is Punished More Harshly?', *American Sociological Review,* 65:705–29.

Steffensmeier, D. and Hebert, Ch. (1999), 'Women and Men Policymakers: Does the Judge's Gender Affect the Sentencing of Criminal Defendants?', *Social Forces,* 77: 1163–96.

Steffensmeier, D., Ulmer, J. and Kramer, J. (1998), 'The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male', *Criminology,* 36: 763–98.

Tiede, L., Carp, R. and Manner, K. (2010), 'Judicial Attributes and Sentencing-Deviation Cases: Do Sex, Race, and Politics Matter?' *Justice System Journal,* 31: 249–72.

Ulmer, J. T. and B. D. Johnson. (2004), 'Sentencing in Context: A Multilevel Analysis', *Criminology,* 42: 137–77.

Welch, S., Combs, M. and Gruhl, J. (1988), 'Do Black Judges Make a Difference?', *American Journal of Political Science,* 32: 126–36.