Sentencing Council

On the Measurement of Consistency in Sentencing

Index

- Consistency in Sentencing
- Methods
- Discussion

Consistency in Sentencing

Sentencing Council

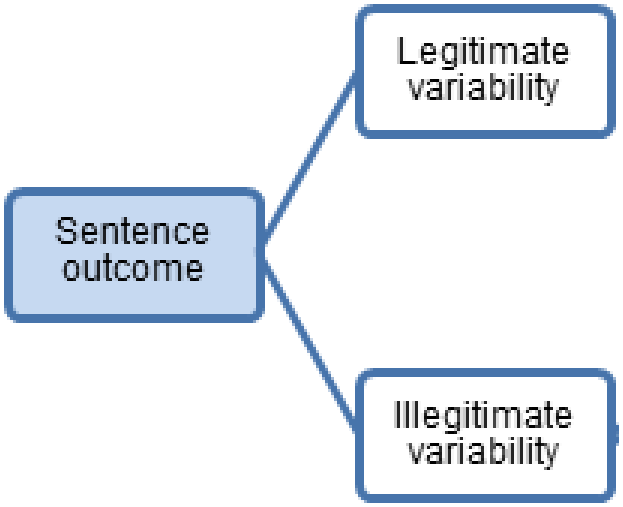
- The most commonly used definition is very intuitive: “consistency as the extent to which like cases are treated alike”.

- The operationalisation of the concept is normally a little more complicated.

- To facilitate doing so we widened our perspective to consider the mechanisms behind variability in sentencing.

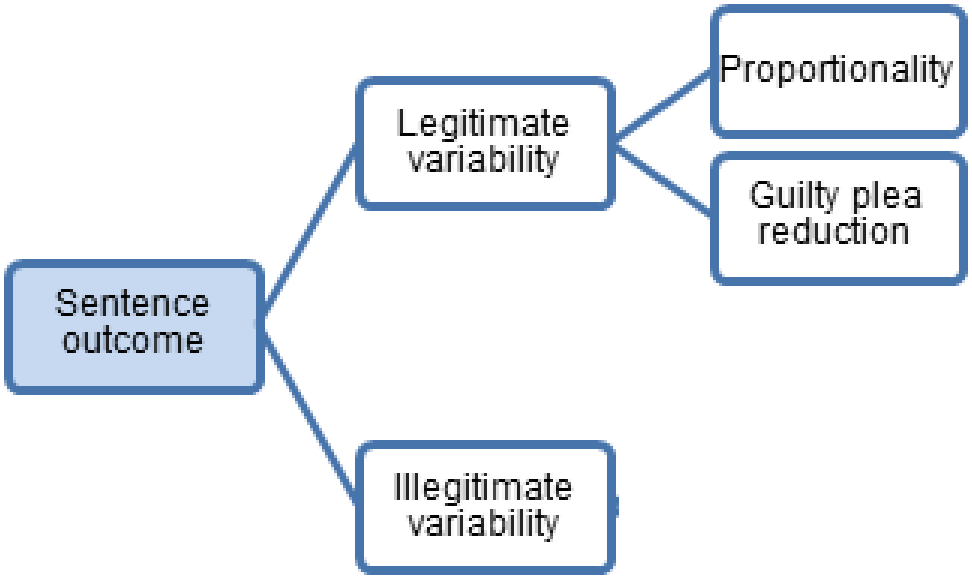
Consistency in Sentencing

Sentencing Council



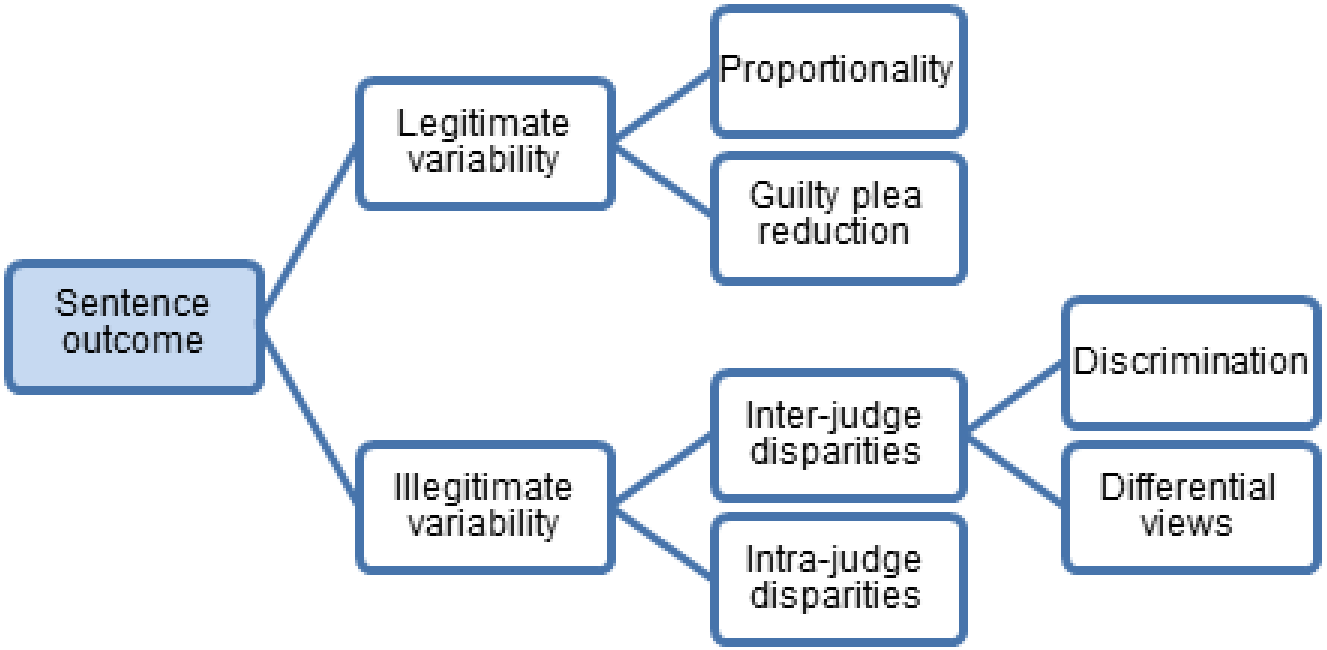
Consistency in Sentencing

Sentencing Council



Consistency in Sentencing

Sentencing Council



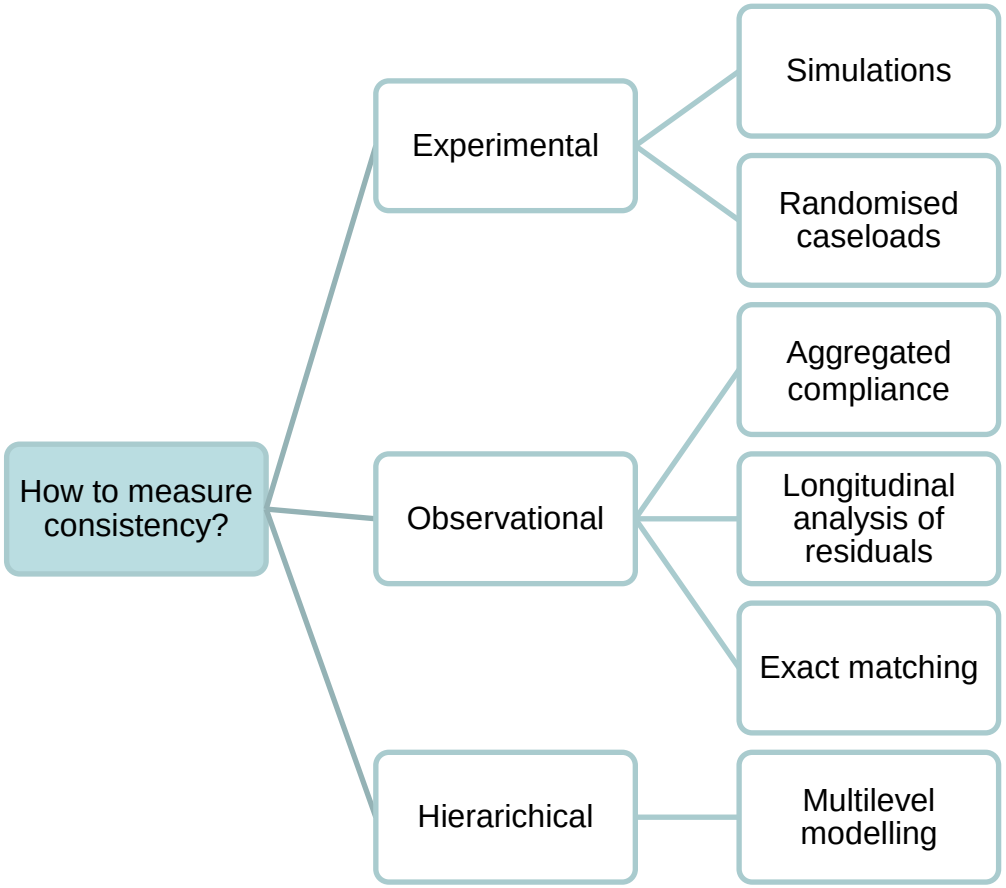
Methods

Sentencing Council

- We are only interested in methods that can offer a –more or less valid- measure of consistency.

- We classify methods according to the type of data that they require.

Methods


How to measure consistency?
Experimental
Simulations
Randomised caseloads
Observational
Aggregated compliance
Longitudinal analysis of residuals
Exact matching
Hierarichical
Multilevel modelling

Methods: Experimental: Simulations

Sentencing Council

- Different judges are asked to sentence a hypothetical case.
- The standard deviation of these sentences is taken as a measure of inconsistency.
- This is a straightforward and simple process; currently being used by the Supreme Court of the United Kingdom.
- Amongst its limitations: a) requires collaboration from sentencers, b) the conditions of the real sentencing process are not perfectly replicated, c) low external validity.

Methods: Experimental: Randomised

Sentencing Council

- This method relies on the practices followed by most federal courts in the US where judges in the same location are assigned cases randomly.
- This random process guarantees that over a large number of cases, each judge in the same court is assigned in average the same sort of cases.
- So we can just take the standard deviation of the different sentence means by judge as a measure of dispersion.
- This is again very straightforward and parsimonious, but limited by similar issues: data scarcely available, only covers inter-judge disparities, generalizable only to the court level.

Methods: Observational: Compliance

Sentencing Council

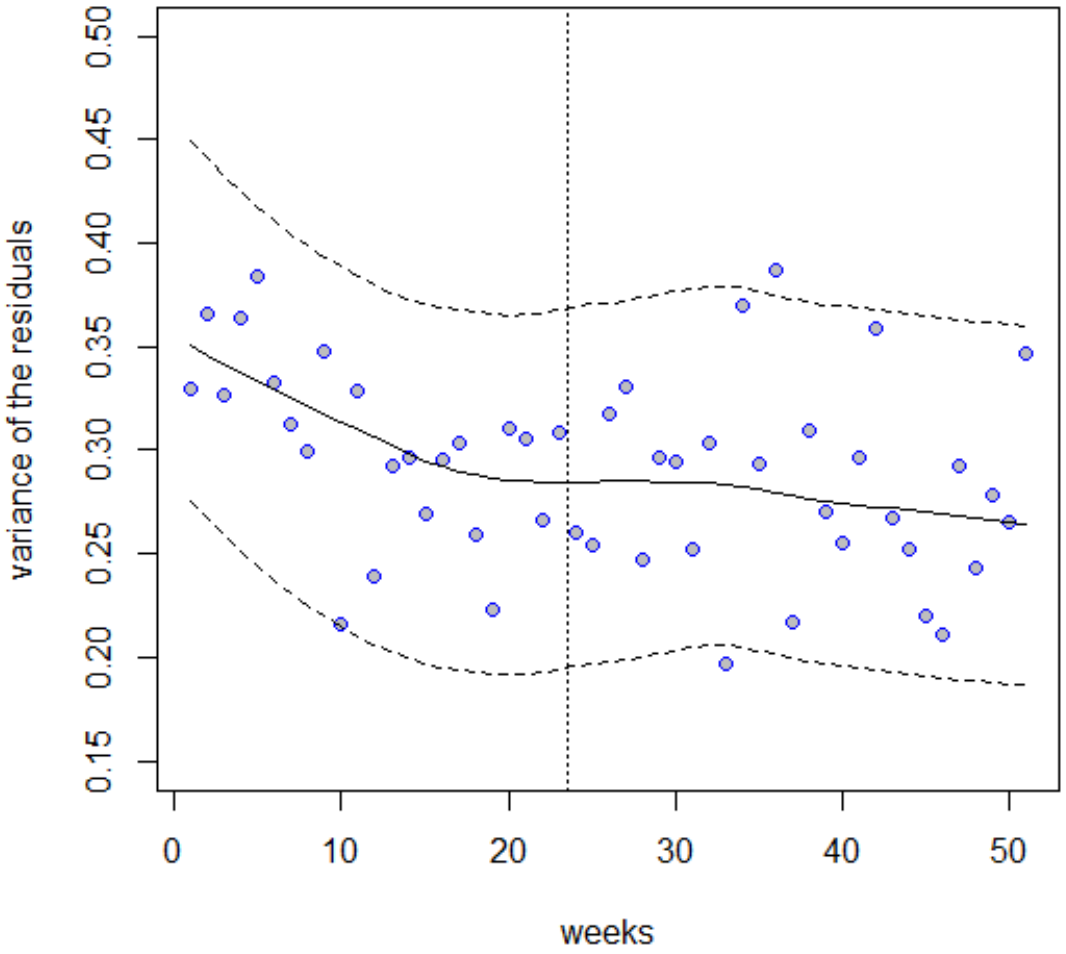
- Different jurisdiction have been recently adopting systems of sentencing guidelines.
- Here we can measure consistency as the proportion of sentences falling within the recommended sentencing bands.
- Routinely published by all US sentencing commissions.
- This is probably the simplest method and the necessary data is widely available.
- However, only measures by specific type of offence can be obtained, and these measures are very coarse.

Methods: Observational: Residuals

Sentencing Council

- At the Sentencing Council we explored the possibility of using regression models for the analysis of consistency.
- We regressed sentence length on a set of nine relevant legal factors using data from assault offences.
- Given that our model is formed exclusively by legal factors, the R^2 can be interpreted as the share of variability in sentence length explained by legitimate reasons, and the rest as just inconsistencies.
- The problem is that this approach confounds genuine inconsistency with problems of misspecification.
- Although, if we assume that the misspecification problems remain constant we can still use this approach to assess changes of consistency across time

Methods: Observational: Residuals



Methods: Observational: Exact Matching

Sentencing Council

- The rationale of this method implies: a) classifying offences into groups that are as homogeneous as possible, and b) comparing changes in the dispersion in sentence length for each group.

- Exact matching is a simpler approach to implement and more intuitive when it comes to conveying results.

- Furthermore, differences in the variances across time can be tested using F-tests.

- One of the major limitations is that it only offers a discrete view of changes in time.

Methods: Observational: Exact Matching

Sentencing Council

Type of Offence	Previous Convictions	Aggravating / Mitigating	Group Size: Before	Group Size: After	Variance: Before	Variance: After	Variance Difference
ABH	0	-	112	112	.37	.42	-.05
ABH	1-3	-	141	89	.36	.30	.05
GBH	1-3	-	78	74	.21	.24	-.03
GBH	0	-	60	59	.32	.24	.08
ABH	1-3	sustained	40	51	.50	.34	.16
GBH	1-3	drugs	48	37	.23	.32	-.08
ABH	1-3	drugs	62	35	.28	.20	.08
Intent	1-3	-	33	33	.30	.14	.16
ABH	1-3	first op.	34	28	.55	.32	.23
GBH	1-3	remorse	28	28	.13	.24	-.11

Methods: Hierarchical: Multi. Modeling

Sentencing Council

- When either judges or courts are identified in the dataset we can also use multilevel modeling.
- We can tests whether some of the legal factors have significant random slopes.
- Furthermore, we can construct 95% confidence intervals using the random slope variability to obtain specific measures of inconsistency.
- The limitations are that we can only look at inter-court disparities.

Methods: Hierarchical: Multi. Modeling

Sentencing Council

Variables	Regression coefficient	Standard error
Fixed effects		
Intercept	**5.356**	.056
Previous conviction	**.429**	.060
(Previous convictions)^2	**-.114**	.015
First opportunity	**-.094**	.017
Remorse	**-.119**	.016
Carer	**-.125**	.037
Gang	.014	.021
Vulnerable	**.139**	.028
Public worker	**-.078**	.036
Sustained	**.201**	.020
Drugs	**.053**	.016
GBH	**.467**	.017
GBH with intent	**1.619**	.021
Random effects		
Intercept	**.005**	.002
Vulnerable	**.015**	.008
Sustained	**.007**	.004
Level 1 residuals	.293	.006

Conclusion

- We have reformulated consistency making it more encompassing: concepts such as discrimination and intra-judge disparities are sources of inconsistency.

Conclusion

- We have reformulated consistency making it more encompassing: concepts such as discrimination and intra-judge disparities are sources of inconsistency.

- Methods relying on experimental data can be considered the most adequate approach to measure consistency but they are not perfect.

Conclusion

Sentencing Council

- We have reformulated consistency making it more encompassing: concepts such as discrimination and intra-judge disparities are sources of inconsistency.

- Methods relying on experimental data can be considered the most adequate approach to measure consistency but they are not perfect.

- In spite of the limitations of observational data –confounding effects-, we can still use them to assess changes in consistency across time.

Conclusion

Sentencing Council

- We have reformulated consistency making it more encompassing: concepts such as discrimination and intra-judge disparities are sources of inconsistency.
- Methods relying on experimental data can be considered the most adequate approach to measure consistency but they are not perfect.
- In spite of the limitations of observational data –confounding effects-, we can still use them to assess changes in consistency across time.
- Hierarchical data opens the possibility of using multilevel methods such as the random slopes model which are less prone to biases due to omitted relevant variables.

Conclusion

Sentencing Council

- We have reformulated consistency making it more encompassing: concepts such as discrimination and intra-judge disparities are sources of inconsistency.
- Methods relying on experimental data can be considered the most adequate approach to measure consistency but they are not perfect.
- In spite of the limitations of observational data –confounding effects-, we can still use them to assess changes in consistency across time.
- Hierarchical data opens the possibility of using multilevel methods such as the random slopes model which are less prone to biases due to omitted relevant variables.
- Any further ideas are most welcome!