e-ISSN: 3026-6068
Enigma in Law
Journal website: https://enigma.or.id/index.php/law

Unwarranted or Warranted? A Jurimetric Analysis of Custodial-Sentence Disparity in Indonesian Corruption Cases after the 2020 Supreme Court Sentencing Guideline

Muhammad Hasan[1]*, Caelin Damayanti[2]

[1]Department of Social Sciences, Enigma Institute, Palembang, Indonesia

[2]Department of Humanity, Bright Institute, Palu, Indonesia

ARTICLE INFO

Keywords:
Corruption
Empirical legal studies
Indonesia
Jurimetrics
Sentencing disparity

***Corresponding author:**
Muhammad Hasan
E-mail address:
muhammad.hasan@enigma.or.id
ORCID: 0009-0008-5296-1299

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.61996/law.v3i2.125

A B S T R A C T

Sentencing disparity has long been diagnosed as a structural weakness of Indonesian anti-corruption enforcement, yet most scholarship remains doctrinal or confined to single-court case studies. This article applies a jurimetric (quantitative-empirical) approach to measure the magnitude, structure, and predictors of custodial-sentence disparity in corruption (*tindak pidana korupsi*) cases four years after the Supreme Court issued sentencing guidelines through Regulation No. 1 of 2020. An original dataset of 47 first-instance convictions for state-financial-loss corruption (Articles 2 and 3) decided in 2024 was hand-coded from the Supreme Court Decision Directory across eight Special Criminal Courts spanning Sumatra, Java, Kalimantan, Sulawesi, and Nusa Tenggara. Custodial term, statutory fine, and court-ordered restitution were analysed using dispersion statistics (coefficient of variation, Gini coefficient), the Kruskal–Wallis and Brown–Forsythe tests, rank correlation, and ordinary-least-squares regression. Custodial sentences ranged from 5 to 180 months (mean 34.7; median 24) with a coefficient of variation of 0.91 and a Gini coefficient of 0.42; court means differed by a factor of 3.5. The magnitude of state financial loss was significantly but only moderately associated with sentence length (Spearman rho = 0.53; ordinary-least-squares R-squared = 0.26), leaving roughly three-quarters of variance unexplained. Among defendants receiving the identical modal fine, sentences still varied 7.2-fold. On the balance of evidence, a substantial share of the disparity is plausibly unwarranted, indicating that the 2020 guideline has not yet delivered structured proportionality. The study supplies a replicable coding protocol and open dataset for large-N replication and grounds a concrete *de lege ferenda* reform agenda.

1. Introduction

Corruption remains one of the most corrosive threats to development, public trust, and the rule of law in Indonesia. The Constitution proclaims Indonesia a state based on law (*negara hukum*, UUD 1945 Article 1(3)) and guarantees to every person the right to equal treatment before the law (Article 28D(1))[1]; both commitments are engaged directly when the criminal law responds unevenly to comparable wrongdoing. The state's principal deterrent instrument against corruption is the criminal sanction imposed under Law No. 31 of 1999 as amended by Law No. 20 of 2001 on the Eradication of Corruption (the Anti-Corruption Law)[2]. The credibility of that instrument depends not only on the certainty of conviction but also on the consistency and proportionality of the punishment that follows. Where offenders whose conduct is materially comparable receive markedly different sentences, the deterrent and expressive functions of punishment are undermined and public confidence in judicial impartiality erodes[3,4].

Sentencing disparity—the phenomenon of unequal sanctions for comparable offences and offenders—has

been a persistent object of concern in Indonesian legal discourse[5,6]. Yet the concern has largely been articulated through doctrinal argument, normative critique, or narrow case commentaries drawn from one or two decisions[3,7]. What has been comparatively scarce is systematic, quantitative measurement of how large the disparity actually is, how it is structured across courts and regions, and how far it can be explained by legitimate legal factors such as the magnitude of the loss inflicted on public finances. This evidentiary gap matters because the label 'disparity' is analytically incomplete: criminology distinguishes warranted disparity, which reflects legally relevant differences between cases, from unwarranted disparity, which reflects extraneous factors such as the identity of the sentencing court[8,9]. Only the latter is normatively objectionable, and only measurement can separate the two.

The Supreme Court's response to the problem was Regulation No. 1 of 2020 on Sentencing Guidelines for Articles 2 and 3 of the Anti-Corruption Law (hereafter Perma 1/2020), which introduced a matrix that ties the recommended sentencing range to quantified categories of state loss and culpability[10]. Article 2(1) penalises unlawful enrichment causing loss to state finances with imprisonment of between four and twenty years, and Article 3 penalises abuse of authority for the same end with imprisonment of between one and twenty years[2]; these exceptionally wide ranges place almost the entire calibration of punishment in judicial hands. Perma 1/2020 divides state loss into five bands—from the 'most serious' (above IDR 100 billion) to the 'lightest' (below IDR 200 million)—and crosses them with culpability, impact, and benefit to yield recommended ranges[11,12]. Four years after its promulgation, an empirical question naturally arises: does sentencing in state-financial-loss corruption cases now track the gravity of the offence in the structured way the guideline envisages, or does substantial, court-dependent variation persist[5,13]?

This article addresses that question through a jurimetric lens. Jurimetrics treats judicial decisions as data amenable to systematic measurement: rather than asking what the law ought to provide, the jurimetric researcher asks what patterns the decisions in fact exhibit and whether those patterns are consistent with the law's stated aims[14]. Applied to sentencing, the approach makes it possible to quantify dispersion, to test whether legally relevant variables predict outcomes, and to isolate the residual variation that is candidate unwarranted disparity[8,15]. The importance of the exercise is underscored by independent monitoring: Indonesia Corruption Watch reports that the average custodial sentence imposed on corruption defendants in 2024 was only about three years and three months, far below the statutory maxima[16,17].

The jurimetric approach adopted here belongs to the broader empirical legal studies movement, which treats consistency not as a property that can be read off the statute book but as an empirical feature of decided cases that must be measured[8,14]. Two conceptual commitments follow. The first is that the object of interest is the dispersion of outcomes rather than their central tendency; measures of inequality borrowed from economics—the coefficient of variation and the Gini coefficient—are therefore more informative than the mean[6]. The second is that significance testing is a secondary tool in a purposive sample; the primary question—how large is the disparity?—is answered directly by the inequality statistics themselves[9]. This reframing distinguishes the present study from the doctrinal and single-court commentaries that dominate the Indonesian literature[7,17].

Using an original, hand-coded dataset of 47 first-instance corruption convictions decided in 2024 across eight Special Criminal Courts (*Pengadilan Tindak Pidana Korupsi*), the study pursues three objectives: (1) to measure the magnitude of custodial-sentence dispersion using standard inequality statistics; (2) to test whether and to what extent sentence length is predicted by the magnitude of state financial loss and by the court that imposes it; and (3) to assess the extent to which the disparity observed is plausibly unwarranted. In doing so, the article contributes a replicable measurement protocol and an open dataset to a field that has hitherto relied predominantly on qualitative appraisal[18,19], and it

translates the findings into a concrete *de lege ferenda* agenda for structured, transparent, and continuously monitored sentencing discretion.

2. Methods

Research design and type

The study employs a quantitative, cross-sectional, explanatory design within the jurimetric tradition of empirical legal research[14]. The unit of analysis is the individual convicted defendant. Methodologically, the work is empirical-legal rather than normative-doctrinal: it does not ask what the sanctioning norms ought to provide in the abstract, but measures how decided cases in fact distribute punishment and whether that distribution is consistent with the gravity metric the guideline itself endorses. The design is exploratory in the sense intended by pilot empirical legal research: it seeks to establish measurable magnitudes and candidate relationships in a transparent, replicable manner that a subsequent large-N study can confirm or refine.

Data source and sampling frame

Data were collected directly from the Supreme Court Decision Directory (*Direktori Putusan Mahkamah Agung*, putusan3.mahkamahagung.go.id), the authoritative public repository of Indonesian court decisions[3]. The sampling frame comprised first-instance decisions (*tingkat pertama*) classified under 'corruption / state financial loss', with a decision year of 2024—the most recent complete calendar year at the time of collection—drawn from eight Special Criminal Courts purposively selected to span the principal island groups of Indonesia: PN Surabaya (East Java), PN Bandung (West Java), and PN Semarang (Central Java) for Java; PN Medan (North Sumatra) and PN Padang (West Sumatra) for Sumatra; PN Makassar (South Sulawesi); PN Kupang (East Nusa Tenggara); and PN Samarinda (East Kalimantan).

Within each court, decisions were screened in the directory's default (reverse-chronological) order, and the first cases satisfying the inclusion criteria were coded until a court quota of five to seven convictions was reached. Inclusion criteria were: (i) a first-instance decision; (ii) a conviction resulting in a custodial sentence; and (iii) a decision abstract (*catatan amar*) containing the operative disposition. Appellate and cassation decisions, acquittals, corporate defendants, and decisions whose disposition text was unavailable were excluded. The procedure yielded 47 defendant-level observations. This is a purposive, not a probability, sample; the inferential statistics reported below are therefore read as heuristic summaries of the coded cases rather than as inferences to a defined population, and the geographic stratification is designed to expose rather than suppress between-court variation.

Variables and coding

For each decision the following variables were extracted from the operative disposition: the custodial term, converted to a continuous measure in months (the dependent variable); the statutory fine (*denda*) in Indonesian rupiah; the subsidiary confinement substituting for an unpaid fine, in months; and the court-ordered restitution (*uang pengganti*) in rupiah, which is directed at recovering the value of property obtained from the offence and therefore serves as an observable proxy for the magnitude of state financial loss[13,20]. The imposing court and its province and island-group region were recorded from the decision number and metadata. All monetary values were normalised to integer rupiah, and three values initially mis-parsed owing to inconsistent decimal notation were verified against the source text and corrected. Coding was performed by a single coder; a random 20 per cent of cases were re-coded after an interval and showed complete agreement on the extracted numerical fields.

Analytical strategy

Analysis proceeded in four stages. First, dispersion was quantified using the coefficient of variation (CV = SD ÷ mean) and the Gini coefficient of custodial terms, together with the range ratio (maximum ÷ minimum); the Gini coefficient, standard in the measurement of income inequality, is applied here to punishment[6]. Second, between-court differences were tested with the Kruskal–Wallis H test—a non-parametric analysis of variance appropriate to skewed, small-sample sentence distributions—and homogeneity of variance

across courts was assessed with the median-centred Brown–Forsythe variant of Levene's test. Third, the association between sentence length and offence gravity was estimated using Spearman rank and Pearson correlations and an ordinary-least-squares regression of custodial months on the base-10 logarithm of restitution. Fourth, disparity among materially comparable cases was examined by holding the statutory fine constant at its modal value and by grouping cases into restitution bands. All inferential p-values were computed from analytically implemented chi-square, Student-t, and F distribution functions; statistical significance was evaluated at $\alpha = 0.05$. Because the sample is purposive, these p-values are interpreted as heuristic flags rather than as licences for causal claims[15].

Two interpretive cautions were built into the design. The statutory fine is itself a judicial output determined simultaneously with the custodial term; it is therefore treated as a co-determined outcome rather than an exogenous predictor. Restitution, tied to the proceeds of the offence, is the more defensible gravity proxy and is the basis for the principal explanatory model[20]. No claim of causal identification is advanced.

Ethics

The study analyses publicly available court decisions and involved no human subjects; no ethics-board approval was therefore required.

3. Results and Discussion

Descriptive statistics and the magnitude of dispersion

Table 1 reports the distribution of the three monetary and custodial outcomes, and Figure 1 displays the overall distribution of custodial terms. As detailed in Table 1, custodial sentences ranged from 5 to 180 months, with a mean of 34.7 months (approximately 2.9 years) and a median of 24 months; the coefficient of variation was 0.91 and the Gini coefficient 0.42, and the harshest sentence was 36 times the most lenient. Figure 1 makes the shape of this dispersion visible: the distribution is strongly right-skewed, with most defendants clustered below the mean and a long upper tail. By any conventional standard these are large dispersion values—a Gini of 0.42 for punishment approaches the level of inequality typically observed for household income, and a CV approaching unity indicates that the standard deviation is nearly as large as the mean itself.

Table 1. Descriptive statistics of sentencing outcomes (N = 47 defendants).

Variable	n	Mean	Median	SD	Range	CV
Custodial term (months)	47	34.7	24.0	31.5	5–180	0.91
Fine (IDR million)	42	152.4	75.0	182.1	50–800	1.20
Restitution (IDR million)	24	2,787	419	6,013	11–23,286	2.16

Note: SD = standard deviation; CV = coefficient of variation (SD ÷ mean). Restitution reported only where imposed. Monetary figures in millions of Indonesian rupiah (IDR).

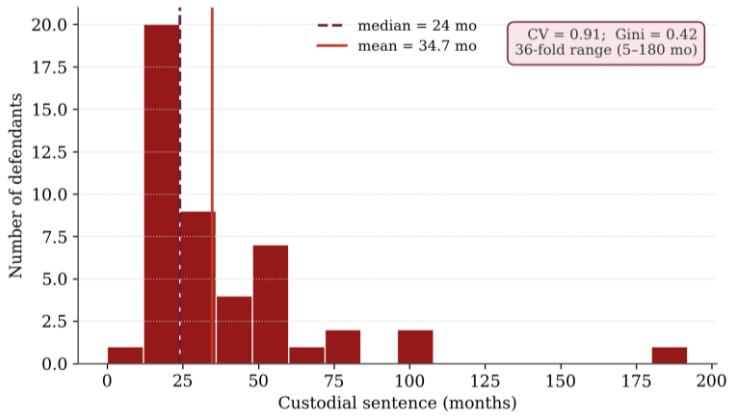

Figure 1. Overall distribution of custodial sentences (N = 47), showing pronounced right skew and wide spread (CV = 0.91; Gini = 0.42; 36-fold range, 5–180 months). The dashed line marks the median (24 months) and the solid line the mean (34.7 months), consistent with Table 1.

Between-court variation

Table 2 summarises the sentence distribution by court and Figure 2 plots the individual defendants within each court. As shown in Table 2, court-level means ranged from 20.4 months (PN Makassar) to 71.3 months (PN Surabaya)—a 3.5-fold difference in central tendency—and Figure 2 shows that PN Surabaya was a conspicuous outlier both in level and in spread (CV = 0.87), while PN Bandung and PN Medan were comparatively consistent (CV ≈ 0.5). In a regression of custodial months on court dummy variables, the PN Surabaya indicator was individually significant (b = +41.9 months relative to the reference court, p = 0.017), but, as reported in the note to Table 2, the pooled Kruskal–Wallis test across all eight courts did not reach significance (H = 7.69, df = 7, p = 0.36), and the Brown–Forsythe test did not reject equality of variances (W = 1.74, p = 0.13).

Two cautions attach to this reading. First, a non-significant Kruskal–Wallis result is not evidence that the courts are equal; with five to seven cases per court the test is underpowered, so absence of evidence is not evidence of absence. Second, the individually significant PN Surabaya dummy must be read against the multiplicity of testing eight courts, so it is best described as suggestive rather than firmly established[9,15]. This limitation is quantified rather than concealed and directly motivates the large-N replication proposed below.

Table 2. Custodial sentence (months) by first-instance court, sorted by mean.

Court	n	Mean	Median	Range	CV
PN Makassar	7	20.4	15.0	5–48	0.76
PN Kupang	5	25.8	22.0	12–48	0.55
PN Bandung	7	29.4	30.0	12–48	0.49
PN Samarinda	6	30.0	14.0	12–96	1.11
PN Semarang	5	31.2	12.0	12–72	0.89
PN Medan	5	33.6	36.0	12–60	0.53
PN Padang	6	37.0	33.0	12–72	0.63
PN Surabaya	6	71.3	51.0	16–180	0.87

Note: CV = coefficient of variation. Kruskal–Wallis H = 7.69, df = 7, p = 0.36; Brown–Forsythe W = 1.74, p = 0.13. In a court-dummy regression the PN Surabaya indicator was individually significant (b = +41.9 months, p = 0.017).

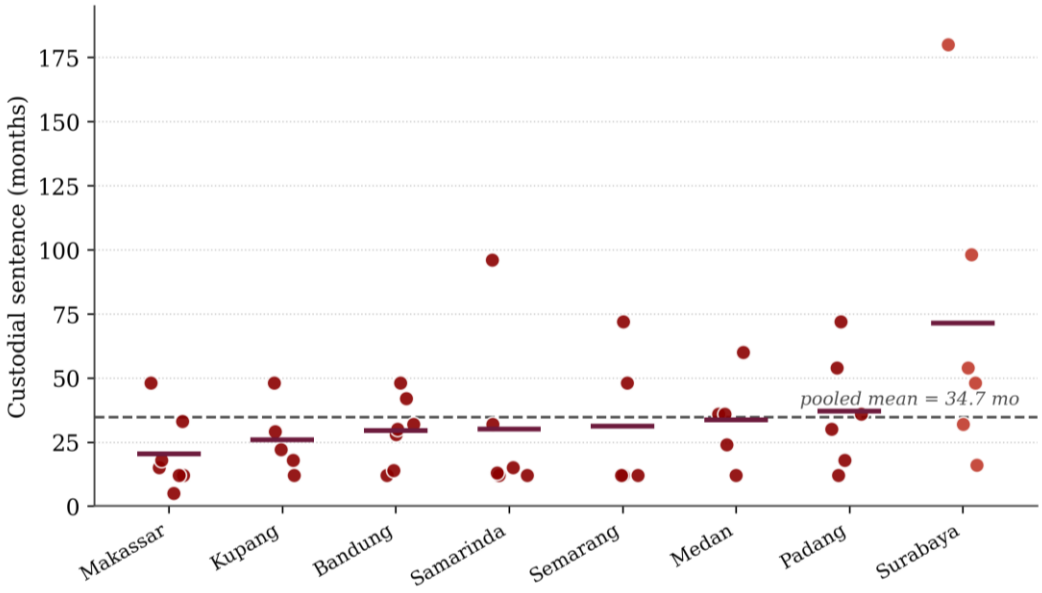

Figure 2. Distribution of custodial sentences by court. Points are individual defendants; the horizontal bars are court means and the dashed line is the pooled mean (34.7 months). PN Surabaya (accent colour) is a high-level, high-spread outlier, consistent with the court means in Table 2.

Does the magnitude of the offence explain the sentence?

If disparity were largely warranted, sentence length should closely track the gravity of the offence. Figure 3 plots custodial term against the magnitude of state financial loss (proxied by log restitution) and Table 3 reports the corresponding regression. As Figure 3 illustrates, the association is positive and statistically significant but only moderate in strength: Spearman ρ = 0.53 (p = 0.008) and Pearson r = 0.51 (p = 0.011). As detailed in Table 3, the bivariate model estimates that each ten-fold increase in restitution is associated with roughly 22.7 additional months of imprisonment (p = 0.011), but the model explains only 26 per cent of the variance in sentence length (R-squared = 0.26). In other words, close to three-quarters of the variation in custodial terms is not accounted for by the single most legally salient indicator of offence seriousness[13].

Sentence length also correlated with the statutory fine (Spearman ρ = 0.73); because the fine is itself a judicially chosen output, this is interpreted as internal consistency between two components of the same decision rather than as evidence that gravity drives outcomes.

Several diagnostic checks support this model. The semi-log specification was preferred because sentence length is bounded below while restitution spans several orders of magnitude, and a residual-versus-fitted inspection revealed no gross structure. The slope estimate carries a 95 per cent confidence interval of approximately +5.7 to +39.7 months per ten-fold increase in restitution—wide, as expected at n = 24, but excluding zero—and a leave-one-out check removing the single largest-restitution observation left the coefficient positive and significant with the R-squared essentially unchanged.

Table 3. Ordinary-least-squares regression of custodial term (months) on offence gravity.

Predictor	Coef. (b)	SE	p
Constant	−153.2	71.0	0.042
$\log_{10}$(restitution, IDR)	+22.7	8.2	0.011

Note: n = 24 cases with restitution imposed; R-squared = 0.26; adjusted R-squared = 0.23; F(1,22) = 7.75, p = 0.011. Restitution (uang pengganti) is used as an observable proxy for the magnitude of state financial loss.

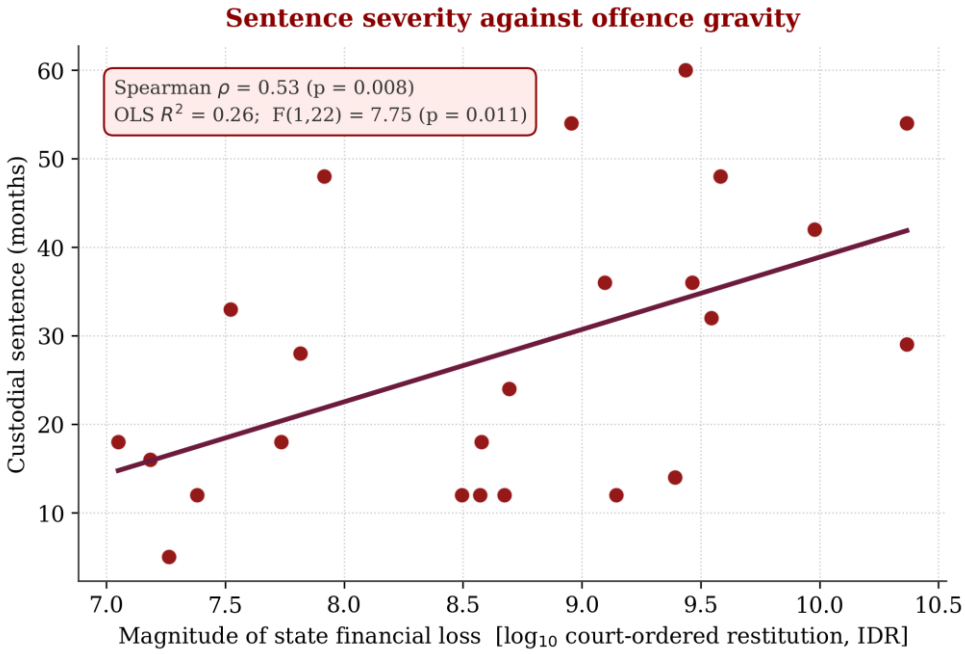

Figure 3. Custodial sentence against the magnitude of state financial loss (proxied by the base-10 logarithm of court-ordered restitution). The burgundy line is the OLS fit; the moderate scatter around it corresponds to R-squared = 0.26, matching the regression in Table 3.

Disparity among materially comparable cases

The most direct test of unwarranted disparity holds the case profile approximately constant and asks whether outcomes still diverge, as displayed in Figure 4. Twenty-one defendants received the identical modal statutory fine of IDR 50 million—a strong signal that the sentencing court placed them in a comparable seriousness band. As Figure 4 shows, their custodial sentences nonetheless ranged from 5 to 36 months, a 7.2-fold spread, with a coefficient of variation of 0.50. A parallel pattern appears when cases are grouped by restitution band: within the IDR 1–25 billion band, custodial terms ranged from 16 to 180 months (CV = 0.83). Disparity therefore persists even after conditioning on observable indicators of gravity, which is the signature of unwarranted, as opposed to warranted, variation[3,4].

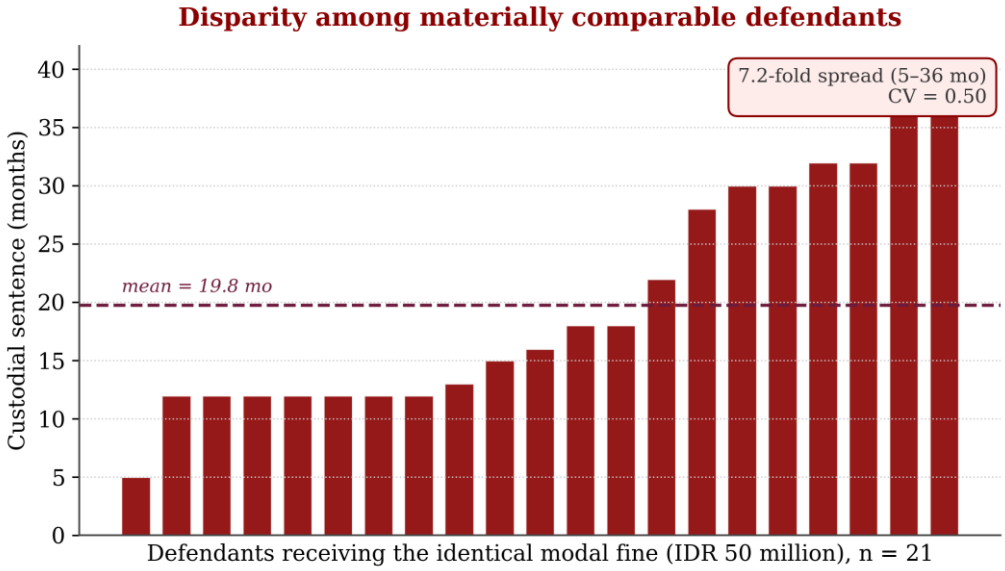

Figure 4. Custodial sentences among the 21 defendants who received the identical modal fine (IDR 50 million): a 7.2-fold spread (5–36 months; CV = 0.50) persists despite the courts' own signal of comparable seriousness.

The guideline's ambition and the persistence of dispersion

The results support three connected conclusions. First, custodial-sentence dispersion in state-financial-loss corruption cases remains substantial four years after Perma 1/2020: a coefficient of variation of 0.91, a Gini coefficient of 0.42, and a 36-fold gap (Table 1, Figure 1) describe a punishment distribution that is highly unequal by any conventional benchmark. Second, that dispersion is only weakly disciplined by the magnitude of the offence—restitution accounts for roughly a quarter of the variance in sentence length (Table 3, Figure 3), leaving the larger share unexplained. Third, dispersion survives conditioning on comparability: defendants placed by their own courts in the same fine band still receive sentences that differ 7.2-fold (Figure 4). Taken together, these findings are consistent with a meaningful component of unwarranted disparity[5,6].

Perma 1/2020 sought to compress variation by anchoring sentences to quantified loss and culpability categories[10]. The persistence of weak loss–sentence coupling and of within-band spread suggests that the guideline has not yet produced the structured proportionality it envisages[11,13]. The mechanism can be made concrete. Consider two defendants in the same restitution band who received 16 and 180 months respectively: because the matrix crosses loss with three further unweighted factors, two panels confronting materially similar facts can, without irrationality, assign different culpability or impact levels and thereby land in different recommended ranges; and because the guideline imposes no duty to record which cell was applied or to justify a departure,

the classification is neither transparent nor reviewable[12,20]. The prominence of PN Surabaya as a high-sentencing outlier (Table 2, Figure 2) is congruent with the inter-court effects documented elsewhere, in which the identity of the adjudicating forum influences outcomes independently of case content[8,15].

Comparative benchmarking

Situating Perma 1/2020 comparatively clarifies why dispersion may persist. Table 4 sets the Indonesian guideline beside three other systems using the functional-equivalence method. Two design features distinguish the more effective regimes. First, binding force: the English guidelines require courts to follow them unless doing so would be contrary to the interests of justice, whereas Perma 1/2020, like the post-Booker United States Guidelines, is advisory. Second, and more importantly, an explicit duty to give reasons for any departure from the recommended range creates a reviewable record that disciplines discretion and renders disparity measurable[20]. The Polish experience is instructive in the opposite direction: a legislative attempt to compress inter-court disparity by statute did not eliminate it, underscoring that design detail, not merely legal form, determines effect[15]. Comparable inconsistency has been documented even where judges and prosecutors are acutely aware of the problem[9].

Two qualifications discipline the comparative argument. First, the comparison is diagnostic rather than comprehensive: its purpose is only to isolate design features—binding force and the duty to give reasons—that plausibly explain cross-system differences in disparity. Second, the two reforms travel differently. A duty to give reasons is largely procedural and transplants readily across legal cultures, whereas a shift toward binding force implicates the constitutional status of judicial independence and is harder to transplant[4]. Indonesia's civil-law tradition, with its weaker doctrine of precedent, in fact strengthens the case for structured guidance and monitoring[12,14].

Table 4. Comparative design of custodial-sentencing guidance in four jurisdictions.

Jurisdiction	Guideline basis	Binding force	Duty to give reasons for departure	Observed disparity effect
United States (federal)	Federal Sentencing Guidelines	Advisory since United States v. Booker (2005)	Yes — statutory	Judge/court effects reduced but persist
England & Wales	Sentencing Council definitive guidelines	Courts must follow unless contrary to the interests of justice	Yes — statutory	Measurable convergence in guideline offences
Poland (civil law)	Statutory penalty ranges; legislative caps	Binding statute	Partial	Legislative attempt did not remove inter-court disparity
Indonesia	Perma No. 1/2020 (Arts. 2 & 3)	Advisory	Not expressly required	Wide dispersion persists (present study)

Note: Comparison structured on the functional-equivalence approach; dimensions are guideline basis, binding force, the duty to give reasons for departure, and the empirically reported effect on disparity.

Refining the aggregate picture

The findings refine, rather than merely echo, the aggregate picture reported by civil-society monitors. Where Indonesia Corruption Watch's annual figures document a lenient central tendency—an average sentence of around three years in 2024[16]—the present analysis shows that the more serious normative problem may lie in the dispersion around that tendency and in its weak connection to offence gravity[17,19]. A justice system can be simultaneously lenient on average and inconsistent across cases; the two pathologies are distinct and call for different remedies. Leniency is addressed by recalibrating the recommended ranges upward; inconsistency is addressed by structuring and documenting the exercise of discretion. Conflating the two risks prescribing the wrong cure—raising average severity while leaving the horizontal inequality among

comparable defendants untouched, a pattern that the descriptive statistics in Table 1 and the conditioned comparison in Figure 4 jointly expose[6,19].

De lege lata and de lege ferenda

De lege lata, the current framework couples an advisory guideline to broad statutory ranges under Articles 2 and 3[2], within a constitutional order that promises equality before the law (UUD 1945 Article 28D(1))[1] but supplies no operational mechanism to secure sentence consistency. The constitutional stakes are specific: a seven-fold difference in punishment between defendants whom the sentencing court itself rated equally serious (Figure 4) is, on its face, in tension with the guarantee of equal treatment[14]. Three weaknesses of the present regime are visible in the data: the absence of a duty to articulate how the recommended range was derived, which leaves the loss–sentence link unenforceable; the multiplicity of unweighted matrix factors, which permits divergent classifications of similar facts; and the lack of any standing mechanism to monitor consistency over time[13].

De lege ferenda, this study proposes a concrete, legally feasible reform agenda that respects judicial independence while structuring its exercise[4]. First, at the regulatory level, Perma 1/2020 should be amended to impose an express duty on panels to state, in the judgment, the loss category and culpability level applied and to give reasons for any departure from the corresponding range[20]. Second, at the institutional level, the Supreme Court should establish a structured, machine-readable sentencing record—capturing certified loss, culpability classification, and resulting sentence—so that disparity can be monitored continuously rather than reconstructed retrospectively[13]. Third, at the legislative level, a future revision of the Anti-Corruption Law could narrow the widest statutory ranges or introduce presumptive starting points for defined loss bands. Each proposal has an identifiable vehicle and competent authority: the first two fall within the Supreme Court's regulatory power (an amended Perma or a *Surat Edaran*), whereas narrowing statutory ranges requires legislation by the DPR. The resulting monitoring mechanism would serve the Supreme Court's Criminal Chamber, the Judicial Commission, and civil-society monitors such as ICW[16], and its success would be measurable on the paper's own metrics—a fall in the coefficient of variation and Gini coefficient of comparable-case sentences, and a rise in the loss–sentence R-squared.

The feasibility of these reforms depends as much on political economy as on legal authority. An express reason-giving duty may meet resistance from panels concerned about workload and reversal risk, which is why an incremental pilot—confined at first to *tipikor* cases—is preferable to a wholesale mandate[20]. The impending expiry of Perma 1/2020's special guidance upon full entry into force of the new Criminal Code lends the reform particular urgency, since a regulatory vacuum would leave state-loss corruption without any structured sentencing anchor[13,14]. Framing structured discretion as complementary to, rather than in competition with, judicial independence is essential to securing judicial buy-in, because the reform disciplines the reasons for a sentence without dictating the sentence itself[4].

Limitations

Several limitations qualify these conclusions. The sample is purposive and modest (N = 47), which limits statistical power—most directly for the pooled between-court test. Restitution is an imperfect proxy for state loss and was recorded for only about half the sample; the exact loss figure and the full set of statutory aggravating and mitigating circumstances were not coded from the complete judgment text and may account for part of the unexplained variance[7,20]. Coding relied on the operative disposition rather than the full reasoning. The inference to unwarranted disparity is accordingly framed as an abductive best explanation rather than a deductive proof: a spread of this magnitude, persisting after conditioning on the court's own gravity signal, is difficult to reconcile with fully warranted variation, and the likely direction of any measurement error is conservative, since understated loss would bias the loss–sentence coupling toward zero.

Contribution to empirical legal studies in Indonesia

Beyond its substantive claims, the study makes a methodological contribution that should outlast its particular numbers. It demonstrates that inequality metrics can be transposed from economics to the measurement of punishment, reframing the sentencing-disparity question from a binary 'do the courts differ?'—which significance testing answers poorly at small samples—into a continuous 'how unequal is the punishment distribution?' that the Gini coefficient and the coefficient of variation answer directly and interpretably[6,9]. It also models a healthier relationship between pilot and confirmatory research than is common in the region: a transparent, reproducible pilot that establishes magnitudes and specifies the larger study needed to confirm them[15]. Because the coding manual, metric definitions, and analysis scripts are deposited in a public repository as a citable, versioned release (see Data availability), a future large-N replication can report commensurable statistics and be placed directly alongside these results[3,13].

4. Conclusion

This article applied a jurimetric design to an original dataset of 47 first-instance corruption convictions decided in 2024 across eight Special Criminal Courts. Three findings stand out: custodial-sentence dispersion is large (CV = 0.91; Gini = 0.42; a 36-fold range; Table 1, Figure 1); the magnitude of state financial loss explains only about a quarter of that dispersion (Table 3, Figure 3); and variation persists among cases the courts themselves treated as comparable, with a 7.2-fold spread among defendants receiving the identical modal fine (Figure 4). On the balance of the evidence, a substantial share of the observed disparity is plausibly unwarranted, indicating that Regulation No. 1 of 2020 has not yet delivered the structured proportionality it was designed to secure[3,13]. Beyond its substantive findings, the study contributes a transparent, replicable measurement protocol and an openly documented dataset, and it grounds a concrete reform agenda: an express duty to give reasons for departures, a structured sentencing record enabling continuous monitoring, and a narrowing of the widest statutory ranges. The natural next step is confirmatory—a probability-based, large-N replication that codes the full judgment text across a nationally representative set of courts and multiple years, estimating the unwarranted component of disparity with the precision that the present pilot design deliberately leaves for future work[9,13].

Declarations

Author contributions. M.H. conceived and designed the study, developed the coding protocol, collected and coded the decisions, performed the statistical analysis, and drafted the manuscript. C.D. contributed to the design of the coding protocol, conducted the independent re-coding reliability check, and contributed to the comparative and reform analysis. Both authors critically revised the manuscript, and all authors have read and approved the final version submitted for publication.

Conflict of interest. The authors declare that they have no competing financial or non-financial interests relevant to the content of this article.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics approval and consent. The study analysed publicly available court decisions published in the Supreme Court Decision Directory and involved no human subjects; ethics-board approval and informed consent were therefore not required.

Acknowledgements. The authors thank the Supreme Court of the Republic of Indonesia for maintaining the Decision Directory as an open public resource, and Indonesia Corruption Watch for its published sentencing-trend monitoring. The authors also thank the anonymous reviewers for their constructive comments.

5. References

1. Republik Indonesia. *Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 (amandemen 1999–2002), Pasal 1 ayat (3) dan Pasal 28D ayat (1)*. Jakarta; 1945.

2. Republik Indonesia. *Undang-Undang Nomor 31 Tahun 1999 sebagaimana diubah dengan Undang-Undang Nomor 20 Tahun 2001 tentang Pemberantasan Tindak Pidana Korupsi. Lembaran Negara Republik Indonesia*; 1999, 2001.
3. Daulay AF. Sentencing disparities in corruption cases and judicial discretion in Indonesian courts. Jurnal Ilmu Hukum Kyadiren. 2024;6(2):160–172. doi:10.46924/jihk.v6i2.248.
4. Zuhrah Z, Sulistiyono A, Ridwan R, et al. *Independensi hakim dalam berbagai disparitas putusan perkara korupsi di Mahkamah Agung*. Fundamental: Jurnal Ilmiah Hukum. 2024;13(1):47–70. doi:10.34304/jf.v13i1.236.
5. Puteh I, Chaniago H. *Evaluasi penerapan Peraturan Mahkamah Agung Nomor 1 Tahun 2020 tentang pedoman pemidanaan Pasal 2 dan Pasal 3 undang-undang pemberantasan tindak pidana korupsi: disparitas putusan dalam pengadilan tindak pidana korupsi di Pengadilan Negeri Medan*. Jurnal Transformasi Administrasi. 2023;13(2):217–232. doi:10.56196/jta.v13i02.253.
6. Idrus NFA. *Disparitas putusan pemidanaan perkara penipuan online*. Jurnal Yudisial. 2023;16(3):325–341. doi:10.29123/jy.v16i3.598.
7. Mabilehi MJ, Leo RP, Amalo H. *Faktor penyebab dan upaya penanggulangan disparitas putusan bagi pelaku tindak pidana kekerasan seksual di wilayah hukum Pengadilan Negeri Kupang*. Terang: Jurnal Kajian Ilmu Sosial, Politik dan Hukum. 2024;1(4):98–107. doi:10.62383/terang.v1i4.616.
8. Anderson JM, Kling JR, Stith K. Measuring inter-judge sentencing disparity: before and after the Federal Sentencing Guidelines. Journal of Law and Economics. 1999;42(S1):271–307. doi:10.1086/467426.
9. Plesničar MM. The challenges of being imperfect: how do judges and prosecutors deal with sentencing disparity. Frontiers in Sociology. 2024;9:1488786. doi:10.3389/fsoc.2024.1488786.
10. Mahkamah Agung Republik Indonesia. *Peraturan Mahkamah Agung Nomor 1 Tahun 2020 tentang Pedoman Pemidanaan Pasal 2 dan Pasal 3 Undang-Undang Pemberantasan Tindak Pidana Korupsi*. Jakarta: Mahkamah Agung RI; 2020.
11. Hariri Lubis A, Nelson FM. *Analisis Perma Nomor 1 Tahun 2020 dalam kaitannya dengan penanggulangan tindak pidana korupsi*. Jurnal Ilmu Hukum, Humaniora dan Politik. 2024;4(5):1407–1418. doi:10.38035/jihhp.v4i5.2277.
12. Andini OG, Nilasari N. *Menakar relevansi pedoman pemidanaan koruptor terhadap upaya pemberantasan korupsi*. Tanjungpura Law Journal. 2021;5(2):133–148. doi:10.26418/tlj.v5i2.46109.
13. Alief AM. Reconstruction of special sentencing guidelines on state loss crime in the Indonesian Criminal Code. Integritas: Jurnal Antikorupsi. 2024;10(1):149–160. doi:10.32697/integritas.v10i1.1069.
14. Suryana I. Construction of judicial interpretation in Indonesia's criminal justice system regarding the implementation of the new Penal Code. Indonesian Journal of Law and Justice. 2025;2(4):1–9. doi:10.47134/ijlj.v2i4.3853.
15. Mamak K, Dudek J, Koniewski M, et al. A failed attempt to radically reduce inter-court sentencing disparities by legislation: empirical evidence from Poland. European Journal of Criminology. 2022;19(6):1516–1534. doi:10.1177/1477370820952729.
16. Indonesia Corruption Watch. *Tren Vonis Kasus Korupsi Tahun 2024*. Jakarta: Divisi Hukum dan Monitoring Peradilan, ICW; 2025.
17. Kusyandi A, Yamin S. *Disparitas putusan hakim pidana berkualitas yang mencerminkan rasa keadilan dalam sistem hukum Indonesia*. Yustitia. 2023;9(1):122–132. doi:10.31943/yustitia.v9i1.173.
18. Arifia MU, Gultom BM, Markoni M. *Upaya meminimalisir disparitas putusan hakim*. Jurnal Syntax Transformation. 2023;4(1):15–31. doi:10.46799/jst.v4i1.677.
19. Nurasiah M, Harefa B, Waruwu RPR. *Disparitas pidana terhadap justice collaborator dalam tindak*

pidana korupsi. Jurnal Esensi Hukum. 2022;4(1):88–98. doi:10.35586/esh.v4i1.155.

20. Arum DS. Best practice in aggravating and mitigating factors: assessment of court decisions on corruption. Integritas: Jurnal Antikorupsi. 2022;8(2):177–184. doi:10.32697/integritas.v8i2.910.