Enigma in Law
https://enigma.or.id/index.php/law

e-ISSN 3026-6068
Vol. 4, No. 2 (2026) · pp. 82–95

ORIGINAL RESEARCH ARTICLE

Assessed at Audit, Overturned at Appeal: A Jurimetric Analysis of Transfer-Pricing Comparability and Intra-Group Service Evidence in Indonesia's Tax Court and Supreme Court

Arya Ganendra[1], Hanifah Yasin[2*], Henry Peter Paul[3]

[1] Department of Literacy, Emerald Language Institute, Balikpapan, Indonesia
[2] Department of Microeconomy, Enigma Institute, Palembang, Indonesia
[3] Department of Behavior Economy, Miami History Museum, Miami, United States

ARTICLE INFO

Received: October 18, 2026
Revised: November 5, 2026
Accepted: November 30, 2026
Available online: December 17, 2026
Published: December 30, 2026
Keywords Arm's-length principle; Empirical legal studies; Intra-group services; Jurimetrics; Transfer pricing litigation
Correspondence Hanifah Yasin · hanifah.yasin@enigma.or.id
ORCID 0009-0009-5497-8710
DOI
https://doi.org/10.61996/law.v4i2.155

ABSTRACT

Background. Indonesia consolidated its arm's-length rules under Law No. 7 of 2021, Government Regulation No. 55 of 2022, and Minister of Finance Regulation No. 172 of 2023, ahead of the 31 December 2026 transfer of Tax Court guidance to the Supreme Court (Constitutional Court Decision No. 26/PUU-XXI/2023). No jurimetric study has measured how far assessments are reduced in litigation.

Objective. To measure the erosion of transfer-pricing assessments in Indonesian courts and to examine evidentiary factors associated with it.

Methods. Eighty-seven non-overlapping income-tax transfer-pricing judgments totaling Rp 4.085 trillion were coded: fifty Tax Court appeals (Stratum I) and thirty-seven Supreme Court civil review (Peninjauan Kembali, PK) decisions (Stratum II). Analysis combined fractional logit, Firth and chamber-clustered logit, statutory interpretation, and comparative analysis.

Results. In Stratum I, 78.0% (39/50) of assessments were overturned in full (54.0%) or in part (24.0%); the value-weighted Sustained Adjustment Ratio (SAR) was 0.251 (74.9% erosion). Primary substantiation (executed contracts plus verifiable deliverables) was associated with higher overturn odds (Firth OR = 2.05, 95% CI 0.94–4.50, p = 0.073; clustered p = 0.021), whereas standalone transfer-pricing studies showed no advantage (OR = 0.51). SAR was 0.474 where comparable screening was contested versus 0.226 otherwise (unadjusted p = 0.051; not significant after adjustment). The Supreme Court rejected 89.2% (33/37) of PK petitions.

Conclusion. In this purposive corpus, assessments lost most of their value at trial and were rarely restored on review. Primary substantiation, rather than standalone studies, was associated with overturn, so six reforms are proposed to strengthen evidence quality at audit.

Copyright © 2026 The Author(s). The authors retain copyright. Published by Enigma Institute.
Access license Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International (CC BY-NC-SA 4.0). Noncommercial sharing and adaptation are permitted with attribution. Adaptations must use the same license.

1. Introduction

1.1 Constitutional, statutory, and institutional background

Article 23A of the 1945 Constitution of the Republic of Indonesia (*Undang-Undang Dasar Negara Republik Indonesia Tahun 1945*) sets the constitutional mandate for lawful taxation. Three other provisions sit beside it: the rule-of-law (*negara hukum*) principle in Article 1(3), judicial independence under Article 24(1), and the guarantee of just legal certainty in Article 28D(1). Together they require tax administration and tax adjudication to balance two aims. One is protecting sovereign revenue against cross-border profit shifting. The other is keeping evidentiary rules predictable and lawful. Indonesia is an open emerging economy. It is tied into multinational supply, licensing, management, and treasury networks. In that setting, Article 18(3) of the Income Tax Law (*Undang-Undang Pajak Penghasilan*, UU PPh; Law No. 7 of 1983 as amended by Law No. 36 of 2008 and Law No. 7 of 2021 on the Harmonization of Tax Regulations, *Undang-Undang Harmonisasi Peraturan Perpajakan* [UU HPP]) empowers the Director General

of Taxes (*Direktur Jenderal Pajak*, hereinafter DJP) to redetermine income and deductions among parties sharing a special relationship (*hubungan istimewa*) in accordance with the arm's-length principle (*prinsip kewajaran dan kelaziman usaha*, PKKU)[1,2]. The executive regime has changed recently. Over the past three years, Indonesia has recodified it through Government Regulation (*Peraturan Pemerintah*, PP) No. 55 of 2022 and Minister of Finance Regulation (*Peraturan Menteri Keuangan*, PMK) No. 172 of 2023 (effective 29 December 2023). These instruments consolidate comparability standards and contemporaneous three-tiered documentation (Master File, Local File, and Country-by-Country Report, originally introduced from fiscal year 2016 under PMK No. 213/PMK.03/2016). They also add mandatory constructive-dividend secondary adjustments (Article 37 PMK No. 172/2023). A second addition is the cumulative preliminary benefit gateway (*tahapan pendahuluan*, Article 13). The gateway was inherited from earlier DJP audit regulations (PER-43/PJ/2010 and PER-32/PJ/2011, which already contained a benefit test for intra-group services in Article 14, and PER-22/PJ/2013, the transfer-pricing audit guideline)[3,4].

A deep institutional paradox separates Indonesia's administrative audit stage from its judicial appeal stage. Consider the sequence. Under Law No. 14 of 2002 on the Tax Court (*Undang-Undang Pengadilan Pajak*), assessments are upheld at DJP's internal objection (*keberatan*) tier, after passing through narrow pre-assessment review by DJP's Audit Quality Assurance Teams (*Tim Quality Assurance Pemeriksaan*, Tim QAP). Then they are adjudicated *de novo* on the facts. Specialized three-judge chambers (*Majelis*) of the Tax Court hear them, governed by the doctrine of substantive material truth (*asas kebenaran materiil* / substance over form, Articles 69, 76, and 78). Tax Court judgments are immediately final and executory (*inkracht van gewijsde*, Articles 77(1) and 80)[5]. No ordinary cassation (*kasasi*) is allowed. Only extraordinary civil review (*Peninjauan Kembali*, PK) before the Supreme Court (*Mahkamah Agung*) remains open, and only on narrow statutory grounds such as newly discovered decisive documents (*novum*, Article 91(b)) or manifest contravention of prevailing legislation (*nyata-nyata tidak sesuai dengan ketentuan peraturan perundang-undangan yang berlaku*, Article 91(e))[6]. Constitutional Court (*Mahkamah Konstitusi*) Decision No. 26/PUU-XXI/2023 (pronounced 25 May 2023) compounds this structural tension. The decision declared the two-roof governance of the Tax Court under Article 5(2) of Law No. 14/2002 conditionally unconstitutional. It held that the phrase "*Departemen Keuangan*" in Article 5(2) is unconstitutional unless read as the Supreme Court, with organizational, administrative, and financial guidance transferred in phases to the Supreme Court under a unified one-roof judicial system (*sistem satu atap*) no later than 31 December 2026[7–9]. On the eve of that December 2026 deadline, Indonesian tax governance faces an urgent empirical question. Why were so many high-value transfer-pricing adjustments assessed by DJP auditors cancelled in full or in part by the Tax Court in the judgments studied? And why did Supreme Court civil review restore so few of them (4 of 37 PK petitions)? Table 1 maps the primary domestic, historical, international, and comparative legal instruments that frame this inquiry. It also notes their *tempus regit actum* applicability.

Table 1. Primary Indonesian, international, and comparative legal instruments analyzed.

Instrument	Year	Jurisdiction	Key provisions & tempus regit actum scope	Status
1945 Constitution of the Republic of Indonesia (UUD NRI 1945)	1945/2002	Indonesia	Art. 1(3) *negara hukum*; Art. 23A statutory basis of taxation; Art. 24(1) judicial independence; Art. 28D(1) just legal certainty	In force
Income Tax Law (Law No. 7/1983 as amended by Law No. 36/2008 & Law No. 7/2021, UU HPP)	1983/2021	Indonesia	Art. 4(1)(g) disguised dividends; Art. 6(1)(a) 3M deductibility; Art. 18(1) limits on deductible borrowing costs (debt–equity ratio, DER, in the pre-2021 wording); Art. 18(3)–(4) arm's-length redetermination (applied across all 87 dockets)	In force
General Tax Provisions and Procedures Law (Law No. 6/1983 as amended by Law No. 7/2021, UU KUP) & Audit QAP Rules (PMK 184/2015)	1983/2023	Indonesia	Art. 13(1) five-year assessment period; Art. 26A(4) preclusion of records withheld at audit (objection stage); Art. 28 bookkeeping; Elucidation Art. 29(2) para. 3 competent proof; Tim QAP described by the authors as limited to legal-basis (non-factual) audit review	In force
Historical Audit & Financing Rules (SE-04/PJ.7/1993; KMK 254/1985; PER-43/2010 jo. PER-32/2011; PER-22/PJ/2013; PMK 169/2015; PMK 213/2016)	1985/2016	Indonesia	Governed 2003–2015 audited tax years ratione temporis; KMK 254/1985 suspended DER until PMK 169/2015 (4:1 from FY 2016); PMK 213/2016 mandated ex-ante TP Doc from FY 2016	Superseded / Consolidated into PMK 172/2023
Government Regulation (PP) No. 55 of 2022 & Minister of Finance Regulation (PMK) No. 172 of 2023	2022/2023	Indonesia	PP 55/2022 Arts. 32–43 (Art. 36(6) secondary adjustment); PMK 172/2023 Art. 13 cumulative preliminary benefit gateway, Art. 37 mandatory secondary adjustment, Arts. 41–54 MAP; Arts. 55–71 APA	In force since 29 Dec 2023
Law No. 14 of 2002 on the Tax Court (UU *Pengadilan Pajak*)	2002	Indonesia	Art. 5(2) dual governance; Art. 69(1) evidence; Art. 76 active judicial fact-finding; Art. 77(1) finality; Art. 78 material truth; Art. 77(3) PK standing; Art. 91 PK grounds; Arts. 92–93 time limits and decision periods	In force (one-roof transfer due 31 Dec 2026)
Constitutional Court Decisions No. 26/PUU-XXI/2023 & No. 24/PUU-XXII/2024	2023/2024	Indonesia	26/PUU-XXI/2023: mandates Tax Court one-roof transfer to Supreme Court by 31 Dec 2026; 24/PUU-XXII/2024: bars government agency PK in general administrative (PTUN) suits	Binding constitutional mandates
Law No. 28 of 2014 on Copyright (UU Hak Cipta)	2014	Indonesia	Art. 42(d): no copyright in court decisions or judges' orders (*putusan pengadilan atau penetapan hakim*)	In force
OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations	2022	International (OECD/G20)	Ch. I–III comparability; Ch. IV (4.95–4.133) bilateral safe-harbor design; Ch. VI DEMPE intangibles; Ch. VII (7.43–7.65) elective 5% LVAS simplified approach	Global soft-law benchmark

Instrument	Year	Jurisdiction	Key provisions & tempus regit actum scope	Status
Income-tax Act, 1961 & Income-tax Rules, 1962	1961/2024	India	s. 92CA Transfer Pricing Officer; s. 144C collegial Dispute Resolution Panel (DRP) that may confirm, reduce or enhance the variations proposed in the draft order; Rules 10TD–10TG Safe Harbors	Actively applied
Taxation Administration Act 1953 & ATO Practical Compliance Guidelines (PCG 2017/2 & PCG 2017/4)	1953/2024	Australia	Subdiv. 815-B ITAA 1997 ALP reconstruction; PCG 2017/2 simplified transfer-pricing record-keeping options; PCG 2017/4 quantitative traffic-light risk zones for affiliate loans	Actively applied
Income Tax Act 1947 & IRAS Transfer Pricing Guidelines (7th Edition)	1947/2024	Singapore	s. 34D arm's-length principle; s. 34E 5% surcharge; s. 34F mandatory contemporaneous TP documentation; IRAS TPG 7th ed. elective 5% cost-plus routine support service safe harbor	Actively applied

Note. Primary statutes, regulations, constitutional judgments, and administrative guidelines are cited by official title and provision in the text and mapped in Table 1; the numbered reference list (1–40) contains scholarly works with verified Digital Object Identifiers (DOIs).

1.2 Litigation-selection theory and research gap

Economic theories of litigation selection and international tax jurisprudence give sharp analytical lenses for diagnosing asymmetric trial overturn rates. Start with Priest and Klein. In their divergent-expectations model, symmetric stakes and unbiased error variance push trial win rates toward 50%. Persistent departures from 50% instead point to one of two causes. One is differential stakes, such as multinational groups defending multi-year regional pricing architectures. The other is structural divergence in standards between administrative assessment and judicial adjudication[10,11]. Empirical tests of selection effects on tax decisions have returned mixed results[12]. Doctrinal and statistical scholarship on transfer pricing shows that arm's-length comparability rarely yields a unique point truth, because functional-asset-risk (FAR) characterization, multi-year averaging, loss-maker screening, and interquartile range (IQR) boundaries each open wide zones of defensible pricing[1,13–15]. Revenue auditors facing tight statutory limitation periods (five years under Article 13(1) of the General Tax Provisions and Procedures Law, *Undang-Undang Ketentuan Umum dan Tata Cara Perpajakan*, UU KUP) may bypass quantitative comparability altogether, and they disallow 100% of intra-group service fees or royalty deductions to Rp 0, asserting categorically that "economic benefit was not proven". In this corpus, trial chambers governed by substance over form overturned 19 of 24 such unbenchmarked zeroing dockets (79.17%; Table 7).

Recent empirical studies in Enigma in Law show that jurimetrics has diagnostic value for quantifying divergences between statutory design and institutional practice, with examples including unexplained variance in post-guideline corruption sentencing[16], prohibited standard clauses persisting across digital contracts[17], and administrative integrity bottlenecks in regional governance[18]. The tax literature is smaller. Tambunan, Rosdiana, and Irianto studied manufacturing-sector Tax Court decisions from 2014–2019. Their topics were audit performance (2015–2019)[19], intangibles[20], intra-group financing[21], and settlement pathways[22]. Content analyses of Tax Court transfer-pricing decisions from 2020–2023 follow a similar descriptive approach[23]. Four empirical and doctrinal gaps remain. First, prior studies rely on binary win/loss descriptions rather than continuous Sustained Adjustment Ratios (SAR, the sustained share of the disputed adjustment), and this masks how *Mengabulkan Sebagian* (partial split) verdicts allocate fiscal value. Second, to our knowledge no study has jointly modeled Tax Court appeals alongside Supreme Court PK substantive decisions while controlling for cross-stratum selection and summary-length heterogeneity. Third, prior commentary has not tested econometrically why pre-PMK 213/2016 standalone transfer-pricing reports failed in court. The comparison is with primary commercial contracts and *Uji Bukti* deliverables. Fourth, to our knowledge no study has connected this pre-reform judicial baseline (tax years 2003–2015, adjudicated 2011–2020) to the institutional risks of PMK No. 172/2023's cumulative benefit gateway (Article 13) and mandatory secondary dividend adjustment (Article 37) on the eve of the 31 December 2026 one-roof transfer.

1.3 Aims and hypotheses

To address these gaps, we examine eighty-seven (N = 87) non-overlapping Indonesian income-tax transfer-pricing substantive judgments covering Rp 4.085 trillion in disputed adjustments. We test four hypotheses, stated as expectations about direction and approximate magnitude. H1 (Asymmetric Trial Erosion and Partial-Split Skew) holds that Tax Court appeals exhibit audit overturn rates exceeding 70% and a value-weighted Sustained Adjustment Ratio (SAR) near 0.25, both at the whole-docket net-income level and on isolated transfer-pricing line items, while partial-split rulings cancel roughly 70%–75% of disputed adjustments. H2 (Primary Transactional Substantiation vs. Ex-Post TP Studies) holds that primary substantiation (executed intercompany contracts plus verifiable performance deliverables) is associated with higher overturn odds and lower SAR, controlling for stratum fixed effects, whereas standalone transfer-pricing studies confer no litigation advantage . H3 (Quantitative Benchmarking vs. Intra-Group Service Gateway Cliffs) holds that DJP sustainment ratios are higher where the record shows quantitative comparability screening, or where statutory preclusion or bright-line financing rules apply, than under unbenchmarked zeroing (variable D3 codes a record in which comparable screening is contested; Table 2). Intra-group service (IGS) disputes, in turn, show lower overturn rates, which would be consistent with preliminary benefit-gateway effects. H4 (Supreme Court Article 91(e) Finality Ratchet) concerns the final stage. DJP initiates nearly 60% of PK petitions. Even so, the Supreme Court affirms nearly 90% of Tax Court rulings symmetrically under Article 91(e).

2. Methods

2.1 Integrated jurimetric, doctrinal, and comparative design

This study uses an integrated empirical–doctrinal legal research design that combines archival jurimetrics[16,17], four-canon statutory interpretation *de lege lata*, and comparative functional-equivalence analysis[24] across four international benchmarks: the OECD Transfer Pricing Guidelines (2022), Singapore, India, and Australia. The design has a limit. Publicly accessible court decisions form a purposive archive of litigated disputes, and they are not a random draw from confidential tax audits (protected by tax secrecy under Article 34 UU KUP). All inferential quantities are therefore interpreted candidly. These quantities are p-values, 95% confidence intervals, and odds ratios. We treat them as heuristic measures of estimation precision and abductive evidential weight within the adjudicated corpus.

2.2 Corpus assembly, screening, and companion-cluster tracking

The harvest respected HTTP 403 and SSL perimeter constraints on putusan3.mahkamahagung.go.id and setpp.kemenkeu.go.id, without any circumvention. Across eleven transfer-pricing queries, we assembled 356 unique full-text candidate tax judgments (228 Supreme Court PK and 128 Tax Court appeal records) from publicly indexed judicial repositories, including the two portals named above. Article 42(d) of Law No. 28 of 2014 on Copyright provides that no copyright subsists in court decisions or judges' orders (*putusan pengadilan atau penetapan hakim*), so all 356 records were stripped of third-party interface chrome. Only public-domain judicial texts remained. We verified them via SHA-256 digests (manifest available on request).

Monthly Value-Added Tax (PPN) derivative assessments can create pseudo-replication. In practice, one corporate audit may spawn 12 monthly VAT dockets. To prevent this, deterministic PRISMA-style screening (Table 2; exclusion log available on request) excluded 269 records. These were 115 monthly VAT/final-tax dockets E3; 73 purely domestic non-ALP disputes E5; 25 customs/land-tax dockets E1; 21 payroll/other taxes E4; 15 procedural *Gugatan* .99 dockets E2; 14 duplicates E7; and 6 formal *Niet Ontvankelijk Verklaard* [NO] dismissals E8. The result was N = 87 non-overlapping substantive income-tax transfer-pricing decision dockets (TP01–TP50 in Stratum I full-text Tax Court appeals; TP51–TP87 in Stratum II Supreme Court PK decisions; 0 cross-stratum docket overlaps). A PK petition is filed by the party that lost at trial, so Stratum II is selected on the trial outcome. In Stratum II, the trial outcome (overturn) is coded from the condensed account of the Tax Court judgment in each PK decision (Table 2). Some dockets remained related. Related audit cycles within the same stratum also produced multi-year or corporate-withholding (PPh Badan + PPh Pasal 23/26) companion dockets (9 secondary sibling dockets flagged via audit_cluster_id). Chamber-clustered (CR1) standard errors account for them. They absorb intra-chamber correlation across all 16 chambers. Sensitivity Check S9 then re-estimates outcomes on the strictly de-duplicated N = 78 primary-cluster corpus.

Table 3. Operational jurimetric coding scheme, statutory anchors, and measurement rules (N = 87).

Variable code	Legal / methodological construct	Statutory / doctrinal anchor	Operational measurement rule
sar & tp_specific_sar	Whole-Docket & Pure TP-Item Sustained Adjustment Ratios (Stratum I, $n_1 = 50$)	Art. 18(3) UU PPh; Art. 80(1) Law 14/2002	SAR_i = Sustained_i / Disputed_i ∈ [0, 1] (0.0 full taxpayer win, 1.0 full DJP win, exact final-table fraction for all 12 partial splits; tp_specific_sar isolates affiliate TP items only)
pp_overturn_binary	Tax Court audit overturn / erosion (N = 87)	Art. 80(1)(b)–(c) Law 14/2002	1 if Tax Court cancelled DJP's adjustment in full (*Mengabulkan Seluruhnya*) or part (*Mengabulkan Sebagian*); 0 if rejected (*Menolak Banding*)
pk_affirmed_binary	Supreme Court PK affirmance (Stratum II, $n_2 = 37$)	Art. 91(e) & Art. 93 Law 14/2002	1 if Supreme Court rejected PK (*Menolak PK*), affirming Tax Court; 0 if granted (*Mengabulkan PK* & Mengadili Kembali)
stratum_ii_indicator	Stratum II (MA_PK) cohort fixed effect	Art. 91 Law 14/2002; Priest–Klein selection	1 if observation belongs to Stratum II (TP51–TP87, condensed PK transcript); 0 if Stratum I (TP01–TP50, full trial transcript)
tx_archetype	Primary disputed transaction archetype	PMK 172/2023 Arts. 4, 13	Mutually exclusive dominant dispute: Intangibles_Royalty (n=34), Tangible_Goods_Margin (n=26), Affiliate_Financing (n=14), Intra_Group_Services (n=13)
d1_unbenchmarked_zeroing	Unbenchmarked categorical audit zeroing	Art. 18(3) UU PPh; Art. 13 PMK 172/2023	1 if audit disallowed 100% to Rp 0 on existence/benefit/dividend grounds without an external comparable range (n=24); 0 otherwise (n=63)
d2_audit_preclusion_brightline	Audit document preclusion or bright-line financing rule	Art. 26A(4), Art. 28 UU KUP; Art. 18(1) UU PPh	1 if DJP invoked Art. 26A(4)/28 audit-withholding preclusion or debt–equity / S-165 bright-line loan rules (n=16); 0 otherwise (n=71)
d3_comparability_contested	Quantitative comparability benchmark contested	PMK 172/2023 Arts. 8–12; OECD TPG Ch. III	1 if record contests empirical comparable screening (loss-makers, single- vs multi-year window, functional mismatch; n=15); 0 otherwise (n=72)
e1_contemporaneous_tpdoc	Standalone TP Study / Documentation tendered	PER-32/2011; PMK 213/2016; PMK 172/2023	1 if taxpayer tendered a formal TP Study, Transfer Pricing Report, or Local File (n=20; predominantly ex-post in 2003–2015 tax years); 0 if absent (n=67)
e2_intercompany_contract	Executed written intercompany agreement	Art. 69(1)(a) Law 14/2002	1 if taxpayer tendered an executed license, technical assistance, management, loan, or sales agreement (n=19); 0 otherwise (n=68)
e3_verifiable_deliverables	Primary commercial / technical deliverables	Arts. 69, 76, 78 Law 14/2002	1 if taxpayer tendered primary performance deliverables (*Uji Bukti* IP certificates, timesheets, PEB/BL customs records, audited FS, vouchers; n=38); 0 otherwise (n=49)
primary_substantiation_score	Primary transactional substantiation (0–2)	Art. 78 Law 14/2002 (*kebenaran materiil*)	Additive primary documentary index E2 + E3 (0, 1, or 2) measuring executed contractual + verifiable physical/commercial performance proof

Note. All 87 × 38 coded values and verbatim Bahasa Indonesia holdings are available from the corresponding author on reasonable request.

2.3 Operational coding, dual SAR metrics, and reliability audit

To limit outcome-concomitant circularity, judicial holding sentences were stripped before upstream audit and documentary indicators were extracted (Table 3). Some partial-split appeals are multi-issue. Domestic non-transfer-pricing items sit there beside affiliated transactions. For these appeals, we construct both whole-docket Net-Income SAR and isolated Transfer-Pricing-Specific SAR (pure TP-item SAR), removing purely domestic inventory or tax-credit adjustments from the numerator and denominator (Sensitivity Check S8).

Table 2. PRISMA screening flow, two-tier stratum architecture, Tax Court chamber distribution, and disputed fiscal stakes (N = 87).

Corpus dimension / Chamber group	Candidate / Stratum I (n_1)	Stratum II (n_2)	Pooled Merits (N)	Disputed Fiscal Stakes (Rp Billions)	Share of Stakes (%)
Panel A: PRISMA screening from harvested repository (356 records)					
Total full-text candidate records harvested & SHA-256 hashed	128	228	356	—	—
Excluded: E3 Monthly VAT (PPN) / Final Tax pseudo-replicated dockets	22	93	115	—	—
Excluded: E5 Domestic deductibility / non-ALP income-tax disputes	29	44	73	—	—
Excluded: E1 Customs (Bea Cukai) & E4 Other non-CIT/WHT taxes	16	30	46	—	—
Excluded: E2 *Gugatan* (.99), E7 Duplicates (14), E8 Formal NO (6)	11	24	35	—	—
Included transfer-pricing merits dockets (87 dockets in 78 clusters)	50	37	87	4,084.55	100.00%
Panel B: Tax category & bookkeeping currency (N = 87)					
Corporate Income Tax (PPh Badan, .15)	37	33	70	3,968.87	97.17%
Domestic / Cross-Border Affiliate Withholding (PPh Pasal 23 & Pasal 26)	13	4	17	115.68	2.83%
Indonesian Rupiah (IDR) bookkeeping dockets	29	21	50	2,866.11	70.17%
US Dollar (USD) functional-currency bookkeeping dockets	21	16	37	1,218.44	29.83%
Panel C: Tax Court chamber (Majelis) coverage (16 chambers)					
High-volume chambers: M.I (11), M.VI (9), M.XI (9), M.XVIII (9)	26	12	38	886.88	21.71%
Mid-volume chambers: M.II (8), M.III (7), M.V (5), M.XII (5)	16	9	25	1,126.94	27.59%
Specialized/other chambers: M.IV, VII, VIII, X, XIII, XIV, XV, XVI (24)	8	16	24	2,070.73	50.70%

Note. Audited tax years span 2003–2015 and judicial decision years span 2011–2020 (TP01–TP50 = Stratum I; TP51–TP87 = Stratum II). Median disputed adjustment across N = 87 is Rp 19.700 billion (IQR: Rp 6.440–44.467 billion; *Gini* = 0.6681; CV = 1.7107).

Rule-extraction validity was checked through a case-by-case Algorithmic–Human Construct Reliability Audit. The audit drew a stratified random 25.29% subsample (n = 22 judgments: 14 Stratum I + 8 Stratum II) and compared automated lexical coding with full-text human legal reading by one author (see Section 4.4), archived row by row (available on request)[25,26]. Table 4 reports the results. Dispositive judicial outcomes reached 100.00% agreement (κ = 1.000). Binary primary substantiation reached 90.91% (κ = 0.820, PABAK = 0.818), unbenchmarked zeroing 86.36% (κ = 0.645, PABAK = 0.727), and four-class transaction archetypes 72.73% (κ = 0.634, multi-category PABAK = 0.636). Every one of these cleared the 0.60 agreement threshold of Landis and Koch[25]. The audit also exposed, transparently, the bundled multi-issue dockets. TP67 is one. It combines a Rp 22.79B inter-location service recharge with a Rp 7.40B royalty.

Table 4. Case-by-case algorithmic–human construct reliability audit (n = 22 judgments).

Coded construct	Subsample (n)	Discrepancies	Raw agreement (%)	Cohen's κ	PABAK	Substantive source of human–regex edge cases
Dispositive outcome (overturn / affirm)	22	0	100.00%	1.000	1.000	Exact 100% match across operative amar and recalculation tables
Primary substantiation (E2 + E3 ≥ 1)	22	2	90.91%	0.820	0.818	Concise MA_PK summaries (TP21) or unaccepted invoice exhibits (TP09)
Unbenchmarked audit zeroing (D1)	22	3	86.36%	0.645	0.727	Multi-item dockets zeroing royalties (TP11, TP35, TP52) alongside benchmarked items
Transaction archetype (4 classes)	22	6	72.73%	0.634	0.636	Bundled multi-issue dockets (TP14 Rp 26.0B loan FX + Rp 1.1B royalty; TP67 recharge + royalty)

2.4 Econometric specification and cross-stratum control

Model 1A fits the bounded Sustained Adjustment Ratio $SAR_i \in [0, 1]$ in Stratum I ($n_1 = 50$; boundary mass points at 0.0 and 1.0 plus 12 interior fractions) with Papke and Wooldridge's Fractional Logit Quasi-Maximum Likelihood Estimator (QMLE)[27]:

$$\ell_i(\beta) = SAR_i \cdot \ln \Lambda(x_i^T\beta) + (1 - SAR_i) \cdot \ln [1 - \Lambda(x_i^T\beta)], \quad \Lambda(z) = 1 / (1 + e^{-z})$$

with Liang–Zeger chamber-clustered sandwich standard errors, using the CR1 small-sample correction $[G/(G-1)][(n-1)/(n-k)]$ across G = 13 active Stratum I chambers[28,29], and a CR1 linear OLS counterpart (Model 1B) is paired with it and was re-estimated on isolated pure TP-item SAR (Sensitivity S8). Model 2 takes binary Tax Court overturn $Y_i \in \{0, 1\}$ across the pooled corpus (N = 87, G = 16 chambers)

and estimates two versions: Firth's Jeffreys-prior penalized logistic regression[30,31] and a CR1 chamber-clustered logit. Both use an identical five-variable substantive vector (the primary substantiation score E2 + E3, the standalone TP study indicator E1, the comparable-screening-contested indicator D3, the IGS archetype indicator, and mean-centered $\log_{10}$ disputed adjustment in Rupiah, with USD dockets converted at Bank Indonesia exchange rates; Sensitivity S7 uses a constant rate). Each also adds an explicit Stratum II cohort fixed effect. This term absorbs systematic differences in transcript condensation (mean primary substantiation = 0.880 in Stratum I vs. 0.324 in Stratum II) and Priest–Klein PK appeal selection, while Sensitivity S6 also reports unadjusted N = 87 and Stratum I-only estimates. Multicollinearity is negligible. Max VIF = 1.275 in Model 1. Max VIF = 1.242 in Model 2.

3. Results

3.1 Two-tier empirical baseline

Table 5 and Figure 1A report the empirical baseline, which supports H1 in this corpus. In Stratum I (n_1 = 50), taxpayers secured complete cancellation (*Mengabulkan Seluruhnya*) in 27 / 50 appeals (54.00%) and partial cancellation (*Mengabulkan Sebagian*) in 12 / 50 appeals (24.00%), leaving DJP with complete sustainment (*Menolak Banding*) in only 11 / 50 appeals (22.00%). The trial overturn/erosion rate was therefore 78.00% (39 / 50; exact 95% CI: 64.04%–88.47%) (72.41%, 63 / 87, 95% CI: 61.79%–81.46% across the pooled N = 87 baseline; 73.08% on the N = 78 cluster-deduplicated corpus in S9).

Table 5. Tier I (*Pengadilan Pajak*) appellate dispositions and whole-docket vs. pure transfer-pricing Sustained Adjustment Ratios (SAR).

Judicial outcome / Fiscal metric	Stratum I Trial Corpus (n_1 = 50)	Pooled Trial Baseline (N = 87)	Exact 95% CI / Counterpart	Fiscal & doctrinal interpretation
Taxpayer Full Win (*Mengabulkan Seluruhnya*, SAR = 0.0)	27 (54.00%)	42 (48.28%)	S1 [39.32%, 68.19%]; pooled [37.42%, 59.25%]	Complete judicial nullification of audited TP adjustment
Partial Split (*Mengabulkan Sebagian*, 0 < SAR < 1)	12 (24.00%)	21 (24.14%)	S1 [13.06%, 38.17%]; pooled [15.60%, 34.50%]	Selective item or quartile recalibration under Art. 80(1)(b)
DJP Full Win (*Menolak Banding*, SAR = 1.0)	11 (22.00%)	24 (27.59%)	S1 [11.53%, 35.96%]; pooled [18.54%, 38.21%]	Full sustainment of audit redetermination
Combined Audit Overturn / Erosion Rate	39 (78.00%)	63 (72.41%)	S1 [64.04%, 88.47%]; pooled [61.79%, 81.46%]	Nearly four-fifths of audited TP assessments fail intact
Unweighted Mean SAR (Whole-Docket vs. Pure TP-Item)	0.2807	—	Pure TP SAR = 0.2916	Removing non-TP domestic items leaves 70.84% cancelled
Mean SAR within Partial Splits (n = 12, Whole vs. Pure TP)	0.2529	—	Pure TP SAR = 0.2983	Partial splits cancel 70.17%–74.71% of disputed adjustments
Total Disputed Adjustments (IDR)	Rp 1,635.87 B	Rp 4,084.55 B	*Gini* = 0.6681 (Pooled)	Top 15% of dockets hold 62.3% of Rp 4.085T stakes
Total Sustained vs. Cancelled (Stratum I, IDR)	Rp 410.02 B	—	Rp 1,225.86 B cancelled	Rp 1.226 trillion extinguished across 50 trial appeals
Value-Weighted SAR (Whole-Docket vs. Pure TP-Item)	0.2506 (25.06%)	—	Pure TP SAR = 0.2552	74.48%–74.94% of audited Rupiah stakes wiped out

Note. Pure TP-Item SAR isolates affiliated transfer-pricing items inside multi-issue partial splits (e.g., coding TP07 PUT-110673 as 1.0000 on sustained royalty while coding TP40 Put.69420, TP11 Put-44616, and TP28 Put.63933 as 0.0000 on 100% cancelled royalties).

Both whole-docket SAR (mean = 0.2807, value-weighted = 0.2506) and isolated Pure TP-Item SAR (mean = 0.2916, value-weighted = 0.2552) show the same pattern: three-quarters (74.48%–74.94%, or Rp 1.226 trillion) of audited transfer-pricing adjustments evaporate at trial. Even inside *Mengabulkan Sebagian* verdicts (n = 12), the Tax Court cancels 70.17%–74.71% of disputed adjustments (mean whole-docket SAR = 0.2529; mean pure-TP SAR = 0.2983). In Put-115639.15/2014/PP/M.XIB Tahun 2019 (TP06), *Majelis* XIB annulled 100% of the Rp 15.688B service-fee and Rp 314.88M trademark disallowances made by DJP, and only Rp 770.29M of unverified coordination fees was sustained (SAR = 0.0409; pure TP-item SAR = 0.0421). In Put.69420/PP/M.XVB/15/2016 (TP40), *Majelis* XVB cancelled 100% of DJP's USD 3,002,761.25 technical-assistance royalty disallowance (pure TP-item SAR = 0.0000) and sustained only a domestic USD 48,521.31 tax-credit adjustment (whole-docket SAR = 0.0159). Docket PUT.69375/PP/M.XIIB/15/2016 (TP43) went the same way. *Majelis* XIIB rejected DJP's Rp 43.970B know-how disallowance. It recalibrated the arm's-length margin to the third quartile (2.87%). That cancelled Rp 20.257B of the Rp 25.018B net-income dispute (SAR = pure TP-item SAR = 0.1903).

3.2 Divergence across transaction archetypes

Table 6 and Figure 1B disaggregate outcomes across the four transaction archetypes (N = 87, Rp 4,084.55 billion). Intangibles and Royalties constitute the most frequent litigation flashpoint (N = 34; Rp 1,240.84 billion), with 76.47% (26 / 34; 80.00% in Stratum I) overturned and mean SAR = 0.2475 (pure TP-item SAR = 0.2756). In dockets where taxpayers produced registered trademark/patent certificates, executed license agreements, and PPh Pasal 26 withholding vouchers, as in Put-44616/PP/M.III/15/2013 (TP11, Australian-registered "GRANITO" trademark, pure TP-item SAR = 0.0000) and 486/B/PK/Pjk/2018 (TP73, affirming cancellation of Rp 193.29B royalty and Rp 104.66B interest adjustments), chambers rejected DJP's categorical Rp 0 disallowances. Conversely, where a subsidiary tendered only a photocopy blowmolding evaluation note and a visitor visa without proving patent deployment (PUT-110673.15/2013/PP/M.VB Tahun 2018, TP07), *Majelis* VB sustained 100% of the Rp 3.947B royalty disallowance (pure TP-item SAR = 1.0000, even though cancelling a separate Rp 8.830B domestic inventory loss yielded whole-docket SAR = 0.3089):

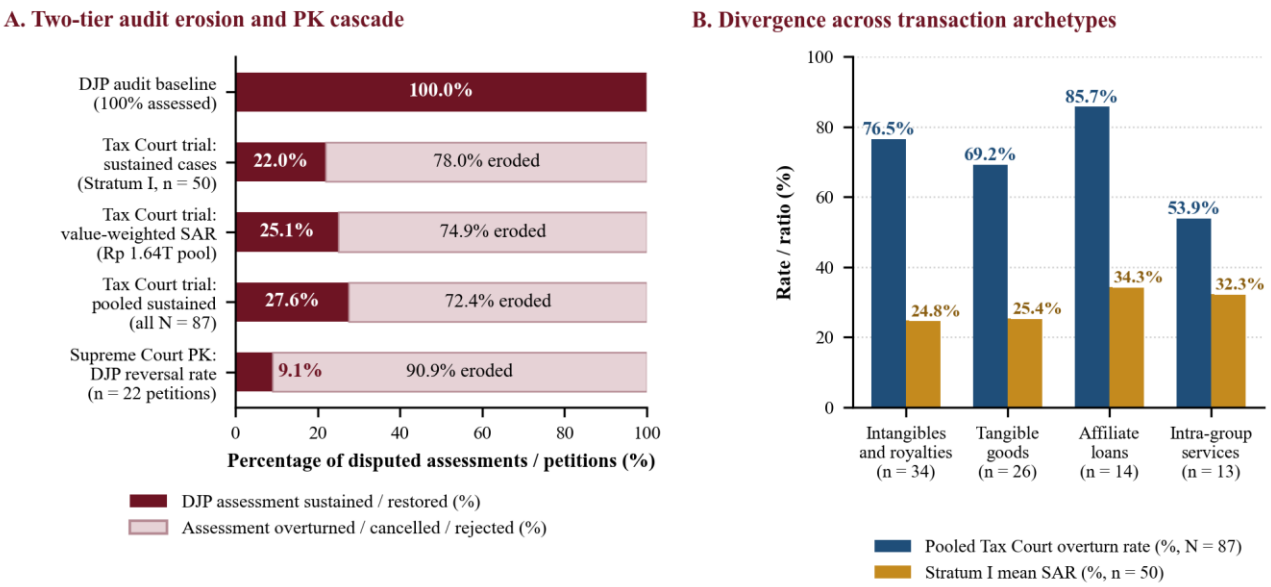

Figure 1. Two-tier transfer-pricing adjudication attrition and transaction-archetype divergence. (A) Cascade from 100% DJP audit baseline through Stratum I case sustainment (22.0%), Rupiah value-weighted sustainment (25.1%), pooled trial sustainment (27.6%), and Supreme Court PK reversal of taxpayer trial wins (9.1%). (B) Comparison of pooled Tax Court overturn rates (N = 87) and Stratum I mean Sustained Adjustment Ratios (SAR, $n_1 = 50$) across the four disputed transaction archetypes.

> *"bahwa bukti berupa fotokopi Report Hasil Evaluasi NG Mesin Blowmolding ... adalah bukti yang tidak cukup kuat untuk membuktikan bahwa hal-hal tersebut terkait dengan hal-hal yang diperjanjikan dalam Perjanjian Asistensi Teknik ... maka Majelis berketetapan bahwa koreksi Terbanding atas Biaya Royalti sebesar Rp3.947.008.826, tetap dipertahankan." (PUT-110673.15/2013/PP/M.VB Tahun 2018, TP07).*

Table 6. Disputed transaction archetypes, audit overturn rates, dual SAR metrics, and landmark docket rulings (N = 87).

Transaction archetype	Pooled N (S1 / S2)	Disputed Stakes (Rp B)	Pooled Overturn k (%) [95% CI]	S1 Overturn (%)	S1 Mean SAR (Pure TP SAR)	Representative docket & core ratio decidendi
Intangibles & Royalties	34 (20 / 14)	1,240.84	26 (76.47%) [58.83, 89.25]	80.00%	0.2475 (0.2756)	TP11 Put-44616 (GRANITO IP registry cancels royalty) & TP43 PUT.69375 (Q3 2.87% know-how calibration) vs. TP07 PUT-110673 (tp_sar = 1.0)
Tangible Goods & Operating Margin	26 (12 / 14)	2,262.10	18 (69.23%) [48.21, 85.67]	75.00%	0.2540 (0.2509)	TP46 Put-57477 (USD 38.29M full win) & TP15 PUT.28315 (Rp 96.42B full win) vs. TP55 2198/B/PK/Pjk/2020 (DJP PK win via 5-yr 5.14% ROS)
Affiliate Loans & Financing	14 (11 / 3)	288.90	12 (85.71%) [57.19, 98.22]	81.82%	0.3432 (0.3450)	TP24 PUT-106925 (USD 282k Art. 18(3) debt–equity sustained) vs. TP36 Put.73688 (2–1 dissent cancelling USD 81.3k interest during pre-2016 DER suspension)
Intra-Group Services (IGS)	13 (7 / 6)	292.71	7 (53.85%) [25.13, 80.78]	71.43%	0.3231 (0.3231)	TP06 Put-115639 (Rp 15.69B service fee cancelled on deliverables) vs. TP35 Put.48349 (Rp 25.27B Dutch parent commission sustained 100%)
All Transfer-Pricing Merits	87 (50 / 37)	4,084.55	63 (72.41%) [61.79, 81.46]	78.00%	0.2807 (0.2916)	Kruskal–Wallis across S1 SAR distributions: H = 0.662, p = 0.8820

Tangible Goods and Operating Margin dockets (N = 26, holding Rp 2,262.10 billion or 55.38% of total fiscal stakes) record a 69.23% pooled overturn rate (mean SAR = 0.2540). Comparable-screening battles under the transactional net margin method (TNMM) drive this result (TP46, USD 38.29M full win). Affiliate Loans and Financing (N = 14) show the highest Stratum I mean SAR = 0.3432 (pure TP-item SAR = 0.3450). They teach a *tempus regit actum* lesson. Minister of Finance Decree KMK No. 254/KMK.01/1985 had suspended Indonesia's quantitative debt–equity ratio (DER) for thirty years, until PMK No. 169/PMK.010/2015 activated the 4:1 DER in fiscal year 2016, and pre-2016 financing audits therefore divided panels sharply. Recharacterization was sustained where shareholder debt dwarfed paid-in capital (TP24), Rp 95.63B was cancelled where audited loan documentation held (TP13), and one panel split 2–1 with a formal dissenting opinion on affiliate loan interest in Put.73688/PP/M.XVIIIA/15/2016 (TP36, fiscal year 2011). Intra-Group Services (IGS, N = 13) show the lowest pooled overturn rate (53.85%, 7 / 13), which is consistent with a benefit-gateway effect[32–35] (the mechanism is not measured; IGS versus other archetypes, Fisher p = 0.1745), and disallowances stood intact where subsidiaries failed to separate shareholder stewardship from subsidiary-specific services. One was a Rp 25.27B commission to 99% parent FF BV (TP35 Put.48349). Another was a Rp 5.96B management fee (TP52 3646/B/PK/Pjk/2020).

3.3 Econometric determinants of overturn and sustainment

Tables 7 and 8 and Figure 2 present our aligned bivariate and multivariate tests of H2 and H3, incorporating the Stratum II

cohort fixed effect to account for cross-stratum differences (the fixed effect does not remove selection on the trial outcome).

Table 7. Bivariate non-parametric (Stratum I SAR) and exact (Pooled N = 87 Overturn) associations with Holm multiplicity adjustment.

Methodological / Evidentiary indicator	S1 1 / 0 (n)	S1 Mean SAR (1 vs. 0)	ΔSAR	Mann–Whitney U (p_raw / p_Holm)	Pooled 1 / 0 (N)	Pooled Overturn (1 vs. 0)	Fisher OR (p_raw / p_Holm)
Executed intercompany contract (E2)	15 / 35	0.1123 vs. 0.3529	-0.2406	213.0 (0.2555 / 1.0000)	19 / 68	89.47% vs. 67.65%	4.065 (0.0820 / 0.4918)
Verifiable deliverables (E3)	30 / 20	0.2640 vs. 0.3058	-0.0418	316.0 (0.7365 / 1.0000)	38 / 49	81.58% vs. 65.31%	2.353 (0.1460 / 0.7299)
Standalone TP Study / Doc (E1)	17 / 33	0.3967 vs. 0.2210	+0.1757	361.0 (0.0725 / 0.3624)	20 / 67	70.00% vs. 73.13%	0.857 (0.7815 / 1.0000)
Quantitative benchmark contested (D3)	11 / 39	0.4740 vs. 0.2262	+0.2478	291.0 (0.0510 / 0.3062)	15 / 72	73.33% vs. 72.22%	1.058 (1.0000 / 1.0000)
Unbenchmarked audit zeroing (D1)	17 / 33	0.2203 vs. 0.3118	-0.0916	257.0 (0.6056 / 1.0000)	24 / 63	79.17% vs. 69.84%	1.641 (0.4348 / 1.0000)
Audit preclusion / bright-line (D2)	10 / 40	0.4135 vs. 0.2475	+0.1660	228.0 (0.4647 / 1.0000)	16 / 71	75.00% vs. 71.83%	1.176 (1.0000 / 1.0000)

Note. ΔSAR < 0 reflects stronger pro-taxpayer erosion; ΔSAR > 0 reflects stronger DJP sustainment.

H2 is partly supported. Primary transactional substantiation (E2 + E3) is positively associated with overturn and negatively with SAR in every specification, whereas standalone transfer-pricing studies show no advantage. The associations are borderline (p = 0.0699 to 0.0820) except in the chamber-clustered model (p = 0.0213), and no bivariate contrast survives Holm adjustment (Table 7). Tendering an executed intercompany agreement (E2 = 1) cuts Stratum I mean SAR by more than two-thirds, from 0.3529 to 0.1123 (ΔSAR = -0.2406), and it lifts pooled overturn from 67.65% to 89.47% (Fisher OR = 4.065, p = 0.0820). In Model 1A (Table 8, Papke–Wooldridge Fractional Logit, n_1 = 50), each unit of the primary substantiation score reduces expected SAR by -13.66 percentage points (AME = -0.1366, β = -0.7697, CR1 z = -1.813, p = 0.0699; Pure TP-Item AME = -0.1177). Model 2 (N = 87) controls explicitly for the Stratum II cohort fixed effect. There, each unit of primary substantiation was associated with 2.05-fold odds of Tax Court overturn (Firth OR = 2.052, 95% CI 0.936–4.502, β = +0.7190, p = 0.0728; chamber-clustered CR1 β = +0.8270, SE = 0.3592, p = 0.0213, OR = 2.286, AME = +14.81 percentage points). The estimates are similar in unadjusted pooled estimation (OR = 2.248, p = 0.0368) and within Stratum I alone (OR = 2.002, Sensitivity S6).

Why, by contrast, did standalone Transfer Pricing Studies (E1 = 1) fail to protect taxpayers (Model 2 Firth OR = 0.507, p = 0.3022; Model 1 AME = +0.1574, p = 0.0718)? The *tempus regit actum* context of our 2003–2015 audited tax years offers an institutional explanation. Before PMK No. 213/PMK.03/2016, which mandated ex-ante Master/Local Files certified within four months of year-end from fiscal year 2016 onward, transfer-pricing studies tendered in 2003–2015 disputes were mostly ex-post defense reports commissioned after audit initiation, and Tax Court chambers are governed by *asas kebenaran materiil* (Articles 69, 76, and 78 Law No. 14/2002). They discount ex-post consultant benchmarking (TP16 PUT-80425). The discount lifts only when the benchmarking is anchored to executed commercial contracts and primary *Uji Bukti* deliverables (E2 + E3). Examination practice under PMK No. 172/2023 is reported to remain *ex post* in orientation despite the ex-ante documentation mandate[36].

Table 8. Aligned multivariate econometric estimates of Stratum I SAR (Model 1, n_1 = 50) and Stratum-Controlled Pooled Overturn (Model 2, N = 87).

Explanatory variable	Model 1A: Fractional Logit on SAR (n_1 = 50, G = 13) β (CR1 SE)	Model 1A AME (p-val)	Model 1B: CR1 OLS on SAR β (CR1 SE, p-val)	Model 2A: Firth Logit on Overturn (N = 87) β (SE)	Model 2A Firth OR [95% CI] (p-val)	Model 2B: CR1 Logit (G = 16) β (CR1 SE, p-val)
Intercept	-0.9809 (0.7911)	— (β p = 0.2150)	+0.2870 (0.1572, 0.0928)	+1.0036 (0.5039)	2.728 [1.016, 7.324] (0.0464)	+1.0593 (0.6030, 0.0790)
Primary Substantiation (E2 + E3, 0–2)	-0.7697 (0.4246)	-0.1366 (0.0699)	-0.1357 (0.0858, 0.1396)	+0.7190 (0.4008)	2.052 [0.936, 4.502] (0.0728)	+0.8270 (0.3592, 0.0213)
Standalone TP Study / Doc (E1, 0/1)	+0.8866 (0.4924)	+0.1574 (0.0718)	+0.1676 (0.1010, 0.1231)	-0.6784 (0.6576)	0.507 [0.140, 1.841] (0.3022)	-0.7327 (0.6299, 0.2447)
Quantitative Benchmark Contested (D3)	+0.9429 (1.0516)	+0.1674 (0.3699)	+0.1944 (0.2149, 0.3837)	-0.0764 (0.6804)	0.926 [0.244, 3.516] (0.9105)	-0.0422 (0.7916, 0.9575)
Intra-Group Services Archetype (IGS)	+0.6783 (0.7999)	+0.1204 (0.3964)	+0.1152 (0.1663, 0.5015)	-0.9898 (0.6510)	0.372 [0.104, 1.331] (0.1284)	-1.0697 (0.5252, 0.0417)
$\log_{10}$(Disputed IDR) (mean-centered)	-0.0042 (0.4999)	-0.0007 (0.9933)	+0.0029 (0.0828, 0.9730)	+0.0284 (0.3455)	1.029 [0.523, 2.025] (0.9346)	+0.0087 (0.4140, 0.9832)
Stratum II (MA_PK) Cohort Fixed Effect	—	—	—	-0.3972 (0.5585)	0.672 [0.225, 2.009] (0.4769)	-0.4159 (0.4851, 0.3912)
Model Diagnostics	QLL = -26.677	max VIF = 1.28	R^2 = 0.1359	ℓ_Firth = -46.784	max VIF = 1.24	AME(Primary) = +0.1481

Note. Identical five substantive covariates are aligned across Model 1 and Model 2, with Model 2 additionally controlling for the Stratum II cohort fixed effect.

H3 is not supported after adjustment. DJP's sustained adjustment ratio rose from 0.2262 to 0.4740 (Mann–Whitney p = 0.0510; Holm-adjusted p = 0.3062) where the record shows contested comparable screening (D3 = 1), compared with unbenchmarked zeroing (D1 = 1, SAR = 0.2203). The adjusted difference is small and imprecise (Model 1A AME = +16.74 percentage points, p = 0.3699), and pooled overturn does not differ (73.33% versus 72.22%, OR = 1.058, p = 1.0000). The ratio reaches 0.4135 where auditors invoke Article 26A(4) KUP audit-withholding preclusion (TP31 PUT-81715) or statutory loan rules (D2 = 1; n = 10; ΔSAR = +0.1660, Mann–Whitney p = 0.4647). Intra-Group Services (IGS) were associated with lower overturn odds in the chamber-clustered model (CR1 β = -1.0697, p = 0.0417, AME = -19.15 percentage points), but not in the Firth model (OR = 0.372, 95% CI 0.104–1.331, p = 0.1284) or in the pooled Fisher test (OR = 0.375, p = 0.1745). The adjusted SAR difference for IGS is not significant (Model 1A AME = +12.04 percentage points, p = 0.3964). Monetary scale (log_{10} disputed adjustment) shows no association with overturn (OR = 1.029, p = 0.9346).

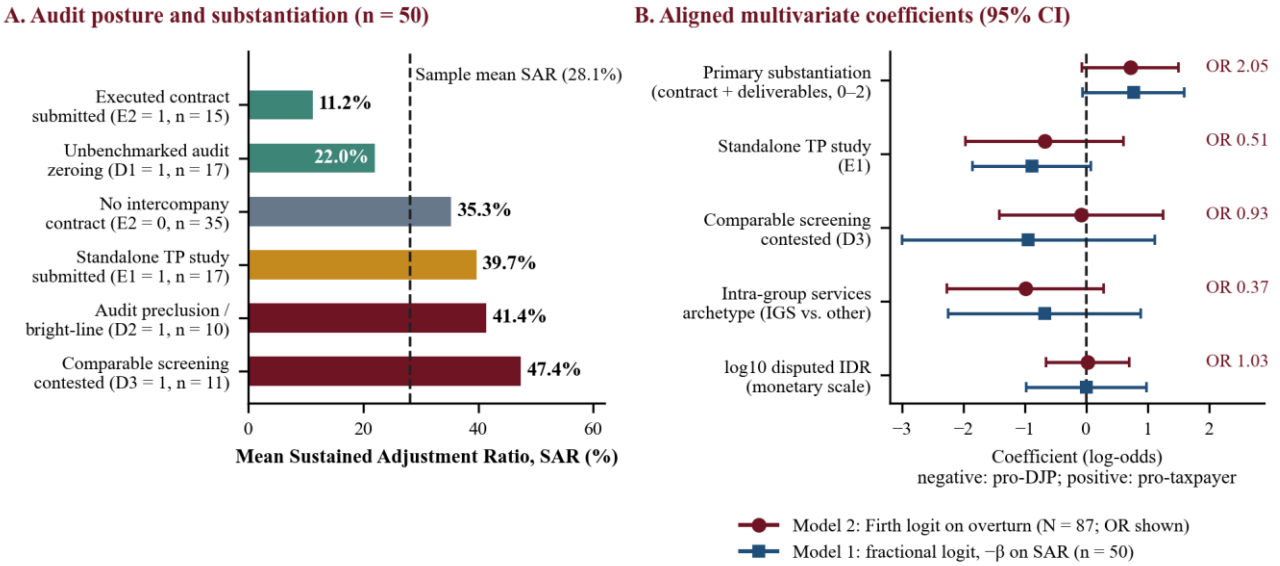

Figure 2. Upstream audit methodology, primary transactional substantiation, and aligned multivariate estimates. (A) Mean Sustained Adjustment Ratio (SAR, Stratum I, n_1 = 50) across documentary and audit-posture regimes. (B) Aligned five-predictor coefficient forest plot comparing Stratum-FE controlled Model 2 (Firth Logit on Pooled Overturn, N = 87) and inverted Model 1 (Papke–Wooldridge Fractional Logit -β on SAR, n_1 = 50, CR1 95% CIs; Max VIF = 1.24).

3.4 Supreme Court civil review and standing

Table 9 and Figure 3 examine Pareto fiscal concentration (*Gini* = 0.6681, where the top 15% of dockets hold 62.3% of the Rp 4.085 trillion pool) alongside Supreme Court *Peninjauan Kembali* outcomes (Stratum II, n_2 = 37), which is consistent with H4. Overall, 89.19% (33 / 37; 95% CI: 74.58%–96.97%) of PK petitions are rejected (*Menolak PK*), affirming the Tax Court symmetrically across DJP petitions (90.91% [20 / 22] rejected) and taxpayer petitions (86.67% [13 / 15] rejected; Fisher's exact OR = 1.538, p = 1.0000). In 97.30% (36 / 37) of PK rulings, the Supreme Court invoked Article 91(e) of Law No. 14/2002 (paired in 54.05% with substance over form), treating comparability and benefit appraisals as factual evaluations that fall outside the narrow grounds of Article 91[37].

Table 9. Tier II Supreme Court *Peninjauan Kembali* (n_2 = 37) lock-in analysis and expanded nine-model sensitivity battery (S1–S9).

Panel / Analytical dimension	Subgroup / Sample (n)	Primary metric / Estimate	95% CI / Counterpart	Substantive & methodological conclusion
Panel A: Supreme Court PK Merits (Stratum II, n_2 = 37)				
Overall PK Affirmance (*Menolak PK*)	37	33 / 37 (89.19%)	[74.58%, 96.97%]	Tax Court is de facto court of last resort in ~9 of 10 TP cases
DJP as PK Petitioner (59.46% of PK docket)	22	20 / 22 Rejected (90.91%)	[70.84%, 98.88%]	Only 2 DJP wins: TP55 (2198/B/PK/Pjk/2020) & TP61 (426/B/PK/Pjk/2020)
Taxpayer as PK Petitioner (40.54% of PK docket)	15	13 / 15 Rejected (86.67%)	[59.54%, 98.34%]	Only 2 Taxpayer wins: TP75 (282/B/PK/Pjk/2018) & TP76 (283/B/PK/Pjk/2018)
Petitioner Symmetry & Art. 91(e) Citation	22 vs. 15	Fisher OR = 1.538 (p=1.0)	Art. 91(e): 97.30%	Symmetrical lock-in despite asymmetric DJP filing volume
Panel B: Expanded Sensitivity & Robustness Battery (S1–S9)				
S1 Corporate Income Tax (PPh Badan) dockets only	N=70 / n_1=37	Overturn 71.43%	Mean SAR = 0.2964	Removing PPh 23/26 withholding dockets leaves baseline intact
S2 IDR-denominated bookkeeping dockets only	N=50 / n_1=29	Overturn 74.00%	Mean SAR = 0.3116	Excluding USD functional-currency dockets preserves >70% erosion
S3 Post-PER-22/PJ/2013 decision cohort (2015–2020)	N=65 / n_1=29	Overturn 72.31%	Mean SAR = 0.2722	Overturn stays above 70% in the post-PER-22 decision cohort

Panel / Analytical dimension	Subgroup / Sample (n)	Primary metric / Estimate	95% CI / Counterpart	Substantive & methodological conclusion
S4 Leave-One-Chamber-Out (LOCO, 16 chambers)	16 iterations	Overturn 69.62%–76.92%	SAR: 0.2098–0.3253	No single *Majelis* drives trial overturn or SAR
S5 Trimmed 5%–95% monetary-outlier sample	N = 77	Overturn 72.73%	vs. 72.41% baseline	Mega-dockets (>Rp 290B) and micro-dockets do not distort rates
S6 Stratum-FE vs. Unadjusted vs. Stratum I-Only OR	N=87 / n_1=50	OR = 2.052 (FE)	2.248 (Unadj) / 2.002 (S1)	Primary substantiation OR ~ 2.0–2.25 is similar across strata
S7 Constant exchange rate (1 USD = Rp 13,500)	n_1 = 50	Value-Weighted SAR = 0.2611	vs. 0.2506 BI FX	~74%–75% Rupiah fiscal erosion is invariant to FX assumptions
S8 Pure Transfer-Pricing Item SAR (tp_specific_sar)	n_1 = 50	Mean 0.2916; VW 0.2552	AME(Primary) = -0.1177	Isolating pure TP items in Partial Splits confirms 74.5% erosion
S9 De-duplicated Companion-Cluster Corpus	N=78 / n_1=45	Overturn 73.08%	Mean SAR = 0.3119	Dropping 9 sibling CIT–WHT / multi-year dockets preserves results

DJP filed 22 of 37 PK petitions (59.46%), and 20 of them (90.91%) were rejected. A petition is filed by the party that lost at trial, so this share follows from the 22 PK dockets in which the taxpayer had won. The tension with Decision No. 24/PUU-XXII/2024 is doctrinal, not an observed asymmetry of behavior. In general administrative litigation (*Peradilan Tata Usaha Negara*), Constitutional Court Decision No. 24/PUU-XXII/2024 barred government officials from filing PK petitions, in order to protect citizens' final judicial victories against protracted state relitigation[38]. In tax disputes, DJP still files PK. It relies on the lex specializ phrase "*pihak-pihak yang bersengketa*" in Article 77(3) of Law No. 14/2002 (with Article 89(1)). It succeeded in only two of 22 petitions (TP55 and TP61), and TP55 turned on a multi-year quantitative benchmark: in 2198/B/PK/Pjk/2020 (TP55), DJP proved that the taxpayer's Return on Sales (ROS) exceeded 5.14% across four consecutive prior years (2008–2011). That justified a USD 4,358,830 five-year quartile adjustment:

> *"bahwa dalam pengujian kewajaran transaksi hubungan istimewa dan penghitungan rentang laba wajar yang mendasarkan pada Quartile data pembanding tahun 2012 sebesar 5,14% sudah tepat ... didukung berdasarkan data pembanding selama 5 (lima) tahun dapat dilihat bahwa Return on Sales (ROS) Pemohon Banding selama 4 tahun sebelumnya, yaitu dari tahun 2008 s.d. 2011 selalu berada di atas 5,14%." (2198/B/PK/Pjk/2020, TP55; contrast taxpayer PK reversals TP75 282/B/PK/Pjk/2018 and TP76 283/B/PK/Pjk/2018 where M.XIV had ignored verified cross-border performance proof).*

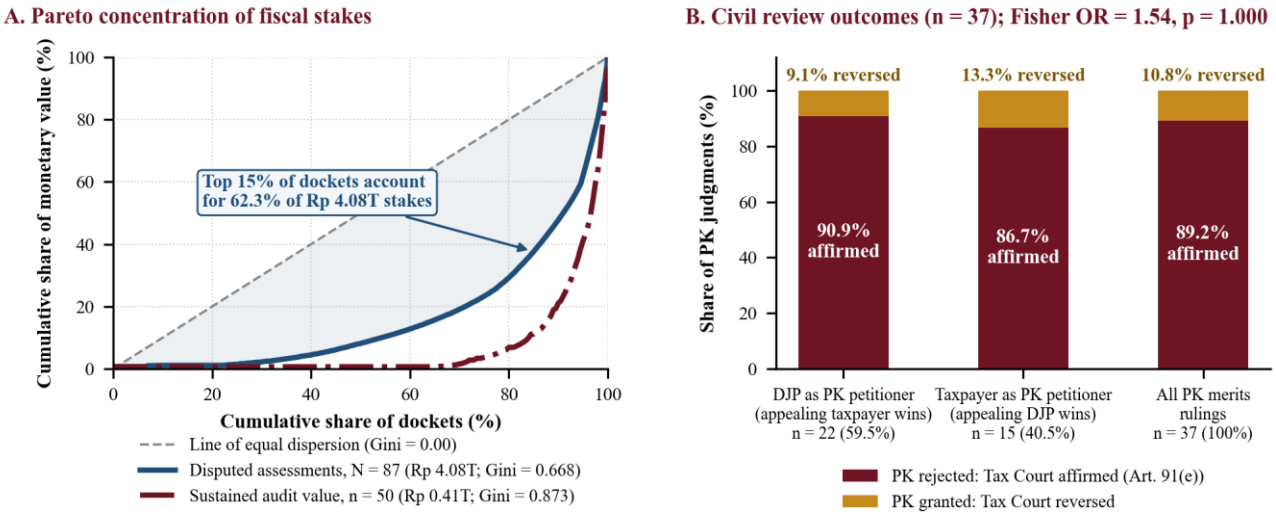

Figure 3. Pareto concentration of disputed fiscal stakes and Supreme Court civil review lock-in. (A) Lorenz curves of pooled disputed transfer-pricing assessments (N = 87, Rp 4.08T, *Gini* = 0.668) and Stratum I sustained audit value (n_1 = 50, Rp 0.41T, *Gini* = 0.873). (B) Symmetrical Supreme Court PK affirmance rates across DJP (90.9%) and taxpayer (86.7%) petitions (n_2 = 37).

4. Discussion

4.1 Four-canon doctrinal diagnosis of PMK No. 172/2023

The four civil-law canons, applied in turn, support an argument that PMK No. 172 of 2023 may intensify the audit–appeal chasm documented in our 2011–2020 judicial baseline rather than cure it. The argument is predictive: the audited years precede the regulation.

1. Grammatical Interpretation (Interpretasi Gramatikal): Article 18(3) UU PPh authorizes DJP "*untuk menentukan kembali besarnya penghasilan dan pengurangan*" ("to redetermine the amount of income and deductions") according to arm's-length business custom. Read grammatically, "*menentukan kembali besarnya*" mandates re-pricing to an arm's-length benchmark. It does not permit punitive confiscation to Rp 0[3,39]. Article 13(1)–(2) of PMK No. 172/2023, however, carries forward the cumulative eight-limb preliminary test (*tahapan pendahuluan*) of PER-32/2011 and PER-22/2013, with no safe harbor or proportional fallback, and this may encourage auditors to disallow 100% of genuine commercial royalties or services (SAR target = 1.0) over minor documentation gaps. Chambers strike such disallowances down under Article 6(1)(a) UU PPh (the "3M" test, practitioner shorthand for the wording of Article 6(1): expenses incurred to earn, collect, and maintain income).

2. Systematic Interpretation (Interpretasi Sistematis) & Secondary Dividend Escalation (Article 37 PMK 172/2023): Read systematically, two inter-statutory collisions compound litigation risk. The first pits Article 26A(4) UU KUP (audit-withholding preclusion) against Articles 69 and 76 of Law No. 14/2002. Those empower Tax Court judges to admit *Uji Bukti* primary deliverables *ex officio* for *kebenaran materiil*[5,6]. The second follows from new rules. Article 36(6) of PP No. 55/2022 and Article 37 of PMK No. 172/2023 codify mandatory secondary adjustments, and these deem every primary transfer-pricing adjustment a constructive dividend (*dividen terselubung*) subject to PPh Pasal 23/26 withholding tax, even toward non-shareholder affiliates, unless cash is repatriated or the matter is resolved through the mutual agreement procedure (MAP). Disputes over this secondary adjustment have already reached the Tax Court[40]. An unbenchmarked Rp 0 primary disallowance, which was overturned in 79.17% (19/24) of the pooled zeroing dockets, may now also generate secondary withholding assessments under Article 37. This effect is predicted from the text of the regulation and is not observed in the corpus.

3. Teleological Interpretation (Interpretasi Teleologis): The Elucidation to Article 29(2) paragraph 3 UU KUP requires assessments to rest on competent objective proof, and disallowances in Stratum I lost 74.94% of their Rupiah value (Rp 1.226T cancelled), which may undermine revenue certainty and raises a question under Article 28D(1) UUD 1945 that the present data cannot resolve. The contemporaneous Local File mandate of PMK No. 213/2016 and PMK No. 172/2023 likewise fails teleologically. That holds unless it is paired with primary contracts and operational deliverables (E2 + E3, OR = 2.052).

4. Historical Interpretation (Interpretasi Historis): The Tax Court grew out of the *Majelis Pertimbangan Pajak*. Law No. 14/2002 then placed it under the shared oversight of the Ministry of Finance and the Supreme Court. For a long time, observers suspected it of leaning toward the executive[7,8,41,42]. Our jurimetric evidence (78.00% trial overturn) does not support that suspicion for the dockets studied; it does not test judicial independence. It also points to an institutional gap ahead of the 31 December 2026 transfer deadline set by Constitutional Court Decision No. 26/PUU-XXI/2023: in the authors' assessment, the Supreme Court has no unified evidentiary guidelines for transfer pricing[8,9].

4.2 Comparative functional analysis

Courts in some jurisdictions have rejected tax authorities' transfer-pricing or royalty positions, as in a 2025 judgment of the Brussels Court of First Instance on a royalty deduction[39], an earlier Belgian Court of Appeal ruling that tested a DEMPE analysis[43], German Federal Tax Court rulings on intercompany financing[44], and the High Court of Australia's PepsiCo decision on embedded royalties ([2025] HCA 30, decided 13 August 2025)[45]. Figure 4 sets out the contrast. Four functional comparators show institutional safeguards that Indonesia currently lacks, in the authors' assessment:

- **OECD Guidelines (2022) & Singapore (IRAS 7th Ed. 2024), Elective 5% LVAS Safe Harbors:** Chapter VII (paras. 7.43–7.65) of the 2022 OECD Guidelines and Singapore's IRAS Transfer Pricing Guidelines (7th ed. 2024, operationalizing sections 34D and 34F of the Income Tax Act 1947) establish an elective 5% cost-plus safe harbor for routine low value-adding intra-group support services (LVAS)[46,47]. They verify the cost pool. Testing the benefit to each recipient individually is not required. Bilateral treaty (MAP) compatibility is preserved (OECD TPG paras. 4.95–4.133). PMK No. 172/2023 imported OECD's strict Article 13 benefit gateway but left out the 5% LVAS safe harbor.

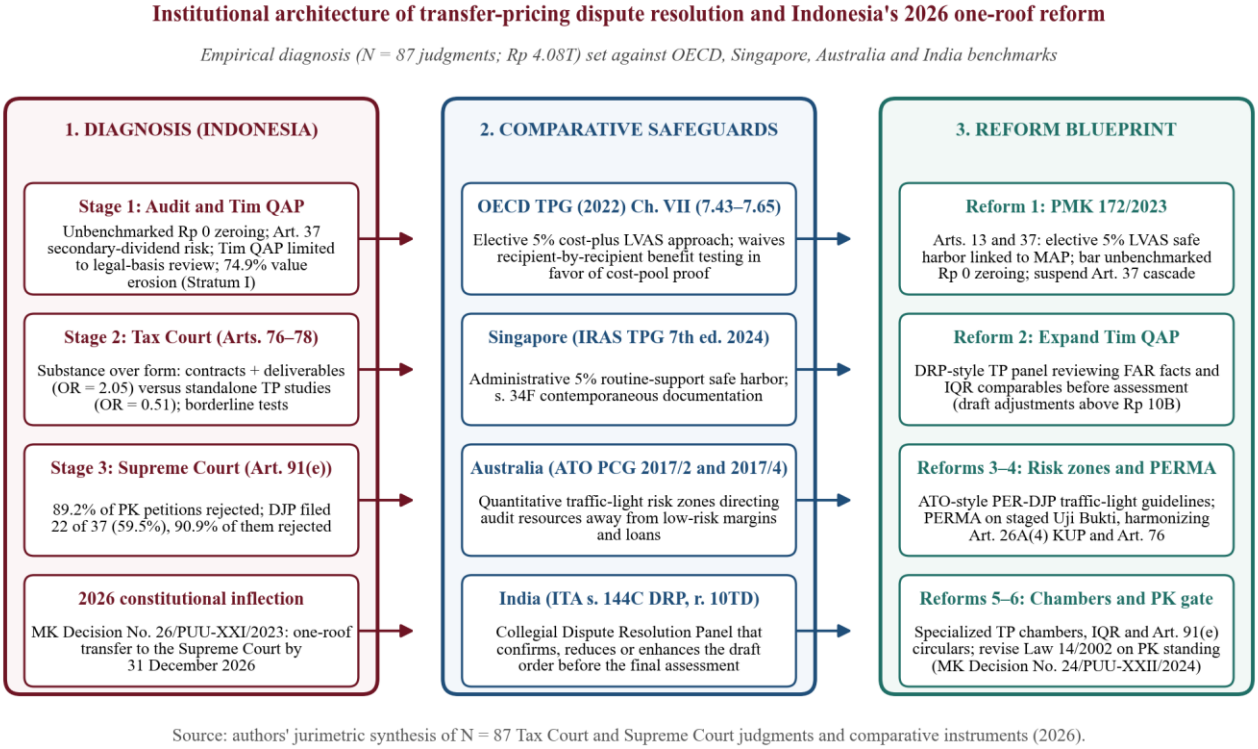

Figure 4. Comparative institutional architecture of transfer-pricing dispute resolution and Indonesia's sequenced 2026 reform blueprint (Reforms 3–4 and 5–6 of Table 10 are combined in the third column).

- **India (ITA s. 144C DRP & Rules 10TD–10TG) vs. Indonesia's Restricted Tim QAP:** Alongside sector safe harbors[48,49], India created collegial Section 144C Dispute Resolution Panels (DRP) with authority to confirm, reduce, or enhance the variations proposed in a draft Transfer Pricing Officer order before it becomes final[50,51]. Indonesia already has a pre-assessment collegial body: DJP's *Tim Quality Assurance Pemeriksaan* (Tim QAP, defined in PMK No.

184/PMK.03/2015 and now in Articles 1(37) and 19 of PMK No. 15/2025). Yet trial overturn reached 78% in these dockets. Tim QAP review is limited to the legal basis of corrections not yet agreed at the final discussion (*dasar hukum koreksi*), which bars it from reviewing factual FAR evidence or IQR comparable selection; whether the 50 audits passed through Tim QAP is not observed.

• **Australia (ATO PCG 2017/2 & PCG 2017/4), Quantitative Risk-Zone Triage:** Australia's ATO publishes quantitative "traffic-light" compliance risk zones for routine distribution/service margins and related-party debt pricing[52,53]. It concentrates audit litigation on high-risk structures. Capacity is not spread over scale-invariant disputes (*Gini* = 0.668).

4.3 De lege ferenda: six sequenced reforms

We propose six sequenced *de lege ferenda* reforms (Table 10). They come ahead of the *satu atap* integration due on 31 December 2026 under Constitutional Court Decision No. 26/PUU-XXI/2023:

1. Insert an Elective 5% LVAS Safe Harbor, Anti-Zeroing Rule, and Secondary-Dividend Suspension into PMK No. 172/2023 (Minister of Finance, Immediate 0–6 months): Article 13 of PMK No. 172/2023 should be amended to adopt an elective, treaty-compatible 5% cost-plus safe harbor for routine intra-group support services (≤ Rp 15 billion/year), integrated with the mutual agreement procedure (Articles 41–54) and advance pricing agreement (APA, Articles 55–71) tracks . The Rp 15 billion and Rp 10 billion thresholds in Reforms 1 and 2 are illustrative and should be calibrated on the distribution of disputed adjustments (median Rp 19.7 billion in this corpus). Unbenchmarked Rp 0 disallowances should be prohibited. This applies whenever executed contracts and primary deliverables (E2 + E3) prove that the services were actually rendered commercially. Article 37 should also be amended. The change would suspend secondary constructive-dividend withholding where primary adjustments are under MAP or a bona fide comparability dispute.

2. Expand DJP's Existing Tim Quality Assurance Pemeriksaan (Tim QAP) into a Full-Fact Transfer-Pricing Gate (Ministry of Finance / DJP, Immediate 0–6 months): Tim QAP regulations should be amended. No new bureaucracy is needed. For all draft Article 18(3) adjustments exceeding Rp 10 billion, a specialized collegial QAP–TP panel would then exercise a review function modeled on India's s. 144C DRP, extended to factual FAR substantiation and multi-year IQR comparable screening. This replicates the methodology that succeeded in TP55 2198/B/PK/Pjk/2020. The aim is to move the sustained SAR from 0.220, observed under unbenchmarked zeroing, toward 0.474, observed where comparable screening was contested (an unadjusted difference).

3. Publish ATO-Style PER-DJP Quantitative Risk-Zone Guidelines (DJP, Short-term 6–12 months): Issue transparent "traffic-light" compliance risk matrices modeled on PCG 2017/2 and PCG 2017/4 for routine manufacturing/distribution margins and post-PMK 169/2015 affiliate financing. They would bind auditors administratively under the general principles of good governance (AAUPB). Audits would then focus on high-risk profit shifting.

4. Promulgate a Post-2026 Supreme Court Regulation (PERMA) Harmonizing Article 26A(4) KUP and Article 76 Law No. 14/2002 (Supreme Court, Short-term 6–12 months): The Supreme Court assumes one-roof governance on 31 December 2026. It should enact a *PERMA Acara Sengketa Pajak*. The PERMA would distinguish wilful withholding of core domestic ledgers (precluded under Article 26A(4) KUP) from supplemental foreign-affiliate performance exhibits, which are admissible in structured pre-trial *Uji Bukti* conferences. Local Files (E1) would also annex primary contract and deliverable schedules (E2 + E3).

5. Designate Specialized Transfer-Pricing Chambers and Issue Unified MA–PP IQR Guidelines (Supreme Court, Medium-term 12–24 months): Dedicated international-tax *Majelis* should be established. They belong within the integrated Tax Court. Supreme Court Administrative Chamber circulars (*SEMA Kamar TUN*) should then standardize multi-year IQR adjustments and Article 91(e) manifest-error thresholds, which would narrow LOCO chamber variance of 69.62%–76.92%.

6. Revise Law No. 14/2002 to Codify One-Roof Governance and Curb State-Initiated PK Filings (DPR RI & President, Structural 24–36 months): Articles 77(3) and 91 of Law No. 14/2002 should be harmonized with the constitutional spirit of Constitutional Court Decision No. 24/PUU-XXII/2024. DJP PK petitions would be restricted. Only authentic *novum* (Article 91(b)) or certified inter-chamber jurisprudential conflict would qualify. Today these petitions make up 59.46% of filings, with a 90.91% rejection rate.

Table 10. De lege ferenda: six sequenced regulatory, administrative, and post-2026 judicial reforms.

#	Reform proposal	Empirical defect targeted (N = 87)	Comparative / Constitutional anchor	Legal instrument & authority	Horizon
1	Elective 5% LVAS safe harbor (≤ Rp 15B), anti-Rp 0 zeroing & Art. 37 secondary-dividend suspension	53.85% IGS overturn (CR1 p = 0.0417; Firth p = 0.1284); 79.17% zeroing overturn (SAR = 0.220)	OECD TPG 7.43–7.65; Singapore IRAS 5% rule	Amendment to PMK No. 172/2023 Arts. 13 & 37 (Minister of Finance)	Immediate (0–6 mos)
2	Upgrade existing DJP Tim QAP to review FAR facts & multi-year IQR comparables (> Rp 10B)	74.94% (Rp 1.23T) value-weighted erosion vs. 0.474 benchmarked SAR; Tim QAP legal-only bar	India Income-tax Act s. 144C DRP	Amendment to MoF Audit QAP Regulation (Ministry of Finance / DJP)	Immediate (0–6 mos)
3	Publish quantitative "Traffic-Light" TP & financing risk-zone guidelines	Scale-invariant losses (OR = 1.029) across *Gini* = 0.668 (Rp 4.085T) stakes	Australia ATO PCG 2017/2 & PCG 2017/4	DJP Director-General Regulation (PER-DJP)	Short (6–12 mos)
4	Enact staged *Uji Bukti* admissibility & primary-exhibit annexure rule	Collision of Art. 26A(4) KUP vs. Art. 76 Law 14/2002; standalone E1 no advantage (OR = 0.507)	Singapore s. 34F contemporaneous substantiation	Supreme Court Regulation (PERMA Acara Pajak) post-31 Dec 2026	Short (6–12 mos)
5	Designate specialized TP chambers & issue MA–PP IQR / Art. 91(e) circulars	Chamber variance (LOCO 69.6%–76.9%) & 2–1 panel splits (TP36)	Specialized tax benches & unified chamber circulars	Supreme Court Chief Justice Decree (SK KMA) & *SEMA Kamar TUN*	Medium (12–24 mos)

#	Reform proposal	Empirical defect targeted (N = 87)	Comparative / Constitutional anchor	Legal instrument & authority	Horizon
6	Revise Law No. 14/2002 for post-26/PUU-XXI/2023 one-roof & restrict DJP PK standing	DJP filed 22 of 37 PK petitions (59.46%); 20 (90.91%) were rejected	MK Decisions 26/PUU-XXI/2023 & 24/PUU-XXII/2024	Statutory revision of Law No. 14/2002 (DPR RI & President)	Structural (24–36 mos)

Note. Reforms 1–5 require no new statute and can be adopted through Ministerial, DJP, and Supreme Court rulemaking (the Supreme Court measures after December 2026).

4.4 Limitations

Eight boundaries qualify our inferences. First, tax audits and objections are confidential under Article 34 UU KUP. Our N = 87 corpus therefore observes only the adjudicated tip of the dispute pyramid (tax years 2003–2015, decisions 2011–2020), so it establishes the pre-reform baseline, before post-2024 audits under PMK No. 172/2023 reach final judgment (2027–2030). Second, Stratum II PK transcripts condense trial pleadings (0.324 vs. 0.880 mean substantiation mentions). Adding the Stratum II cohort fixed effect (Model 2) and testing Stratum I separately (Sensitivity S6: OR = 2.052 vs. 2.002) confirm that this cross-stratum condensation does not drive our findings, and isolating pure transfer-pricing items (S8, pure TP-item SAR = 0.2552) and removing 9 companion sibling dockets (S9, N = 78, 73.08% overturn) show that the results also hold against multi-issue and serial-audit clustering. Third, moderate bivariate contrasts do not survive conservative six-test Holm correction (Table 7), given $n_1 = 50$ and N = 87. That reinforces our abductive framing. Fourth, our case-by-case manual audit (log available on request) achieved 72.73%–100.00% agreement ($\kappa = 0.634$–1.000). One legal scholar did the human verification, though. Future post-2026 one-roof audits should deploy multi-coder panels across $N \geq 250$ dockets. Fifth, the benchmark variable D3 records contested comparable screening and does not identify who constructed the benchmark. Sixth, Stratum II is selected on the trial outcome, because a PK petition is filed by the party that lost at trial. Seventh, cluster-robust inference in Stratum I rests on 13 clusters. Eighth, USD dockets were converted to Rupiah for the monetary covariate at Bank Indonesia exchange rates (Sensitivity S7 tests a constant rate of Rp 13,500 per USD).

5. Conclusion

The corpus covers eighty-seven non-overlapping Indonesian income-tax substantive transfer-pricing judgments, representing Rp 4.085 trillion in disputed adjustments. In these 87 judgments, DJP's transfer-pricing assessments lost most of their value at trial and were rarely restored at civil review. In Stratum I Tax Court appeals ($n_1 = 50$), 78.00% of assessments are overturned in full (54.00%) or in part (24.00%), while the Rupiah value-weighted SAR is 0.2506, which extinguishes 74.94% (Rp 1.226 trillion) of audited stakes (72.41% overturn across N = 87; Table 5). Once stratum fixed effects are controlled, primary transactional substantiation, meaning executed intercompany contracts plus verifiable deliverables, was associated with higher overturn odds (Firth OR = 2.052, 95% CI 0.936–4.502; CR1 p = 0.0213). Pre-2016 standalone transfer-pricing studies showed no advantage (OR = 0.507). Sustained ratios were higher where comparable screening was contested (0.4740 vs. 0.2262, unadjusted p = 0.0510), but the difference does not survive adjustment. At *Peninjauan Kembali* ($n_2 = 37$), the Supreme Court affirms 89.19% of trial rulings (90.91% DJP defeat) under Article 91(e). Indonesia is now executing the 31 December 2026 one-roof transfer of the Tax Court to the Supreme Court, as ordered by Constitutional Court Decision No. 26/PUU-XXI/2023, so three measures should be paired: elective 5% LVAS safe harbors, expanded factual Tim QAP pre-assessment review, and unified Supreme Court evidentiary rules. Together, these measures would help reconcile fiscal protection with rule-of-law certainty.

Declarations

Ethics Approval and Consent

Not applicable. The study analyzed promulgated statutes, regulations, and official public court judgments and involved no human participants, so institutional ethics review was not required.

Use of Artificial Intelligence

The authors declare that no generative artificial intelligence tools were used in the design, analysis, or writing of this manuscript.

Author Contributions

A.G.: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. H.Y.: Formal analysis, Software, Validation, Visualization, Project administration, Writing – review & editing. H.P.P.: Supervision, Writing – review & editing. All authors reviewed and approved the final version of the manuscript.

Funding

None. This research received no external funding.

Competing Interests

None. The authors declare no competing interests.

Data Availability Statement

The de-identified dataset and the analysis scripts supporting the findings are available from the corresponding author on reasonable request.

Acknowledgments

None.

References

1. Collier R, Andrus JL. Transfer pricing and the arm's length principle after BEPS. Oxford: Oxford University Press; 2017. doi:10.1093/law/9780198802914.001.0001
2. Tambunan MRUD. How is the transfer pricing concept arm's length principle applied in Indonesia? Bisnis Birokrasi. 2022;29(3). doi:10.20476/jbb.v29i3.1317
3. Tanisha G, Setiabudi AW. Studi komparatif penerapan *prinsip kewajaran dan kelaziman usaha* pada transaksi afiliasi jasa manajemen sebelum dan sesudah berlakunya PMK Nomor 172 Tahun 2023. Balance J Akunt Audit Keuang. 2024;21(2):145–176. doi:10.25170/balance.v21i2.6450
4. Winarta R, Basani CS. Kepastian hukum dalam penggunaan *prinsip kewajaran dan kelaziman usaha* dikaitkan dengan kepatuhan hukum wajib pajak. J Ilmu Multidisiplin. 2026;5(1):554–568. doi:10.38035/jim.v5i1.1894
5. Hariyasin H. Kekuatan eksekutorial putusan *Majelis Pengadilan Pajak* terhadap sengketa pajak di Indonesia. Mimbar Keadilan. 2023;16(2):269–276. doi:10.30996/mk.v16i2.8909
6. Erwiningsih W. Implementasi penyelesaian sengketa pajak pada *Pengadilan Pajak* Indonesia. Justicia Sains J Ilmu Huk. 2022;6(2):248–262. doi:10.24967/jcs.v6i2.1585
7. Rifandanu F. Implikasi Putusan *Mahkamah Konstitusi* Nomor 26/PUU-XXI/2023 terhadap pembinaan *Pengadilan Pajak*. Amnesti J Huk. 2024;6(1):145–161. doi:10.37729/amnesti.v6i1.3270
8. Dewi NS, Rido A. Tinjauan yuridis kedudukan *Pengadilan Pajak* pasca Putusan *Mahkamah Konstitusi* Nomor 26/PUU-XXI/2023. Amicus Curiae. 2025;2(2):738–745. doi:10.25105/amicus.v2i2.23011
9. Sa'adah N, Aulia R. Reposisi *Pengadilan Pajak* ke dalam sistem kekuasaan kehakiman: Rekonstruksi pasca Putusan *Mahkamah Konstitusi* Nomor 26/PUU-XXI/2023 melalui integrasi ke

Mahkamah Agung. Konferensi Nasional Asosiasi Pengajar Hukum Tata Negara dan Hukum Administrasi Negara. 2026;3(1):173–188. doi:10.55292/2v6qs709

10. Priest GL, Klein B. The selection of disputes for litigation. J Leg Stud. 1984;13(1):1–55. doi:10.1086/467732
11. Sansing R. Voluntary binding arbitration as an alternative to Tax Court litigation. Natl Tax J. 1997;50(2):279–296. doi:10.1086/ntj41789257
12. Gliksberg D. Does the law matter? Win rates and law reforms. J Empir Leg Stud. 2014;11(2):378–407. doi:10.1111/jels.12044
13. Martensen F. Rethinking interquartile ranges. Int Transf Pricing J. 2019;27(1):24–33. doi:10.59403/17st9xw
14. Reuther P, Brown SS, Krøier K, et al. The illusion of precision: Rethinking the use of the interquartile range in transfer pricing. Int Transf Pricing J. 2026;34(1):1–14. doi:10.59403/azvzw2
15. Pamungkas FWA, Putri NE, Hutagaol PMJL. Mitigasi volatilitas transfer pricing: Bukti empiris robustitas interquartile range melalui simulasi data ekstrem. JAPP J Akunt Perpajak Portofolio. 2026;6(1):8–15. doi:10.24269/japp.v6i1.14200
16. Hasan M, Damayanti C. Unwarranted or warranted? A jurimetric analysis of custodial-sentence disparity in Indonesian corruption cases after the 2020 Supreme Court Sentencing Guideline. Enigma Law. 2026;3(2):62–73. doi:10.61996/law.v3i2.125
17. Rahayu D, Badruddin A, Wang E. Void by law, alive in practice: Measuring prohibited standard clauses in twenty Indonesian digital-platform contracts under Article 18. Enigma Law. 2026;3(2):74–94. doi:10.61996/law.v3i2.126
18. Elsyra N, Syahwami, Sunaryo J, et al. Evaluating the effectiveness of legal frameworks in promoting integrity and preventing maladministration in public agencies in Jambi Province. Enigma Law. 2025;3(1):28–37. doi:10.61996/law.v3i1.87
19. Tambunan MRUD. A note to transfer pricing audit performance by Indonesia's tax administration based on Tax Court decision 2015–2019 on manufacturing industry. J Akunt Audit Indones. 2020;24(2):85–96. doi:10.20885/jaai.vol24.iss2.art2
20. Tambunan MRUD, Rosdiana H, Irianto ES. Transfer pricing cases on intangibles performed by manufacturing industry based on Tax Court decisions 2015–2019 in Indonesia. Int Transf Pricing J. 2020;27(5):361–371. doi:10.59403/3df99kh
21. Tambunan MRUD, Rosdiana H, Irianto ES. Transfer pricing on intra-group financing in the manufacturing industry in Indonesia: An essay on Tax Court decisions (2014–2019). Int Transf Pricing J. 2020;27(3):197–207. doi:10.59403/2wj94tb
22. Tambunan MRUD. Transfer pricing settlement in Indonesia: A note for tax authority, Tax Court, and taxpayers based on the Tax Court decisions. Bisnis Birokrasi. 2022;29(2):74–86. doi:10.20476/jbb.v29i2.1306
23. Aisya S, Nuryanah S. Tax dispute analysis of transfer pricing: case study of Tax Court decisions 2020–2023. JMKSP. 2024;9(1):481–495. doi:10.31851/jmksp.v9i1.14343
24. Michaels R. The functional method of comparative law. In: Reimann M, Zimmermann R, editors. The Oxford handbook of comparative law. 2nd ed. Oxford: Oxford University Press; 2019. p. 344–389. doi:10.1093/oxfordhb/9780198810230.013.11
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–174. doi:10.2307/2529310
26. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol. 1993;46(5):423–429. doi:10.1016/0895-4356(93)90018-v
27. Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J Appl Econom. 1996;11(6):619–632. doi:10.1002/(sici)1099-1255(199611)11:6<619::aid-jae418>3.0.co;2-1
28. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73(1):13–22. doi:10.1093/biomet/73.1.13
29. Cameron AC, Gelbach JB, Miller DL. Robust inference with multiway clustering. J Bus Econ Stat. 2011;29(2):238–249. doi:10.1198/jbes.2010.07136
30. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27–38. doi:10.1093/biomet/80.1.27
31. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002;21(16):2409–2419. doi:10.1002/sim.1047
32. Simamora BEPP, Hermawan AA. Transfer pricing analysis on intra-group services and the related transfer pricing disputes from Indonesian tax perspectives. In: Proceedings of the 6th International Accounting Conference (IAC 2017). Atlantis Press; 2018. doi:10.2991/iac-17.2018.5
33. Yeroshenko X. Supreme Administrative Court decision on intra-group services involving the Stock Spirit Group. Int Transf Pricing J. 2023;30(5):1–8. doi:10.59403/35va9bz
34. Zaki ARS, Irawan F. Intra-group services: suatu perspektif fiskus. J Pajak Indones. 2022;6(2):275–286. doi:10.31092/jpi.v6i2.1805
35. Suharsono A, Prasetyoningsih N, Nur Aida Ikrima. A comparison of Supreme Court justices' legal argumentation in deciding disputes on transfer pricing management fees in Indonesia. J Jurisprud. 2024;14(2):213–232. doi:10.23917/jurisprudence.v14i2.5874
36. Irawan F, Gumilang G. Ex ante and *ex post* approaches in transfer pricing documentation: a review of PMK 172 of 2023 and examination practices by the tax officers. Educoretax. 2025;5(5):652–659. doi:10.54957/educoretax.v5i5.1590
37. Gunawan AR, Mulyaningrum ER. Mekanisme peninjauan kembali atas putusan *Pengadilan Pajak* dalam upaya penyelesaian sengketa pajak. J Ilmu Huk Humaniora Polit. 2024;4(6):2662–2670. doi:10.38035/jihhp.v4i6.2849
38. Azzahra SS, Pratama AB. Penafsiran hukum pada Putusan *Mahkamah Konstitusi* Nomor 24/PUU-XXII/2024 terhadap kewenangan pengajuan peninjauan kembali dalam sengketa tata usaha negara. Manabia J Const Law. 2025;5(2):209–228. doi:10.28918/manabia.v5i02.13931
39. Engels A, Parduyns E. Belgian court rejects tax authorities' use of transfer pricing principles to deny deduction of royalty for trade name use. Int Transf Pricing J. 2025;33(1):1–9. doi:10.59403/33y9xwj
40. Santoso I, Ning Rahayu, Gunadi. Review of certainty aspect and revenue productivity principles in the secondary transfer pricing adjustment policy in Indonesia. Lex Localis. 2025;23(S5):2854–2866. doi:10.52152/801674
41. Wahjudin IH, Suwadi P, Rustamaji M. The existence of one roof system at Indonesian Tax Court. Lex Localis. 2025;23(10):1028–1042. doi:10.52152/801251
42. Regika V, Sari M, Muhkam MF, et al. The strengthening of Tax Court independence in the Indonesian judicial system. J Cakrawala Huk. 2024;15(1):80–90. doi:10.26905/idjch.v15i1.14276
43. Michel B. Belgian Court of Appeal wipes the floor with ambulatory use of 2017 transfer pricing guidelines, but not without testing the tax authorities' DEMPE function analysis. Int Transf Pricing J. 2023;30(2). doi:10.59403/xggxc8
44. Bärsch SE, Engelen C. German Federal Tax Court rules in favour of taxpayer in transfer pricing cases on intercompany financing transactions. Int Transf Pricing J. 2022;29(1). doi:10.59403/1n65wcb
45. Butler M, Wang L. High Court dismisses Commissioner's appeal in PepsiCo and confirms no royalty withholding tax or diverted profits tax payable on "embedded royalties". Int Transf Pricing J. 2026;33(2):1–12. doi:10.59403/359h4w3
46. Bose A, Alvarez U. Updated E-Tax Guide on Singapore Transfer Pricing Guidelines. Int Transf Pricing J. 2021;29(1):58–65. doi:10.59403/9apwmj
47. Benz A. Practical application of the guidance on low-value-adding services: case study. Int Transf Pricing J. 2021;28(2). doi:10.59403/10z2m50
48. Agrawal M. One step towards a non-adversarial tax regime: Safe harbour rules in India. Int Transf Pricing J. 2017;24(6):458–466. doi:10.59403/3zsy8fg
49. Patel VT, Savla KP, Punamiya A. India safe harbour intra-group loans per 1 April 2024. Int Transf Pricing J. 2024;31(5):1–7. doi:10.59403/2ya5skf
50. Sengupta DP. Judicial systems and international tax disputes: The view from India. In: Pistone P, de Goede J, editors. Flexible multi-tier dispute resolution in international tax disputes. Amsterdam: IBFD; 2021. p. 345–376. doi:10.59403/22m7kbp016
51. Bilaney SK. Transfer pricing litigation trends and dispute resolution mechanisms. Int Transf Pricing J. 2016;24(1). doi:10.59403/x3g7vy
52. Butler MSJR, Pengelly J, Wang L. Commissioner issues draft Practical Compliance Guideline on cross-border related-party financing arrangements. Int Transf Pricing J. 2017;24(5):379–386. doi:10.59403/1whz81j
53. Joseph A. Arm's length debt test: Australian Tax Office issues guidance. Int Transf Pricing J. 2020;28(1). doi:10.59403/s83gd4