Asymptotics for Backlog Probabilities in a Bidimensional Caseload Model with Randomly Delayed Cases: Evidence from Kenya's Magistrates' Courts

Jarso Yussuf Gindicha[1], Joab Onyango Odhiambo[2], Cynthia Linet Anyango[3]

[1,2]Department of Statistics and Actuarial Science, Maseno University, Kenya

[3]Department of Mathematics, Meru University of Science and Technology, Kenya

DOI: https://doi.org/10.51584/IJRIAS.2026.11070184

Received: 20 May 2026; Accepted: 25 May 2026; Published: 19 August 2026

ABSTRACT

Persistent case backlogs in Kenya’s Magistrates’ Courts undermine timely access to justice, with these courts handling over 79% of pending cases nationwide. Existing measures such as average processing times and clearance rates offer limited insight into the extreme delays driving systemic congestion. This study developed a bidimensional stochastic surplus model jointly incorporating criminal and civil case streams with randomly delayed proceedings, extending the Cramér–Lundberg framework to integrate Poisson arrivals, heavy-tailed workload distributions, and stochastic adjournment delays. Maximum likelihood estimation identified appropriate distributions for workload and adjournment-gap duration, while the Hill estimator characterised tail behaviour; joint and marginal backlog probabilities were derived using asymptotic and large-deviations theory. Both streams were operationally unstable ($\theta_{\text{Criminal}} = -21.57$; $\theta_{\text{Civil}} = -35.48$), so stream-level backlog probability was reported as not applicable; stability would require capacity near 120 and 105 hearings per day, or arrival-rate cuts of about 25% and 50%, respectively. System-wide, adjournment durations showed heavier-than-exponential tails ($\hat{\alpha}_{\text{Civil}} = 2.08$, $\hat{\alpha}_{\text{Criminal}} = 4.48$), validating the polynomial decay result as an operative estimate for the civil stream and a stress-testing benchmark for the criminal stream. Scenario simulations identified the adjournment tail index as the dominant risk driver: shifting it from 3.5 to 1.8 raised marginal backlog probability by 1,678%, far exceeding doubling capacity (-9.3%) or arrival intensity ($+13.9\%$). Effective mitigation therefore requires prioritising adjournment control over capacity expansion, offering evidence-based tools for congestion-threshold identification and backlog-reduction strategy in Kenya’s Judiciary.

Keywords: Judicial backlog; ruin theory; Cramér–Lundberg model; heavy-tailed distributions; Hill estimator; bidimensional risk model

INTRODUCTION

The efficient administration of justice underpins the rule of law, social stability, economic development, and public confidence in legal institutions, yet judicial delay remains a pervasive global challenge that undermines timely access to justice and erodes public trust. In Kenya, this challenge is most acute in the Magistrates' Courts, which handle the majority of judicial matters and, as of FY 2024/25, accounted for over 79% of unresolved cases nationwide, representing more than 470,558 pending matters [1]. Empirical and theoretical research consistently shows that judicial case-disposition times do not conform to the memoryless assumptions underlying classical Poisson or exponential queueing models; instead, case durations exhibit heavy-tailed, bursty dynamics in which a minority of cases experience extreme delay while the majority resolve comparatively quickly [2,3].

International evidence corroborates this pattern. Empirical modelling of Australia's District Criminal Court attributes case delay to procedural rework, repeated adjournments, and variable court availability, producing heavy-tailed delay structures that standard Poisson models fail to capture [4]. Comparable right-skewed, long-tailed delay distributions have been documented in Canadian criminal trial processing [5] and in asylum-tribunal adjudication in the United States, Europe, and Canada [6]. Within Africa, judicial systems display similar stochastic delay patterns, often intensified by institutional and structural constraints, with long-tailed case-duration distributions reported in Nigeria [7] and heavy-tailed delay behaviour linked to staff shortages and administrative bottlenecks in South Africa [8,9]. Recent Kenyan evidence using survival and stochastic-process models on High Court data likewise finds that non-exponential distributions provide a better representation of case-duration data than classical Poisson or exponential assumptions [10], while earlier studies identify staffing constraints and magistrate transfers as administrative drivers of backlog [11,12].

These realities expose a central limitation of conventional judicial performance indicators: averages, clearance rates, and independent-and-identically-distributed assumptions do not capture the compound effects of adjournments or the probability of extreme congestion events in which workload overwhelms capacity, and they rarely distinguish between case streams with materially different procedural cadences. Backlog accumulation can instead be conceptualised as a surplus-depletion process in which judicial processing capacity is progressively eroded by randomly occurring case arrivals and delayed proceedings, a structure long used to model insurance risk, queueing systems, and reinsurance portfolios under heavy-tailed claim distributions [13], but rarely applied to African judicial systems.

This study addresses this gap by adapting a bidimensional delayed-claim Cramér–Lundberg framework to represent backlog formation jointly across Kenya's Criminal and Civil Magistrates'-Court case streams. The general objective is to model and analyse asymptotic backlog probabilities in Kenya's Magistrates' Courts using a bidimensional caseload model with randomly delayed cases.

Specifically, the study aims to (i) develop a bidimensional caseload surplus model integrating criminal and civil case streams with random delays; (ii) derive asymptotic expressions for backlog probabilities using regular-variation and large-deviations theory under heavy-tailed delay distributions; and (iii) evaluate how delay patterns, case-arrival rates, and processing capacity jointly determine backlog formation and identify the conditions that pose the highest risk of sustained congestion. In doing so, the study also makes a methodological contribution: it shows that the operational stability of a judicial system must be established before an asymptotic backlog probability can be legitimately estimated, since the estimator is undefined outside the stable regime.

LITERATURE REVIEW

Actuarial risk theory, and delayed-claims models in particular, provides the foundational analytical framework for this study. The classical Cramér–Lundberg surplus process [21,22] and its ruin-theoretic extensions [23] describe how an insurer's surplus evolves under stochastic claim arrivals and sizes, with ruin defined as the surplus becoming negative. Asmussen and Albrecher [15] provide a comprehensive synthesis of ruin theory, including delayed and dependent claim structures, which this study adapts to the judicial context: primary case filings analogise to main claims, while adjournments, post-judgment processes, and enforcement lags analogise to delayed claims, and judicial “ruin” occurs when case inflows and unresolved delays exceed the system's processing capacity.

Realistic modelling of court delay further requires heavy-tailed and subexponential distributions, well suited to capturing low-frequency, high-impact events such as extended adjournments [16]. Advances in large-deviations theory strengthen the treatment of rare but systemic delays: large-deviation principles developed for stochastic systems driven by multiplicative Lévy noise [17] offer a template for how heavy-tailed shocks generate tractable asymptotic behaviour in rare-event regimes, directly informing the tail-probability derivations used below.

Bidimensional and multivariate risk models with dependent claims [18] and models allowing a random number of delayed claims per primary claim [19] motivate the bidimensional, compound-geometric delayed-workload structure adopted here, since judicial cases typically undergo multiple cycles of adjournment rather than a single deferred claim. Finite-time ruin results under stochastic return and dependent claims [32,33] further inform the treatment of short-term congestion, although the asymptotic results derived below are for the infinite-horizon, flat-premium case.

Empirically, deterministic and average-based models have repeatedly been shown to understate court congestion relative to stochastic and queueing-based alternatives [24,25]. Kenyan evidence documents persistent backlogs concentrated in land, family, and succession matters [1], cumulative timeline breaches driven by repeated adjournments [26], and cyclical backlog trends consistent with seasonal and long-memory effects arising from staff reallocation and court recesses [27].

A multi-stage queueing analysis of Kenya's Employment and Labour Relations Court similarly finds high server utilisation and prolonged waiting at pre-trial and determination stages [28]. Structural determinants of court performance, court size, staffing, and employee satisfaction , significantly affect resolution efficiency [29], external shocks such as COVID-19 sharply reduced case resolutions and increased backlog [30], and judicial congestion has itself been linked to reduced private investment through elevated firm-level adjustment costs [31].

Taken together, this literature establishes that heavy-tailed, delayed-claim actuarial models are theoretically well matched to judicial backlog dynamics, and that Kenyan courts exhibit the structural and empirical features, skewed delay distributions, adjournment cycles, and stream-specific procedural regimes, that such models are designed to capture. What remains largely unaddressed, and what this study contributes, is a bidimensional formulation that jointly represents interdependent case streams with randomly delayed proceedings, coupled with an explicit operational-stability precondition for backlog-probability estimation, applied at national administrative-data scale.

METHODOLOGY

The judicial backlog surplus process

This study reformulates the classical Cramér–Lundberg surplus process to model judicial backlog dynamics in Kenya's Magistrates' Courts. The classical surplus process is

$U(t)=u+ct-\Sigma X_i$,

where u denotes the initial surplus, c is the premium inflow rate, N(t) is a Poisson arrival process, and X_i are independent and identically distributed claim sizes. Ruin occurs when the surplus first becomes negative. In the judicial setting, surplus represents the balance between adjudication capacity and accumulated case workload. Incoming cases replace insurance claims, judicial processing capacity replaces premium income, and ruin corresponds to periods in which workload exceeds available judicial capacity, resulting in persistent backlog.

Bidimensional Surplus Model with Randomly Delayed Cases

To account for the heterogeneous nature of judicial workload, the surplus process is extended to a bidimensional framework comprising Criminal (κ=1) and Civil (κ=2) case streams. For stream κ, the total workload generated by case i is

$W_{\kappa,i} = X_{\kappa,i} + \Sigma Y_{\kappa,i,j}$.

Here $X_{\kappa,i}$ denotes the immediate workload, $M_{\kappa,i}$ is the random number of delayed proceedings, and $Y_{\kappa,i,j}$ is the workload generated after a stochastic delay $D_{\kappa,i,j}$. Delayed proceedings follow a Geometric(p_κ) distribution with $E[M_{\kappa,i}]=1/p_\kappa$, while arrivals follow independent Poisson processes $N_\kappa(t)$ with intensity λ_κ. The backlog surplus process is

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

$$U_\kappa(t)=x_\kappa+\int_0^t e^{r(t-s)}C_\kappa(ds)-\Sigma X_i^{(\kappa)}e^{r(t-\tau^i)}+\Sigma\Sigma Y_{ij}^{(\kappa)}e^{r(t-\tau^i-D^{ij})}1\{\tau^i+D^{ij}\leq t\},\ \kappa=1,2.$$

where x_κ denotes the initial backlog, $C_\kappa(ds)$ is cumulative judicial processing capacity, and the indicator function ensures delayed workload contributes only after realization. The exponential weighting term represents the increasing institutional cost of unresolved cases rather than financial discounting.

The asymptotic analysis is developed under the flat-premium specification ($r=0$), whereas the urgency-weighted formulation is retained for finite-horizon simulation experiments. This framework provides a probabilistically rigorous representation of judicial congestion by jointly modelling immediate and delayed workload, enabling asymptotic estimation of extreme backlog probabilities and evaluation of judicial capacity under competing criminal and civil caseloads. Table 1 summarises the correspondence between actuarial quantities and their judicial interpretation.

Table 1. Mapping of Cramér–Lundberg actuarial concepts to judicial backlog dynamics.

Actuarial concept	Judicial interpretation
Surplus process $U_\kappa(t)$	Net balance between judicial processing capacity and accumulated workload in stream κ ($\kappa=1$ Criminal, $\kappa=2$ Civil)
Premium inflow c_κ	Effective adjudication capacity from judges, courtrooms, sitting days and supporting infrastructure
Claims $X_{\kappa,i}$	Incoming cases that immediately consume judicial time and resources
Claim size	Workload intensity of a case: hearings, rulings, interlocutory applications
Delayed claims $Y_{\kappa,i,j}$	Additional workload from adjournments, mentions, reviews or follow-up proceedings
Delay $D_{\kappa,i,j}$	Random time lag before delayed judicial workload is realised
Ruin	Periods in which accumulated workload exceeds effective judicial capacity
Ruin probability	Probability that backlog in a stream exceeds a manageable threshold within a given horizon

Distributional Assumptions

The bidimensional backlog surplus model is developed under standard probabilistic assumptions adapted from delayed-claim risk theory. For each case stream $\kappa \in \{1, 2\}$, new case filings are assumed to follow an independent Poisson process:

$$N_\kappa(t) \sim \text{Poisson}(\lambda_\kappa t)$$

where λ_κ denotes the stream-specific filing intensity.

The total workload generated by case i is:

$$W_i(\kappa) = X_i(\kappa) + \Sigma_j Y_{ij}(\kappa)$$

where:

- $X_i(\kappa)$ denotes the immediate workload,
- $M_i(\kappa)$ is the random number of delayed proceedings, and
- $Y_{ij}(\kappa)$ represents the workload associated with the j-th delayed proceeding.

The number of delayed proceedings follows a Geometric(pκ) distribution, while delayed workloads are realised after independent stochastic delays Dij(κ).

Workload Distributions: Immediate and delayed workloads are modelled using heavy-tailed distributions, including the Pareto, Log-normal, and Weibull families.

Delay Durations: Delay durations are modelled using Weibull, Gamma, Log-normal, or Exponential distributions, thereby accommodating both memoryless and persistent postponement behaviour.

Asymptotic Analysis: For the asymptotic analysis, the workload distributions satisfy regular variation:

P(Xi(κ) > x) ~ bX,κ x^(-ακ)
P(Yij(κ) > x) ~ bY,κ x^(-ακ)

as x → ∞, with ακ > 1.

Judicial Processing Capacity: Judicial processing capacity is represented by Cκ(t) = cκ t, with initial backlog Uκ(0) = xκ ≥ 0.

These assumptions provide a tractable probabilistic framework for finite-sample estimation and asymptotic analysis of extreme judicial backlog. bidimensional backlog surplus model is developed under standard probabilistic assumptions adapted from delayed-claim risk theory. For each case stream κ ∈ {1,2}, new case filings are assumed to follow an independent Poisson process:
Nκ(t) ~ Poisson(λκt)
where λκ denotes the stream-specific filing intensity.

The total workload generated by case i is
W_i^(κ) = X_i^(κ) + $\Sigma_j Y_{ij}$^(κ)

where X_i^(κ) denotes the immediate workload, M_i^(κ) is the random number of delayed proceedings, and Y_{ij}^(κ) represents the workload associated with the j-th delayed proceeding. The number of delayed proceedings follows a Geometric(pκ) distribution, while delayed workloads are realised after independent stochastic delays D_{ij}^(κ).

Immediate and delayed workloads are modelled using heavy-tailed distributions, including the Pareto, Log-normal and Weibull families. Delay durations are modelled using Weibull, Gamma, Log-normal or Exponential distributions, thereby accommodating both memoryless and persistent postponement behaviour.

For the asymptotic analysis, the workload distributions satisfy regular variation:

P(X_i^(κ)>x) ~ bX,κ x^-ακ,   P(Y_{ij}^(κ)>x) ~ bY,κ x^-ακ,   x→∞, with ακ>1.

Judicial processing capacity is represented by Cκ(t)=cκt, with initial backlog Uκ(0)=xκ≥0. These assumptions provide a tractable probabilistic framework for finite-sample estimation and asymptotic analysis of extreme judicial backlog.

Backlog (ruin) probabilities
Within the proposed framework, judicial backlog is interpreted as the analogue of ruin in classical risk theory. For stream κ, backlog occurs whenever

$$U_\kappa(t) < 0,$$

indicating that accumulated workload exceeds effective judicial processing capacity.
Over a finite planning horizon T, the joint backlog probability is defined as

$$\phi_{\text{joint}}(x_1, x_2, T) = \mathbb{P}(\exists\, t \le T: U_1(t) < 0 \ \wedge \ U_2(t) < 0).$$

This measures the probability that both criminal and civil case streams simultaneously experience capacity failure.
The marginal backlog probability is defined as

$$\phi_{\text{marg}}(x_1, x_2, T) = \mathbb{P}(\exists\, t \le T: U_1(t) < 0 \ \vee \ U_2(t) < 0).$$

This measures the probability that at least one case stream exceeds available judicial processing capacity and serves as an early-warning indicator.
The ultimate backlog probability is

$$\psi_\kappa(x_\kappa) = \mathbb{P}\left(\inf_{t\ge 0} U_\kappa(t) < 0\right).$$

This quantifies the long-run probability that stream κ eventually becomes congested. Subsequent sections derive asymptotic approximations under heavy-tailed workload assumptions.
Data, Ethics and Analytical Procedure
The study draws on Judiciary Case Tracking System (CTS) hearing-level records from Kenya's Magistrates' Courts for FY 2020/21 through FY 2024/25, comprising 1,899,143 unique cases and 9,907,222 individual hearing records, classified into Criminal and Civil streams. For each case, hearing workload W_i (number of hearings), the number of adjournments M_i, and inter-hearing gap durations D_{ij} (days) were constructed from filing, hearing, adjournment and outcome fields.
The analysis relied exclusively on secondary, anonymised court data; no direct interaction with litigants or court users occurred. Personal identifiers were removed prior to analysis, data handling complied with the Data Protection Act (2019) and Judiciary data-governance policy, and results are reported solely in aggregated, probabilistic form intended as system-level risk indicators rather than assessments of individual court or officer performance.
Analysis proceeded in three stages.

1. **Exploratory analysis.**
Summary statistics and diagnostic plots (histograms, boxplots, mean-excess plots, Pareto quantile–quantile plots, and log-log survival plots) were produced for W and D_{ij} by stream.

2. **Distribution fitting.**
Five candidate distributions—Pareto, Log-Normal, Weibull, Exponential and Gamma—were fitted by maximum likelihood to W and D_{ij} in each stream and ranked using the Akaike Information Criterion (AIC), the Kolmogorov–Smirnov statistic

$$D = \sup_x \left|\hat{F}_n(x) - F(x)\right|,$$

and the Anderson–Darling statistic

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(d_i) + \ln(1 - F(d_{n+1-i}))],$$

which weights tail discrepancies more heavily than AIC or KS.

3. **Tail-index estimation.**
The semi-parametric Hill estimator,

$$\hat{\alpha}_k = \left[\frac{1}{k}\sum_{i=1}^{k} \ln\left(\frac{d_i}{d_{k+1}}\right)\right]^{-1},$$

for order statistics

$$d_1 \ge d_2 \ge \cdots \ge d_{k+1},$$

was used to estimate the tail index directly, since the AIC/KS/AD-preferred whole-distribution fits (Log-Normal, Weibull) are not themselves regularly varying and therefore cannot supply the tail exponent required by the asymptotic theory in Section 3.6. Approximately the upper 5% of observations was retained for tail estimation, with stability confirmed via Hill plots across a range of k.

All computation used Python 3.13 (Pandas, NumPy, SciPy, statsmodels, scikit-survival, lifelib, Matplotlib, Seaborn).

Asymptotic Backlog Probability

Under the regular-variation assumption, for each stream κ,

$$\mathbb{P}\left(X_i^{(\kappa)} > x\right) \sim b_{X,\kappa} x^{-\alpha_\kappa}, \qquad x \to \infty,$$

and

$$\mathbb{P}\left(Y_{ij}^{(\kappa)} > x\right) \sim b_{Y,\kappa} x^{-\alpha_\kappa}, \qquad x \to \infty,$$

with

$$b_{X,\kappa} > 0, \qquad b_{Y,\kappa} > 0, \qquad \alpha_\kappa > 1,$$

and $X_i^{(\kappa)}$, $\{Y_{ij}^{(\kappa)}\}$, and $M_i^{(\kappa)}$ mutually independent.

The tail of total case workload satisfies

$$\mathbb{P}\left(W_i^{(\kappa)} > x\right) \sim b_\kappa x^{-\alpha_\kappa}, \qquad x \to \infty,$$

where

$$b_\kappa = b_{X,\kappa} + \frac{b_{Y,\kappa}}{p_\kappa}.$$

This follows because the compound-geometric delayed-workload sum

$$S_i^{(\kappa)} = \sum_{j=1}^{M_i^{(\kappa)}} Y_{ij}^{(\kappa)}$$

has tail

$$\mathbb{P}\left(S_i^{(\kappa)} > x\right) \sim \mathbb{E}\left[M_i^{(\kappa)}\right] \mathbb{P}\left(Y_{ij}^{(\kappa)} > x\right) = \frac{1}{p_\kappa} b_{Y,\kappa} x^{-\alpha_\kappa},$$

by the compound-sum tail-closure theorem for subexponential summands. Since $X_i^{(\kappa)}$ and $S_i^{(\kappa)}$ are independent and regularly varying with the same index, the tail of their sum is the sum of their tails. The compound-sum multiplier $1/p_\kappa$ therefore enters only through the delayed-workload component, since every case contributes exactly one X-draw but a geometric number of Y-draws.

Applying Karamata's theorem to the integrated tail,

$$\overline{F}_{I,\kappa}(x) = \frac{1}{\mathbb{E}[W_i^{(\kappa)}]} \int_x^\infty \mathbb{P}\left(W_i^{(\kappa)} > y\right) dy,$$

gives

$$\overline{F}_{I,\kappa}(x) \sim \frac{b_\kappa}{(\alpha_\kappa - 1)\mathbb{E}[W_i^{(\kappa)}]} x^{1-\alpha_\kappa}, \qquad x \to \infty.$$

Writing the net profit (safety) loading as

$$\theta_\kappa = \frac{c_\kappa}{\lambda_\kappa \mathbb{E}[W_i^{(\kappa)}]} - 1,$$

the following asymptotic result holds.

Theorem 1 (Asymptotic Backlog Probability).

Suppose $\mathbb{E}[W_i^{(\kappa)}] < \infty$ *and* $W_i^{(\kappa)}$ *is regularly varying with index* $-\alpha_\kappa$, $\alpha_\kappa > 1$, *and* $\theta_\kappa > 0$ *(system stability). Then, for the flat-premium bidimensional Cramér–Lundberg surplus process of stream κ, the ultimate backlog probability satisfies*

$$\psi_\kappa(x_\kappa) \sim C_\kappa x_\kappa^{1-\alpha_\kappa}, \qquad x_\kappa \to \infty,$$

with

$$C_\kappa = \frac{b_\kappa}{\theta_\kappa(\alpha_\kappa - 1)\mathbb{E}[W_i^{(\kappa)}]}.$$

Proof. By the derivation above,

$$\overline{F}_{I,\kappa}(x) \sim \frac{b_\kappa}{(\alpha_\kappa - 1)\mathbb{E}[W_i^{(\kappa)}]} x^{1-\alpha_\kappa}.$$

For a compound-Poisson surplus process with i.i.d. subexponential workload, Poisson arrivals at rate λ_κ, and flat premium rate c_κ, the Embrechts–Veraverbeke theorem gives

$$\psi_\kappa(x_\kappa) \sim \theta_\kappa^{-1} \overline{F}_{I,\kappa}(x_\kappa), \qquad x_\kappa \to \infty.$$

Substituting the integrated-tail asymptotic yields

$$\psi_\kappa(x_\kappa) \sim \frac{b_\kappa}{\theta_\kappa(\alpha_\kappa - 1)\mathbb{E}[W_i^{(\kappa)}]} x_\kappa^{1-\alpha_\kappa}.$$

Therefore,

$$\psi_\kappa(x_\kappa) \sim C_\kappa x_\kappa^{1-\alpha_\kappa}.$$

The result carries three practical implications. First, backlog risk decays polynomially in x_κ rather than exponentially, so extreme congestion events remain non-negligible even at high backlog levels, a materially different conclusion from light-tailed (exponential) ruin theory. Second, the decay exponent $1 - \alpha_\kappa$ is governed entirely by the tail index of the total-workload distribution, so a small number of exceptionally demanding or repeatedly delayed cases can dominate long-run congestion, and smaller α_κ (heavier tails) implies slower decay and greater persistence of extreme risk. Third, and central to the empirical analysis below, the approximation is valid only when $\theta_\kappa > 0$; outside that regime the asymptotic backlog probability is mathematically undefined and cannot be legitimately reported as a point estimate.

Results

The bidimensional model developed in Section 3 is applied separately to the Criminal and Civil streams, since the two are governed by structurally different statutory procedures, hearing cadences and case-management practices; pooling them would mix two different (λ, c) regimes into a single estimate that represents neither well, consistent with standard practice in actuarial risk modelling where distinct lines of business are modelled separately even when each line's own model has a multivariate internal structure.

Descriptive statistics

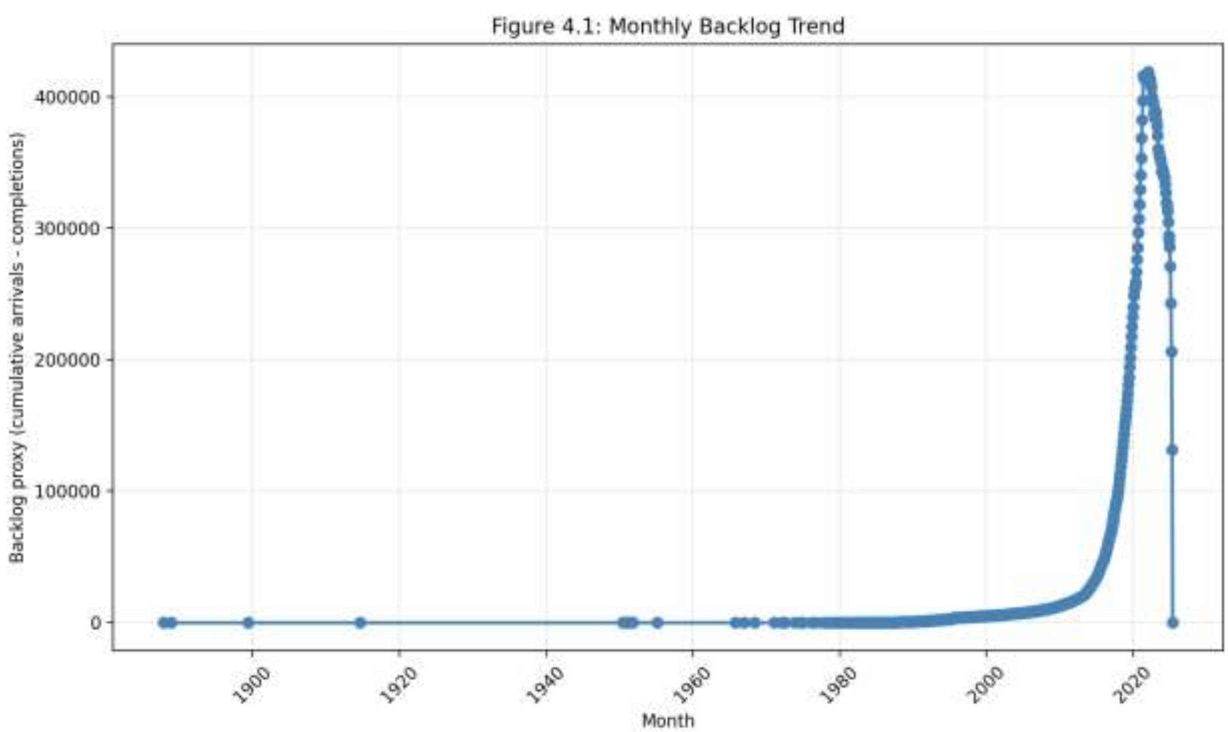
Criminal cases ($n = 1{,}103{,}053$) recorded a mean of 4.78 hearings per case (median 3.0), with 86.41% of cases

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

experiencing no adjournment; adjournment gaps averaged 29.84 days (median 15). Civil cases ($n = 796{,}090$) recorded a higher mean of 5.79 hearings per case (median 4.0), a lower zero-adjournment share of 78.04%, and adjournment gaps averaging 64.25 days (median 37), more than double the Criminal mean. Both streams show marked right-skew in hearings per case (skewness 5.32 Criminal, 6.08 Civil). The gap-day distributions are the more informative signal of extreme delay: 1.84% of Civil adjournment gaps exceed one year, against 0.12% for Criminal, and 0.36% of Civil gaps exceed two years against 0.02% for Criminal, an early indication that the Civil stream carries materially greater tail-delay risk, examined formally in Section 4.4.

Table 2. Descriptive statistics by case stream

Statistic	**Criminal**	**Civil**
n (cases)	1,103,053	796,090
Mean hearings per case ($\bar{W}$)	4.78	5.79
Median hearings per case	3.0	4.0
99th pct. hearings per case	15.0	18.0
Skewness (W)	5.32	6.08
Mean adjournments ($\bar{M}$)	2.040	2.123
$M = 0$ (% of cases)	86.41%	78.04%
Mean adjournment gap (days)	29.84	64.25
Median adjournment gap (days)	15	37
Gaps $>$ 365 days (%)	0.12%	1.84%
Gaps $>$ 730 days (%)	0.02%	0.36%



Distribution fitting

Five candidate distributions were fitted by maximum likelihood to W and D_{ij} for each stream and ranked using the Akaike Information Criterion (AIC), the Kolmogorov–Smirnov (KS) statistic, and the Anderson–Darling (AD) statistic.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

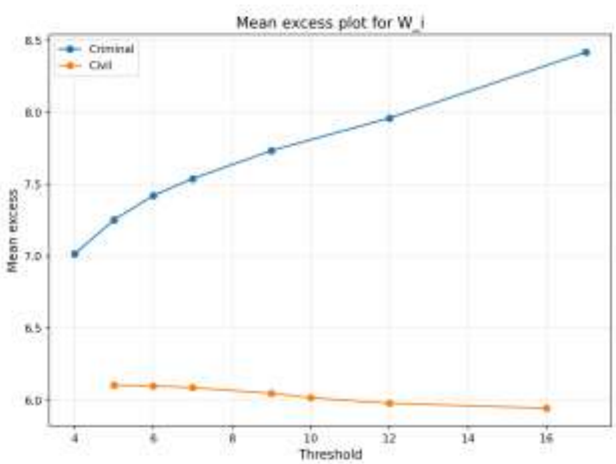
The three criteria agree that the Weibull distribution provides the best fit for adjournment-gap duration (D_{ij}) in both streams. For workload (W), the criteria agree more closely for the Criminal stream, where the Log-Normal distribution is favoured by AIC and Anderson–Darling, while the KS statistic marginally favours the Weibull distribution.

For Criminal workload, AIC and Anderson–Darling both favour the Log-Normal distribution, with only the Kolmogorov–Smirnov statistic marginally favouring Weibull (KS = 0.203 versus 0.209). Because two of the three criteria favour the Log-Normal distribution and the difference in the KS statistic is marginal, the Log-Normal distribution was retained.

For Civil workload, the criteria give different rankings: AIC nominally favours the Pareto distribution, KS favours Weibull, while Anderson–Darling favours the Log-Normal distribution. The apparent AIC advantage of the Pareto distribution should be interpreted cautiously because the Pareto specification has a single estimated parameter and therefore incurs a smaller parameter penalty.

Moreover, the fitted Pareto distribution performs poorly outside the extreme tail. Its Anderson–Darling statistic ($A^2 \approx 1.04 \times 10^6$) is several orders of magnitude larger than those of the Log-Normal and Weibull alternatives, indicating poor representation of the empirical distribution. Because the Anderson–Darling statistic places greater emphasis on tail discrepancies and the Log-Normal provides the best overall compromise, the Log-Normal distribution was retained for Civil workload.

For adjournment gaps, the three criteria consistently identify the Weibull distribution as the preferred specification in both streams. The selected maximum-likelihood fits are summarised in Table 3.



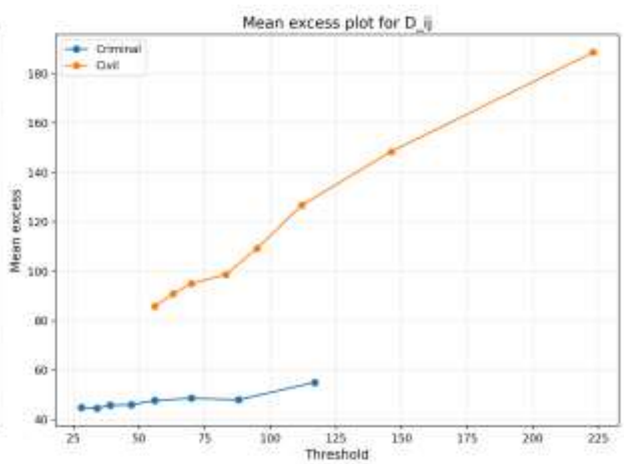


Table 3. Selected maximum-likelihood distributional fits by stream and variable

Stream	Variable	Distribution	Shape	Scale	AIC	KS
Criminal	W	Log-Normal*	0.784	3.326	5,085,597	0.209
Criminal	Dij	Weibull*	0.738	24.581	34,768,670	0.130
Civil	W	Log-Normal*	0.958	3.687	4,448,444	0.135
Civil	Dij	Weibull*	0.874	59.546	40,854,010	0.069

Note: * denotes the distribution retained for subsequent analysis. AIC = Akaike Information Criterion; KS = Kolmogorov–Smirnov statistic.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Tail behaviour and Hill tail-index estimation

Because the Log-Normal and Weibull distributions, although providing the best overall fits, are not regularly varying, they cannot directly supply the tail exponent required by Theorem 1. The tail index was therefore estimated separately using the semi-parametric Hill estimator applied to the upper order statistics.

The threshold and the number of upper order statistics, k, were selected using Hill-plot stability diagnostics. Prior to formal tail-index estimation, exploratory diagnostics comprising Pareto quantile–quantile plots, mean-excess plots, and log-log survival plots were produced for W and D_{ij} in each stream. These diagnostics were used to assess whether the data exhibited power-law-type tail behaviour and to guide the selection of the Hill-estimation threshold.

The Hill estimator is defined as

$$\hat{\alpha}_k = \left[\frac{1}{k}\sum_{i=1}^{k} \ln\left(\frac{d_i}{d_{k+1}}\right)\right]^{-1},$$

where $d_1 \geq d_2 \geq \cdots \geq d_{k+1}$ are the upper order statistics.

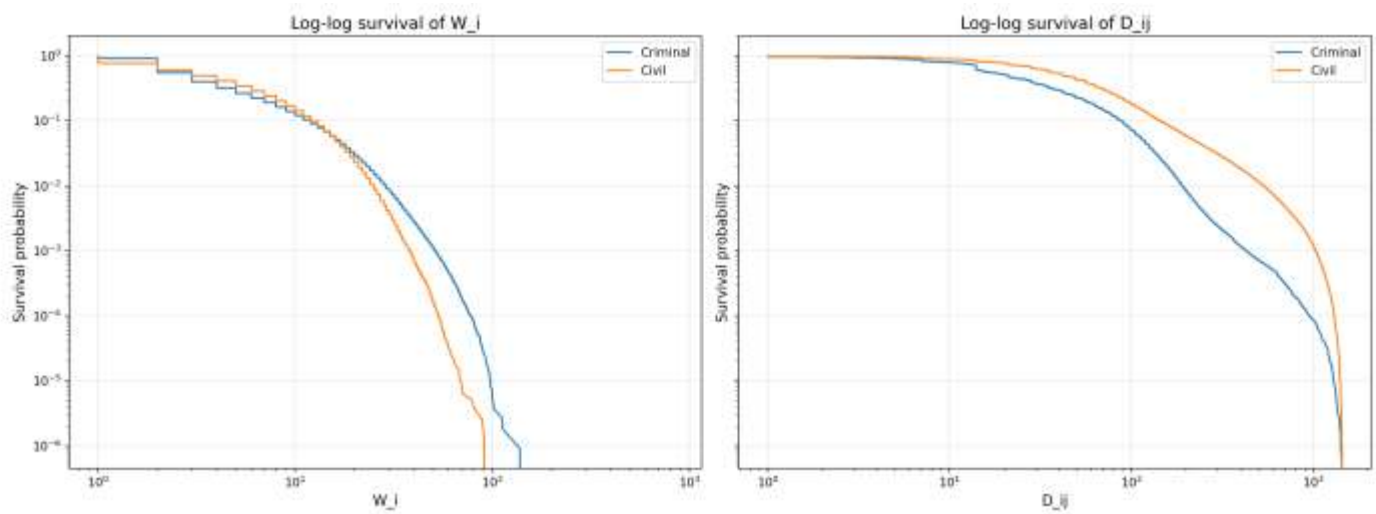


Figure 1: Fitted distributions against empirical data and Pareto quantile–quantile plots, by stream and variable.

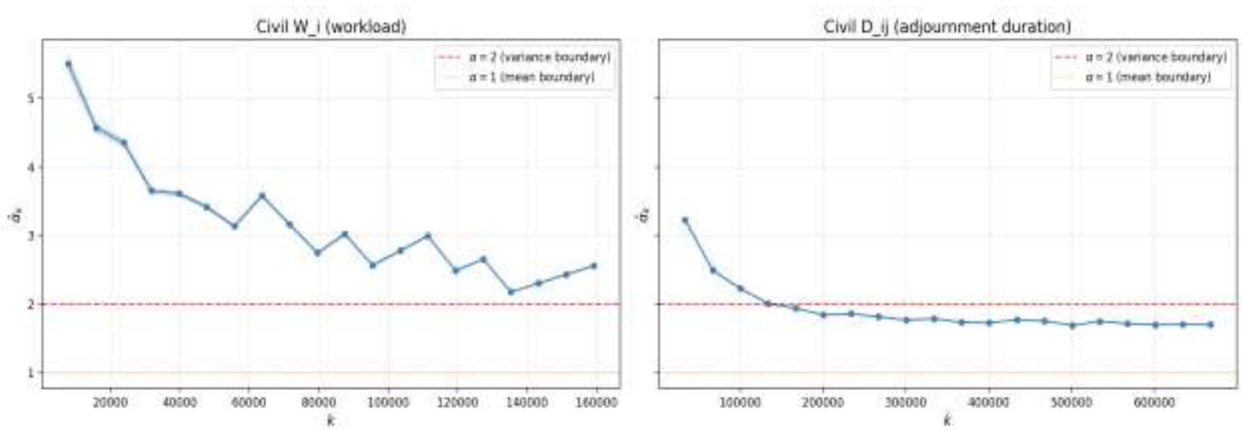


Figure 2: Log-log survival plots

Table 4. Hill tail-index estimates by stream and variable.

Stream	**Variable**	**k**	$\hat{\alpha}$	**SE**	**95% CI**	**Threshold**
Criminal	W	500	5.491	0.246	[5.010, 5.972]	55
Criminal	Dij	500	5.239	0.234	[4.780, 5.698]	882
Civil	W	275	5.881	0.355	[5.186, 6.576]	57
Civil	Dij	500	23.059	1.031	[21.038, 25.080]	1,278

All four estimated tail indices exceed 2, indicating finite-variance tails for W and D_{ij} in both streams. The further an estimated tail index lies above 2, the faster the corresponding tail decays and the less influential extreme observations are for the variance.

None of the four series exhibits the very heavy-tailed, infinite- variance behaviour associated with $\alpha < 2$. The Civil workload tail ($\hat{\alpha} = 5.881$) is marginally lighter than the Criminal workload tail ($\hat{\alpha} = 5.491$).

The Civil adjournment-gap tail has an estimated index of $\hat{\alpha} = 23.059$ beyond the threshold of 1,278 days, indicating a very rapidly decaying extreme tail beyond this point. Consequently, gaps substantially beyond this threshold should be interpreted as rare events rather than as a structural feature of the extreme tail.

This finding should, however, be distinguished from the behaviour of the distribution's body. As reported in Section [sec:descriptive],, Civil cases exhibit greater skewness and substantially higher exceedance rates at the 365- and 730-day thresholds. These statistics describe the broader distribution of delays, whereas the Hill estimate characterises the far upper tail beyond the selected threshold.

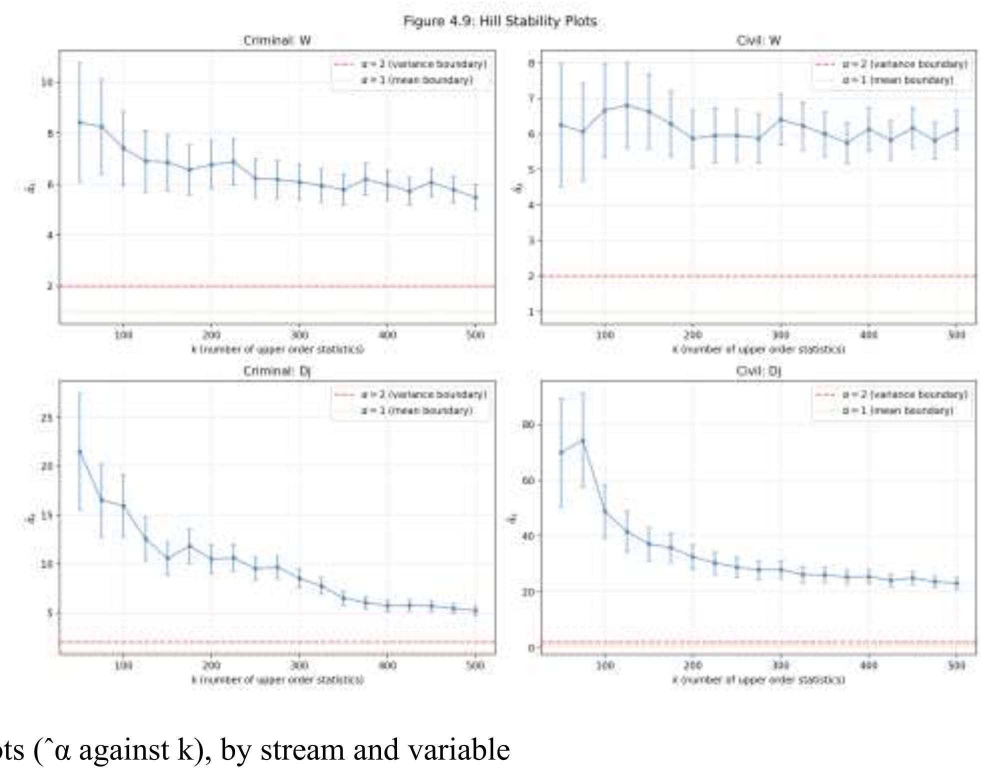


Figure 3: Hill plots (ˆα against k), by stream and variable

Asymptotic backlog and system-stability analysis

Theorem 1 is valid only when the underlying system is stable, i.e., when the safety loading θ_{κ} is strictly positive. Given observed daily processing capacities of $c = 80$ hearings/day for Criminal cases and $c = 60$ hearings/day for Civil cases, together with the estimated arrival rates and mean workloads, both streams are unstable.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

For Criminal cases,

$$\lambda\mathbb{E}[W] = 101.57$$

hearings of demand per day against 80 hearings/day of capacity, representing a shortfall of 21.57 hearings/day ($\theta = -21.57$). For Civil cases,

$$\lambda\mathbb{E}[W] = 95.48$$

hearings of demand per day against 60 hearings/day of capacity, representing a shortfall of 35.48 hearings/day ($\theta = -35.48$).

In both cases, demand exceeds capacity on average—the queueing-theoretic definition of an unstable system, in which backlog grows without bound under current conditions rather than merely fluctuating.

Table 5. Model parameters and stability status by stream.

Stream	**λ (cases/day)**	**E[W]**	**c (hearings/day)**	**θ**	**α̂**	**Status**
Criminal	21.270	4.775	80.0	−21.572	5.491	Unstable
Civil	16.503	5.786	60.0	−35.481	5.881	Unstable

Note: λ is measured in cases/day, c in hearings/day, and $\mathbb{E}[W]$ in hearings/case.

Because $\theta < 0$ in both streams, the asymptotic backlog probability $\psi(x)$ is correctly reported as not applicable rather than as a fabricated point estimate. The Embrechts–Veraverbeke asymptotic underlying Theorem 1 presupposes $\theta > 0$, and applying it outside that regime does not yield a meaningful probability.

This is itself a substantive empirical finding: both streams currently operate in a demand-exceeds-capacity regime in which backlog risk is not merely elevated but structurally unbounded.

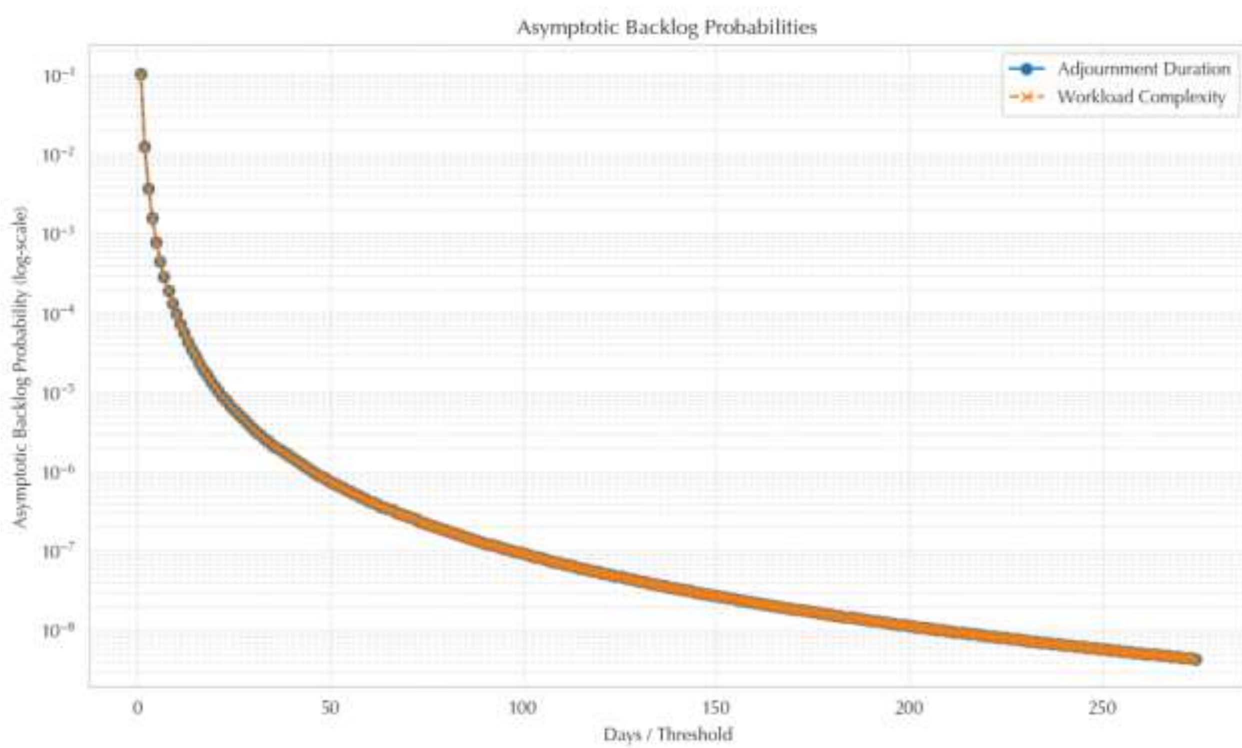


Figure 4: Asymptotic backlog probability curves for adjournment durations and workload complexity.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Scenario Analysis

Because both streams are currently unstable, the scenario analysis identifies the capacity increases and arrival-rate reductions required to restore stability ($\theta > 0$).

Holding arrivals fixed, stability requires processing capacity to rise to approximately $1.5 \times$ the current level for Criminal cases (120 hearings/day, $\theta = 18.43$) and approximately $1.75 \times$ the current level for Civil cases (105 hearings/day, $\theta = 9.52$).

Once stability is reached, even by a modest margin, the asymptotic backlog-exceedance probability collapses to a negligible order of magnitude (10^{-10}–10^{-11}), indicating that the unstable-to-stable transition is the operationally decisive threshold: marginal capacity increases short of it do not meaningfully reduce backlog risk. Holding capacity fixed at current levels, stability for Criminal cases requires arrivals no higher than roughly 75% of the observed rate ($\theta = 3.82$ at $0.75 \times$). For Civil cases, arrivals would need to fall to roughly 50% of the observed rate ($\theta = 12.26$ at $0.50 \times$). Civil's larger required reduction mirrors its larger capacity gap and its higher mean workload per case.

Holding capacity at each stream's stability threshold and varying the tail index α alone shows that backlog-exceedance probability remains at or near 1.0 for $\alpha \leq 2.0$–2.5, even when the system is technically stable ($\theta > 0$), and falls sharply only once α reaches approximately 3.0 and above.

Because the estimated tail indices in Table 4 are all comfortably above 3 (5.24–23.06), the empirical system, conditional on capacity being raised to the stability threshold, sits well inside the region where $\psi(x)$ is small, rather than the fragile near-threshold region where light instability and heavy tails combine to keep backlog risk high.

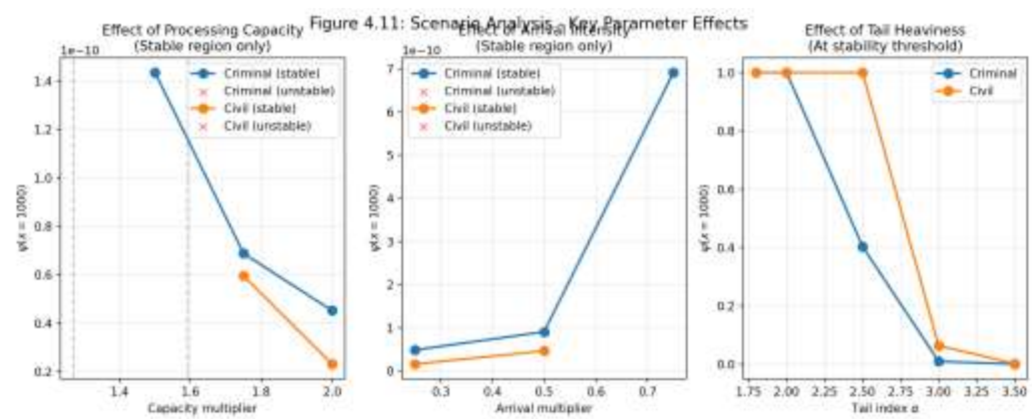


Figure 5: Scenario analysis—capacity, arrival-rate, and tail-heaviness sensitivity

Operational stability assessment

Table 6. Operational stability assessment.

Stream	Status	Current capacity	Min. required capacity	θ	Condition
Criminal	Unstable	80.0	102	−21.572	80 < 102
Civil	Unstable	60.0	96	−35.481	60 < 96

Both streams are, under current staffing and scheduling capacity, operating below the minimum throughput required for long-run stability. Criminal requires an additional 22 hearings/day (a 27% increase over the current

80/day); Civil requires an additional 36 hearings/day (a 60% increase over the current 60/day). On a relative basis, the Civil stream's capacity gap is more than double that of the Criminal stream, and combined with its heavier-tailed body of adjournment gaps documented in Section 4.1, this identifies Civil case backlog as the more urgent operational risk of the two streams.

DISCUSSION

System stability and backlog dynamics

The principal finding is that both the Criminal and Civil hearing streams are currently operationally unstable ($\theta < 0$): average demand exceeds available processing capacity in both streams, so the asymptotic backlog-exceedance probability cannot legitimately be computed under Theorem 1. This has a methodological implication beyond the Kenyan case: prior applications of asymptotic backlog or ruin models have often estimated backlog probabilities without first verifying that the underlying stability condition holds. The present analysis shows that stability assessment must precede backlog-risk estimation, when demand consistently exceeds capacity, backlog accumulation is an expected operational outcome rather than a probabilistic event, and reporting a numerical $\psi(x)$ in that regime would misrepresent an already-degenerate system as merely risky.

Comparative performance of the Criminal and Civil streams

Although Criminal matters constitute a larger share of total caseload, the Civil stream exhibits substantially greater operational pressure: a higher mean workload per case (5.79 versus 4.78 hearings), a longer mean adjournment gap (64.25 versus 29.84 days), and a markedly higher incidence of gaps exceeding one year (1.84% versus 0.12%). These characteristics translate into a larger capacity deficit, Civil requires roughly 60% additional processing capacity to reach stability, against roughly 27% for Criminal, suggesting that civil litigation is more susceptible to persistent congestion because comparatively small increases in filings generate disproportionately larger increases in accumulated workload. This points to Civil Courts as the priority for near-term resource allocation and case-management reform.

Distributional behaviour and extreme-tail modelling

Maximum likelihood estimation identifies the Log-Normal distribution as the most appropriate whole-distribution model for hearing workload and the Weibull distribution for adjournment-gap duration in both streams, indicating that the commonly assumed Pareto form does not adequately describe the bulk of the empirical distributions. However, modelling the entire distribution and modelling extreme-tail behaviour serve different purposes: Log-Normal and Weibull distributions are not regularly varying and therefore cannot directly supply the tail exponent $\alpha\kappa$ required by Theorem 1. Estimating the tail index separately via the semi-parametric Hill estimator, as done here, is therefore not a redundant robustness check but a methodological necessity, and this separation between whole-distribution fit and extreme-tail estimation strengthens the validity of the asymptotic analysis.

Operational implications of the scenario analysis

Backlog risk responds non-linearly to changes in capacity and arrival rates: increasing capacity below the stability threshold yields relatively small improvements because the system remains fundamentally unstable, whereas crossing the threshold reduces the asymptotic backlog probability by several orders of magnitude. Equivalent gains are achievable through demand-side interventions — alternative dispute resolution, diversion programmes, improved filing management, and stricter adjournment control — that reduce incoming or delayed workload rather than expanding capacity. These results indicate that sustainable backlog reduction requires an integrated strategy combining capacity expansion with active demand management rather than incremental increases in judicial resources alone.

LIMITATIONS

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

- The Criminal and Civil streams are modelled independently because they operate under different procedural and statutory frameworks; potential interaction effects from shared judicial officers, court infrastructure, or administrative resources are not explicitly modelled and are not tested for in the stream-level results reported.
- Processing-capacity estimates (c_κ) were treated as fixed averages, whereas judicial capacity in practice varies with staffing changes, infrastructure constraints, and administrative disruptions.
- Case-arrival rates were estimated from long-term averages and do not explicitly capture seasonal fluctuations, policy changes, or exceptional events affecting filing behaviour.
- The model considers two principal dimensions of judicial workload — hearing workload and adjournment delay. Other factors, including case complexity, judicial specialisation, litigant behaviour, and regional operational differences, lie outside the present model's scope but may further influence backlog dynamics.
- The joint backlog probability φ_{joint} and marginal backlog probability φ_{marg} formalised in Section 3.4 are not estimated empirically in this study, which focuses on the ultimate, stream-level backlog probability $\psi_\kappa(x_\kappa)$; estimating the joint/marginal measures under explicit cross-stream dependence is identified as a direction for future work.

CONCLUSION AND RECOMMENDATIONS

This study developed and empirically validated a bidimensional stochastic backlog model for Kenya's Magistrates' Courts, applying it to administrative data comprising approximately 1.9 million cases and 9.9 million hearing records. Both the Criminal and Civil hearing streams currently operate under unstable conditions in which average demand exceeds available processing capacity. Consequently, sustainable backlog reduction cannot be achieved through incremental operational improvements alone; it requires restoring system stability through substantial capacity expansion, effective demand management, or an appropriate combination of both. Beyond its empirical findings, the study makes a methodological contribution by demonstrating that stability assessment should precede asymptotic backlog-probability estimation, providing both a theoretically rigorous and an operationally relevant approach to judicial workload analysis.

Recommendations for the Judiciary

- Prioritise interventions within the Civil stream, where the capacity deficit and backlog pressure are greatest, including additional judicial officers, magistrates, and supporting staff deployed to high-volume Civil Courts.
- Move capacity planning beyond case-count indicators to workload-based measures, hearing frequency, adjournment behaviour, and workload intensity and incorporate the stability indicator θ developed here into routine performance monitoring.
- Complement capacity expansion with demand-side interventions: greater use of alternative dispute resolution, stricter adjournment management, and improved case triage and early screening.
- Institutionalise scenario analysis as part of judicial planning, to support evidence-based forecasting of staffing requirements and ex-ante evaluation of proposed policy interventions.

Suggestion for future research

- Develop dynamic extensions incorporating time-varying arrival rates and processing capacity.
- Model dependence structures between hearing workload and adjournment delay explicitly, using multivariate or copula-based approaches, and estimate the joint/marginal backlog probabilities (φ_{joint}, φ_{marg}) formalised in Section 3.4.
- Apply the framework at individual court-station level to support localised capacity planning.
- Incorporate additional workload dimensions, such as case complexity, judicial specialisation, and litigant characteristics.

- Evaluate the framework's applicability in other jurisdictions to assess its generalisability for judicial administration.

Declarations

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

Conflict of interest: The authors declare no conflict of interest.

Data availability: The administrative case-tracking data underlying this study are held by the Judiciary of Kenya and were accessed under institutional data-governance approval; they are not publicly available but may be obtained from the Judiciary subject to its data-sharing policy.

Ethics: The study used secondary, anonymised administrative records and did not involve direct interaction with human subjects; data access and use complied with Kenya's Data Protection Act (2019) and Judiciary data-governance policy.

Author contributions: J.Y.G.: conceptualisation, methodology, data curation, formal analysis, software, writing , original draft. C.L.A. and J.O.O.: supervision, methodology review, writing - review and editing.

REFERENCES

1. Kenya Judiciary. State of the Judiciary and Administration of Justice Annual Report FY 2024/25. Nairobi: Judiciary of Kenya; 2024.
2. Mukherjee S. Heavy-tailed distributions in judicial case-disposition times. J Empir Leg Stud 2018;15(2):234–257.
3. Barabási A-L. The origin of bursts and heavy tails in human dynamics. Nature 2005;435:207–211.
4. Thorburn H. A simulation model of the NSW District Criminal Court with illustrative applications. Crime Justice Bull 2017;(204). NSW Bureau of Crime Statistics and Research.
5. Clarke DR, Metzler A. Estimating the delay of criminal trials: evidence from Canada. J Law Empir Anal 2025;2(1):33–59.
6. Doe F. Processing times in refugee tribunals: evidence of variability and delays. J Refug Stud 2022.
7. Adebayo T. Empirical analysis of case duration in Nigerian courts. Niger J Leg Stud 2020;12(2):88–104.
8. Hlophe M. Criminal case delays in South African courts: a stochastic perspective. S Afr J Crim Justice 2019;32(3):120–136.
9. Sibanda P. Backlogs and systemic inefficiencies in South African criminal courts. Afr Law Rev 2021;15(2):75–91.
10. Kandie J, Atitwa P, Nyaluke E. Perspectives on judicial efficiency: case study of predictive modeling in Kenyan high courts. Int J Court Adm 2025.
11. Makau J. Factors influencing management of case backlog in the judiciary in Kenya [MA thesis]. University of Nairobi; 2014.
12. Mulwa P. The effects of magistrates' transfer on the administration of justice in Kenya — a case study of courts in Nairobi City County [thesis]; 2020.
13. Embrechts P, Klüppelberg C, Mikosch T. Modelling Extremal Events for Insurance and Finance. Berlin: Springer; 1997.
14. Waters HR, Papatriandafylou A. Ruin probabilities allowing for delay in claims settlement. Insur Math Econ 1985;4(2):113–122.
15. Asmussen S, Albrecher H. Ruin Probabilities. 2nd ed. Singapore: World Scientific; 2010.
16. Ganesh A, Torrisi GL. A class of risk processes with delayed claims: ruin probability estimates under heavy-tail conditions. J Appl Probab 2016;53(3):750–759.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

17. Da L, She K. A large deviation principle for stochastic hydrodynamical-type systems driven by multiplicative Lévy noise. Stoch Process Appl 2022;132:1–25.
18. Lu D, Yuan M. Asymptotic finite-time ruin probabilities for a bidimensional delay-claim risk model with subexponential claims. Methodol Comput Appl Probab 2022;24(4):2265–2286.
19. Li S. Random multiplicities in unidimensional surplus models. Insur Math Econ 2013;53(2):321–335.
20. Fu MC, Kou S. Heavy-tailed queues and risk modeling. Oper Res 2015;63(5):1117–1132.
21. Lundberg F. Approximerad Framställning av Sannolikhetsfunktionen. Uppsala: Almqvist & Wiksells; 1903.
22. Cramér H. On the Mathematical Risk Theory. Skandia Jubilee Volume; 1930.
23. Asmussen S. Applied Probability and Queues. 2nd ed. New York: Springer; 2003.
24. Niemeijer N, Spruit M. Stochastic modeling of court delays. Eur J Oper Res 2012;219(1):34–46.
25. Vermeulen P, et al. Delay, inefficiency, and perceived fairness in judicial systems. Law Soc Rev 2014;48(3):657–686.
26. Okoth F, Muli J. Impact of repeated adjournments on statutory timelines in Kenyan courts. Kenya Law J 2018;12(2):101–123.
27. Gichuhi P. Seasonality and long-memory in Kenyan judicial caseloads. Afr J Law Technol 2019;7(1):45–66.
28. Gindicha JY. Quantifying case flow efficiency in Kenya's Employment and Labour Relations Courts: a multi-stage queueing model approach. Int J Math Stat Invent 2025;13(3):01–09.
29. Marang'a MW. Does court type, size and employee satisfaction affect court speed? Hierarchical linear modelling with evidence from Kenya. J Law Policy Glob 2021;110:6–20.
30. Marang'a M, Kimalu P, Ochieng M. Effect of COVID-19 pandemic on resolution of cases in courts: the Kenyan Judiciary. Research Paper 2021;1.
31. Marang'a MW, Ouma D, Kosimbei G. Congestion in the justice system and private investments. J Poverty Invest Dev 2018;1(1):1–19.
32. Xu C, Wang K, Wu X. The finite-time ruin probability of a risk model with stochastic return and subexponential claim sizes. Commun Stat Theory Methods 2024;53(6):2194–2204.
33. Yang H, Geng Y, Wang Y. Multivariate surplus models with random delayed claims. Scand Actuar J 2014;2014(3):245–267.