Supplementary Information for "Exact Finite-Sample Inference for the Exponential Proportional Hazards Model via the Empirical Characteristic Function"

Nargiza Nurmukhamedova[1*], Guzal Abdujalilova[1], Nargiza Boltaeva[1], Gulhayo Xalilova[1], Dilsuz Khamraeva[2]

[1*]National University of Uzbekistan named after Mirzo Ulugbek, Universitetskaya str. 4, Tashkent, 100174, Uzbekistan.
[2]ZARMED University, Qayum Murtazoyev str. 13A, Bukhara, 200117, Uzbekistan.

*Corresponding author(s). E-mail(s): n.nurmuhamedova@nuu.uz;
Contributing authors: rasulova2406@gmail.com; boltayevan27@gmail.com; xalilovagulhayo1@gmail.com; mohlaroyim261291@gmail.com;

Abstract

This Supplementary Information collects the full proofs of the results stated in the main text, together with the detailed derivations for the normal and Cauchy location models and the proof of the general ECF/score-weight equivalence. Equation, theorem, and lemma numbers refer to the main manuscript.

Keywords: empirical characteristic function, proportional hazards model, exact inference

1 Supplementary Note: Full proofs of results in the main text

This note gives the proofs deferred from the main text. Each proof restates, in its heading, the result it establishes; numbering follows the main manuscript.

Proof of Proposition 4.11

Proof By Theorem 4.7, the joint density of (Z_i, δ_i) factorises as $f(z, \delta; \theta_*, p) = \theta_* e^{-\theta_* z} \cdot p^{\delta}(1-p)^{1-\delta}$, where $\theta_* = (\beta+1)\theta$ and $p = 1/(1+\beta)$. The map $(\theta, \beta) \mapsto (\theta_*, p)$ is a bijection on $(0,\infty)\times(0,\infty) \to (0,\infty)\times(0,1)$. In the new parametrisation the log-likelihood *separates* as $\ln L(\theta_*, p) = \ell_1(\theta_*) + \ell_2(p)$, where

$$\ell_1(\theta_*) = n \ln \theta_* - \theta_* \, n\bar{Z},$$
$$\ell_2(p) = n\bar{\delta} \ln p + n(1-\bar{\delta}) \ln(1-p).$$

Since ℓ_1 and ℓ_2 involve disjoint parameters, each is maximised independently. $\ell_1'(\theta_*) = n/\theta_* - n\bar{Z} = 0$ gives $\hat{\theta}_* = 1/\bar{Z}$ (maximum: $\ell_1'' = -n/\theta_*^2 < 0$). $\ell_2'(p) = n\bar{\delta}/p - n(1-\bar{\delta})/(1-p) = 0$ gives $\hat{p} = \bar{\delta}$ (maximum: $\ell_2'' < 0$). Back-transforming: $\hat{\beta} = 1/\hat{p} - 1 = 1/\bar{\delta} - 1$ and $\hat{\theta} = \hat{p} \cdot \hat{\theta}_* = \bar{\delta}/\bar{Z}$.

Note on the original parametrisation. In (θ, β) coordinates the two score equations $\partial \ln L/\partial\theta = 0$ and $\partial \ln L/\partial\beta = 0$ both reduce to $\theta = \bar{\delta}/[(\beta+1)\bar{Z}]$ and are thus linearly dependent; the system is underdetermined in (θ, β). The reparametrisation (θ_*, p) resolves this degeneracy. □

Proof of Lemma 2.1

Proof The sum $S_n = \sum Z_i \sim \Gamma(n, \lambda)$ with density $\lambda^n z^{n-1} e^{-\lambda z}/\Gamma(n)$, $z > 0$. Since $\bar{Z} = S_n/n$:

$$\mathbb{E}[1/\bar{Z}] = n\mathbb{E}[1/S_n] = \frac{n\lambda^n}{\Gamma(n)} \int_0^\infty z^{n-2} e^{-\lambda z} \, dz = \frac{n\lambda^n}{\Gamma(n)} \cdot \frac{\Gamma(n-1)}{\lambda^{n-1}} = \frac{n\lambda}{n-1}.$$

Similarly, for $n \geq 3$: $\mathbb{E}[1/\bar{Z}^2] = n^2\lambda^n \Gamma(n-2)/[\Gamma(n)\lambda^{n-2}] = n^2\lambda^2/[(n-1)(n-2)]$. For $n = 2$ the integral $\int_0^\infty z^{-1} e^{-\lambda z} \, dz$ diverges at 0^+. □

Proof of Lemma 2.2

Proof For $z > 0$ and the event $\{\delta_i = 1\} = \{X_i \leq Y_i\}$, with $X_i \sim \text{Exp}(\theta)$ and $Y_i \sim \text{Exp}(\beta\theta)$ independent:

$$\Pr(\delta_i = 1,\ Z_i > z) = \Pr(X_i \leq Y_i,\ X_i > z) = \int_z^\infty \theta e^{-\theta x} \int_x^\infty \beta\theta e^{-\beta\theta y} \, dy \, dx$$
$$= \int_z^\infty \theta e^{-\theta x} \cdot e^{-\beta\theta x} \, dx = \theta \int_z^\infty e^{-(\beta+1)\theta x} \, dx = \frac{e^{-\theta_* z}}{1+\beta}.$$

The marginals are $\Pr(\delta_i = 1) = 1/(1+\beta)$ and $\Pr(Z_i > z) = e^{-\theta_* z}$ (since $Z_i = \min(X_i, Y_i) \sim \text{Exp}(\theta_*)$). Therefore $\Pr(\delta_i = 1, Z_i > z) = \Pr(\delta_i = 1)\Pr(Z_i > z)$ for all $z > 0$, which proves $\delta_i \perp Z_i$ (Kalbfleisch and Prentice 2002, Theorem 2.1). □

Proof of Lemma 2.3

Proof Since $X_1, Y_1, \ldots, X_n, Y_n$ are mutually independent, $\delta_i = \mathbf{1}[X_i \leq Y_i]$ is a function of (X_i, Y_i) alone and $Z_j = \min(X_j, Y_j)$ of (X_j, Y_j) alone. Hence $\delta_i \perp Z_j$ for all i, j: for $i \neq j$ this is immediate from the independence of distinct pairs, and for $i = j$ it is Lemma 2.2. Consequently the σ-fields $\sigma(\delta_1, \ldots, \delta_n)$ and $\sigma(Z_1, \ldots, Z_n)$ are independent, and any Borel function of the first vector is independent of any Borel function of the second. Taking the two sample means gives $\bar{\delta} \perp \bar{Z}$. □

Proof of Theorem 2.5

Proof By Lemma 2.3, $\bar{\delta} \perp \bar{Z}$, hence also $\bar{\delta} \perp 1/\bar{Z}$ (as $1/\bar{Z} = g(\bar{Z})$ for the continuous function $g(v) = 1/v$; a Borel function of an independent random variable is independent of the other):

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}[\bar{\delta}] \cdot \mathbb{E}[1/\bar{Z}]. \tag{1}$$

We have $\mathbb{E}[\bar{\delta}] = 1/(1+\beta) = \theta/\theta_*$ and, by Lemma 2.1 with $\lambda = \theta_*$, $\mathbb{E}[1/\bar{Z}] = n\theta_*/(n-1)$. Hence $\mathbb{E}[\hat{\theta}] = (\theta/\theta_*) \cdot n\theta_*/(n-1) = n\theta/(n-1)$, proving (a).

For (b), independence gives $\mathrm{Var}[\hat{\theta}] = \mathbb{E}[\bar{\delta}^2]\mathbb{E}[1/\bar{Z}^2] - (\mathbb{E}[\hat{\theta}])^2$. Computing:

$$\mathbb{E}[\bar{\delta}^2] = \frac{1}{n(1+\beta)} + \frac{n-1}{n(1+\beta)^2} = \frac{\theta}{n\theta_*} + \frac{(n-1)\theta^2}{n\theta_*^2},$$

$$\mathbb{E}[1/\bar{Z}^2] = \frac{n^2\theta_*^2}{(n-1)(n-2)} \quad (n \geq 3).$$

Substituting and simplifying (bringing all terms to denominator $(n-1)^2(n-2)$: $n\theta\theta_*/[(n-1)(n-2)] = n(\beta+1)\theta^2/[(n-1)(n-2)]$, which splits as $n\beta\theta^2/[(n-1)(n-2)] + n\theta^2/[(n-1)(n-2)]$; $n^2\theta^2/[n(n-2)] - n^2\theta^2/(n-1)^2 = n\theta^2[(n-1)^2 - n(n-2)]/[n(n-2)(n-1)^2] = n\theta^2/[(n-2)(n-1)^2]$; and $n\theta^2/[(n-1)(n-2)] + n\theta^2/[(n-2)(n-1)^2] = n\theta^2 n/[(n-1)^2(n-2)] = n^2\theta^2/[(n-1)^2(n-2)]$):

$$\begin{aligned}
\mathrm{Var}[\hat{\theta}] &= \left(\frac{\theta}{n\theta_*} + \frac{(n-1)\theta^2}{n\theta_*^2}\right)\frac{n^2\theta_*^2}{(n-1)(n-2)} - \frac{n^2\theta^2}{(n-1)^2} \\
&= \frac{n\theta\theta_*}{(n-1)(n-2)} + \frac{n^2\theta^2}{n(n-2)} - \frac{n^2\theta^2}{(n-1)^2} \\
&= \frac{n\beta\theta^2}{(n-1)(n-2)} + \frac{n^2\theta^2}{(n-1)^2(n-2)},
\end{aligned}$$

where the last step uses $n\theta\theta_*/[(n-1)(n-2)] = n(\beta+1)\theta^2/[(n-1)(n-2)] = n\beta\theta^2/[(n-1)(n-2)] + n\theta^2/[(n-1)(n-2)]$ and $n\theta^2/[(n-1)(n-2)] + n\theta^2/[(n-2)(n-1)^2] = n^2\theta^2/[(n-1)^2(n-2)]$. This proves (3). As $n \to \infty$, expanding both terms of (3) to leading order, $n^2\theta^2/[(n-1)^2(n-2)] \sim \theta^2/n$ and $n\beta\theta^2/[(n-1)(n-2)] \sim \beta\theta^2/n$, so $\mathrm{Var}[\hat{\theta}] \sim (\beta+1)\theta^2/n$, exactly the asymptotic variance of Theorem 2.9 and the reciprocal Fisher information $[\mathcal{I}^{-1}]_{11}$ of Lemma 2.12; the exact finite-sample formula and the first-order asymptotic theory agree, as they must. Since $\mathrm{Var}[\hat{\theta}] \to 0$, Chebyshev's inequality with the full form $\mathbb{P}(|\hat{\theta} - \theta| \geq \varepsilon) \leq [\mathrm{Var}[\hat{\theta}] + (\mathbb{E}[\hat{\theta}] - \theta)^2]/\varepsilon^2 \to 0$ gives (c). □

Proof of Theorem 2.9

Proof Let $U_i = (\delta_i, Z_i)^\top$ with mean $\mu = (p, \mu_Z)^\top$, $p = 1/(1+\beta)$, $\mu_Z = 1/\theta_*$. Note that $Z_i = \min(X_i, Y_i) > 0$ a.s. (since X_i, Y_i are exponentially distributed), so $\bar{Z} > 0$ a.s. and $g(u,v) = u/v$ is well-defined on the range of $(\bar{\delta}, \bar{Z})$. The covariance matrix $\Sigma_U = \mathrm{Cov}(U_1)$ is finite since $\mathbb{E}[\delta_1^2] = p < \infty$ (δ_1 Bernoulli) and $\mathbb{E}[Z_1^2] = 2/\theta_*^2 < \infty$ (Z_1 exponential with $\theta_* = (\beta+1)\theta < \infty$). By the SLLN, $(\bar{\delta}, \bar{Z}) \xrightarrow{\text{a.s.}} (p, \mu_Z)$. By the CLT for i.i.d. vectors, $\sqrt{n}(\bar{U} - \mu) \rightsquigarrow N(0, \Sigma_U)$. Since $\delta_1 \perp Z_1$ (Theorem 4.7), $\mathrm{Cov}(\delta_1, Z_1) = \mathbb{E}[\delta_1 Z_1] - \mathbb{E}[\delta_1]\mathbb{E}[Z_1] = 0$, giving

$$\Sigma_U = \begin{pmatrix} p(1-p) & 0 \\ 0 & \theta_*^{-2} \end{pmatrix}.$$

The function $g(u,v) = u/v$ is continuously differentiable on $\mathbb{R} \times (0, \infty)$, hence in a neighbourhood of (p, μ_Z) since $\mu_Z = 1/\theta_* > 0$. By the delta method, $\sqrt{n}(g(\bar{U}) - g(\mu)) \rightsquigarrow N(0, \nabla g^\top \Sigma_U \nabla g)$, with gradient at (p, μ_Z): $\nabla g = (\mu_Z^{-1}, -p\mu_Z^{-2})^\top = (\theta_*, -p\theta_*^2)^\top$. Computing:

$$\sigma^2 = \nabla g^\top \Sigma_U \nabla g = \theta_*^2 p(1-p) + p^2\theta_*^4 \cdot \theta_*^{-2} = \theta_*^2 p[(1-p) + p] = \theta_*^2 p = \theta\theta_*.$$

□

Proof of Theorem 2.16

Proof *(a)* Let $U_i = (\delta_i, Z_i^\gamma)^\top$. Since $Z_i^\gamma \sim \mathrm{Exp}(\theta^\gamma(\beta+1))$ and $\delta_i \perp Z_i^\gamma$ (Theorem 4.7 with universal independence applied to $W_i = Z_i^\gamma$), $\mathbb{E}[U_i] = (p,\, 1/(\theta^\gamma(\beta+1)))^\top$. By the SLLN, $(\bar\delta, \overline{Z^\gamma}) \xrightarrow{\text{a.s.}} (p, p/\theta^\gamma)$. By the continuous mapping theorem, $\hat\theta_W = (\bar\delta/\overline{Z^\gamma})^{1/\gamma} \xrightarrow{\text{a.s.}} \theta$.

(b) Note $Z_i^\gamma \sim \mathrm{Exp}(\lambda)$ with $\lambda = \theta^\gamma(1+\beta)$, so $\mathrm{Var}(Z_i^\gamma) = 1/\lambda^2$. The covariance matrix of U_i is $\Sigma_U = \mathrm{diag}(p(1-p),\, p^2/\theta^{2\gamma})$ (zero off-diagonal by independence). Apply the delta method to $g(u,v) = (u/v)^{1/\gamma}$:

$$\nabla g\big|_{(p,\, p/\theta^\gamma)} = \left(\frac{\theta}{\gamma p},\, -\frac{\theta^{\gamma+1}}{\gamma p}\right)^\top.$$

Computing $\sigma_W^2 = \nabla g^\top \Sigma_U \nabla g$:

$$\sigma_W^2 = \frac{\theta^2}{\gamma^2 p^2}\cdot p(1-p) + \frac{\theta^{2\gamma+2}}{\gamma^2 p^2}\cdot\frac{p^2}{\theta^{2\gamma}} = \frac{\theta^2(1-p)}{\gamma^2 p} + \frac{\theta^2}{\gamma^2} = \frac{\theta^2}{\gamma^2 p} = \frac{(\beta+1)\theta^2}{\gamma^2}.$$

(c) We compute the Fisher information matrix of (θ,β) explicitly and invert it. Reparametrise through $\phi = \theta^\gamma$. Under the exponential PHM in the variable $W_i = Z_i^\gamma$ with rate ϕ, Lemma 2.12 gives the information matrix of (ϕ,β),

$$\mathcal{I}_{(\phi,\beta)} = \begin{pmatrix} 1/\phi^2 & 1/[\phi(\beta+1)] \\ 1/[\phi(\beta+1)] & 1/[\beta(\beta+1)] \end{pmatrix}.$$

The map $(\theta,\beta) \mapsto (\phi,\beta) = (\theta^\gamma,\beta)$ has Jacobian $J = \mathrm{diag}(\gamma\theta^{\gamma-1}, 1)$, and the information matrix transforms as $\mathcal{I}_{(\theta,\beta)} = J^\top \mathcal{I}_{(\phi,\beta)} J$. Writing $\phi = \theta^\gamma$ and simplifying,

$$\mathcal{I}_{(\theta,\beta)} = \begin{pmatrix} \gamma^2\theta^{2\gamma-2}/\phi^2 & \gamma\theta^{\gamma-1}/[\phi(\beta+1)] \\ \gamma\theta^{\gamma-1}/[\phi(\beta+1)] & 1/[\beta(\beta+1)] \end{pmatrix} = \begin{pmatrix} \gamma^2/\theta^2 & \gamma/[\theta(\beta+1)] \\ \gamma/[\theta(\beta+1)] & 1/[\beta(\beta+1)] \end{pmatrix}.$$

Its determinant is $\det \mathcal{I}_{(\theta,\beta)} = \gamma^2/[\theta^2\beta(\beta+1)^2]$, so

$$[\mathcal{I}^{-1}_{(\theta,\beta)}]_{11} = \frac{\mathcal{I}_{22}}{\det \mathcal{I}_{(\theta,\beta)}} = \frac{1/[\beta(\beta+1)]}{\gamma^2/[\theta^2\beta(\beta+1)^2]} = \frac{(\beta+1)\theta^2}{\gamma^2}.$$

Hence the Cramér–Rao bound for θ is $(\beta+1)\theta^2/(\gamma^2 n) = \sigma_W^2/n$, which the estimator attains by part (b). □

Proof of Lemma 2.12

Proof $\mathcal{I}_{11} = \mathbb{E}[s_\theta^2] = 1/\theta^2$ (verified in Section 2.3).

$\mathcal{I}_{22} = \mathbb{E}[s_\beta^2]$. Since $\delta \perp Z$: $\mathbb{E}[s_\beta^2] = \mathbb{E}[(1-\delta)^2]/\beta^2 - 2\theta\,\mathbb{E}[(1-\delta)]\mathbb{E}[Z]/\beta + \theta^2\mathbb{E}[Z^2]$. Using $\mathbb{E}[(1-\delta)^2] = \mathbb{E}[1-\delta] = \beta/(\beta+1)$ (Bernoulli), $\mathbb{E}[Z] = 1/\theta_*$, $\mathbb{E}[Z^2] = 2/\theta_*^2$, and $\theta/\theta_* = 1/(\beta+1)$:

$$\mathcal{I}_{22} = \frac{\beta/(\beta+1)}{\beta^2} - \frac{2\beta}{(\beta+1)\beta(\beta+1)} + \frac{2}{(\beta+1)^2} = \frac{1}{\beta(\beta+1)}.$$

$\mathcal{I}_{12} = \mathbb{E}[s_\theta\, s_\beta]$. Since $\mathbb{E}[s_\theta] = 0$ and $\delta \perp Z$:

$$\begin{aligned}
\mathcal{I}_{12} &= \mathbb{E}\big[(1/\theta - (\beta+1)Z)\,(1-\delta)/\beta\big] - \theta\,\mathbb{E}\big[(1/\theta - (\beta+1)Z)\,Z\big] \\
&= \underbrace{\mathbb{E}[s_\theta]}_{0}\cdot\frac{\mathbb{E}[1-\delta]}{\beta} - \theta\Big(\frac{\mathbb{E}[Z]}{\theta} - (\beta+1)\mathbb{E}[Z^2]\Big) \\
&= -\theta\Big(\frac{1}{\theta\theta_*} - (\beta+1)\cdot\frac{2}{\theta_*^2}\Big) = -\frac{1}{\theta_*} + \frac{2\theta_*}{\theta_*^2} = \frac{1}{\theta_*} = \frac{1}{\theta(\beta+1)}.
\end{aligned}$$

We stress that $\mathcal{I}_{12}$ is the expectation of the *entire* product $s_\theta s_\beta$. The contribution $\mathbb{E}[(\beta+1)Z\cdot\theta Z] = \theta(\beta+1)\mathbb{E}[Z^2] = 2/\theta_*$ is only one of the four terms in the expansion; the term $-\theta\,\mathbb{E}[Z]/\theta = -1/\theta_*$ (from $\mathbb{E}[s_\theta] = 0$ applied after expansion) reduces it to $1/\theta_*$. Retaining only the Z^2 term would give the incorrect value $2/\theta_*$. The determinant follows by direct calculation. □

2 Supplementary Note: Normal-model derivations

This appendix contains the elementary derivations for the three normal models referred to in Remark 4.6 ($N(\theta,1)$, $N(0,\theta^2)$, and $N(\theta_1,\theta_2^2)$). All results are standard and are included for completeness and self-containedness only.

Model $N(\theta, 1)$

Score: $\partial \ln f/\partial\theta = x-\theta$. Using (15)–(16): $\omega_\theta(t) = i\,\delta'(t) - \theta\,\delta(t)$. Substituting into (12) and applying **(P4)** with $\hat{\varphi}_n'(0) = i\bar{X}$, $\varphi_\theta'(0) = i\theta$: $(\bar{X}-\theta)=0$, giving $\hat{\theta}=\bar{X}$. Fisher information $I(\theta)=1$; CR bound $1/n = \mathrm{Var}[\bar{X}]$, so the estimator is efficient.

Model $N(0, \theta^2)$

Score: $\partial \ln f/\partial\theta = -1/\theta + x^2/\theta^3$. Using (15) and (17): $\omega_\theta(t) = -\theta^{-1}\delta(t) - \theta^{-3}\delta''(t)$. Applying **(P1)** and **(P4)**: $\theta^{-3}(\overline{X^2}-\theta^2)=0$, giving $\hat{\theta}^2 = \overline{X^2}$. $I(\theta) = 2/\theta^2$; CR bound for $\tau(\theta)=\theta^2$ is $2\theta^4/n$; $\mathrm{Var}[\overline{X^2}] = (3\theta^4-\theta^4)/n = 2\theta^4/n$. Efficient.

Model $N(\theta_1, \theta_2^2)$

For θ_1: same as $N(\theta,1)$ up to the factor θ_2^{-2}, giving $\hat{\theta}_1 = \bar{X}$.

For θ_2: score $-1/\theta_2 + (x-\theta_1)^2/\theta_2^3$. Expanding $(x-\theta_1)^2$ and applying (15)–(17):

$$\omega_{\theta_2}(t) = -\theta_2^{-3}\delta''(t) - 2i\theta_1\theta_2^{-3}\delta'(t) + (\theta_1^2-\theta_2^2)\theta_2^{-3}\delta(t).$$

Substituting into (12) and using $\hat{\theta}_1=\bar{X}$: $\hat{\theta}_2^2 = \overline{X^2} - \bar{X}^2$.

$I_{11} = \theta_2^{-2}$, $I_{22} = 2\theta_2^{-2}$, $I_{12}=0$. $\hat{\theta}_1 = \bar{X}$ is efficient. $S^2 = \frac{n}{n-1}(\overline{X^2}-\bar{X}^2)$ is UMVUE by the Lehmann–Scheffé theorem (Borovkov 1984) and satisfies $\mathrm{Var}[S^2] = 2\theta_2^4/(n-1) > 2\theta_2^4/n$, so the CR bound $2\theta_2^4/n$ is not attained at finite n. The equality condition in the CR inequality (Borovkov 1984, Th. 2.5.1) requires the score $\partial \ln f/\partial\theta_2^2$ to be affine in T a.s.; this fails because $(n-1)S^2/\theta_2^2 \sim \chi^2_{n-1}$ is based on deviations from the estimated mean $\bar{X}$, not the true θ_1.

3 Supplementary Note: Cauchy location family derivations

This appendix collects the derivations for the Cauchy example of Section 4.4: the weight function (25), the reduction of the estimating equation to (26), the Fisher information, and the verification of the regularity conditions (C1)–(C4) of Lemma 4.14.

Finite-sample comparison with maximum likelihood.

Table 1 makes the comparison a fair one, as the reviewer rightly demands. When maximum likelihood is computed *globally*—by enumerating every root of the likelihood equation and retaining the one of largest likelihood—its RMSE is indistinguishable from that of the ECF estimator at every sample size (e.g. 0.339 versus 0.339 at $n=20$), and both sit at the Cramér–Rao limit $\sqrt{2/n}$. This is as it must be: the ECF

Table 1 Supplementary Table S1. Cauchy location model ($\theta = 0$): RMSE of the ECF estimator, of the *global* maximum-likelihood estimator (all likelihood modes enumerated and the largest selected), and of ML with three random restarts, together with the secondary-mode-capture rate of the three-restart procedure. "CR SD" $= \sqrt{2/n}$ is the asymptotic efficiency limit for all three. Global-ML and three-restart figures are from the present enumeration study ($R = 5\,000$ for $n \le 100$, $R = 3\,000$ for $n = 500$); Monte Carlo standard errors of RMSE, RMSE$/\sqrt{2R}$, are at most 0.003 throughout. "Multi-mode" is the fraction of replications where the three-restart ML converged to a mode with $|\hat{\theta}_{\mathrm{ML}} - \hat{\theta}_{\mathrm{ECF}}| > 0.5$.

n	CR SD	ECF	global ML	ML (3 restarts)	Multi-mode (%)
20	0.316	0.339	0.339	1.88	55.9
50	0.200	0.207	0.207	2.71	65.0
100	0.141	0.147	0.147	2.55	74.2
500	0.063	0.064	0.064	—	—

estimating equation for the Cauchy location family *is* the likelihood score equation (Lemma 4.14(C4)), so the two estimators share the same set of stationary points and, once the global optimum is identified, the same value. The ECF estimator carries no efficiency advantage over correctly computed maximum likelihood, and we claim none.

The difference the table records is therefore not one of statistical quality but of *numerical robustness of a default solver*. A naive multi-start ML with a small, fixed number of restarts is not guaranteed to locate the global optimum for the Cauchy likelihood, whose expected number of local maxima grows with n (Barnett 1966; Reeds 1985); with three restarts it captures a secondary mode in a majority of replications at these sample sizes, inflating its RMSE far above the CR limit. Increasing the restart budget removes the effect—with enough restarts the multi-start RMSE returns to the global-ML value—at a cost that grows linearly in the number of restarts and must itself be tuned. The ECF estimating equation, by contrast, is solved by a single one-dimensional root search over the enumerated sign changes of $\sum_k (x_k-\theta)/[1+(x_k-\theta)^2]$, with no restart budget to choose. The practical advantage of the ECF route is thus the absence of a multi-start heuristic and its tuning, not any superiority in estimation error; the two procedures are statistically equivalent and differ only in the reliability of the off-the-shelf optimiser.

Weight function

The function $x \mapsto 2x/(1+x^2)$ belongs to $L^2(\mathbb{R})$ (since $4x^2/(1+x^2)^2 \le 4/x^2$ is integrable at $\pm\infty$) but *not* to $L^1(\mathbb{R})$ (since $2|x|/(1+x^2) \sim 2/|x|$ and $\int_1^\infty 2/x\,dx = \infty$); its Fourier transform is therefore taken in the L^2 (Plancherel) sense. Let $h(x) = 2/(1+x^2) \in L^1(\mathbb{R})$; then $\widehat{h}(t) = 2\pi e^{-|t|}$ (standard residue calculation (Rudin 1991, Ch. 4)). Since $2x/(1+x^2) = x \cdot h(x) \in L^2(\mathbb{R})$, the L^2 Fourier transform satisfies $\widehat{xh}(t) = i\,\widehat{h}'(t)$ (valid because $h \in L^1$ implies $\widehat{h} \in C^\infty$, so $i(\widehat{h})'$ lies in L^2 and the

identity holds by differentiation under the L^2-integral sign). Computing,

$$\widehat{\left(\frac{2x}{1+x^2}\right)}(t) = i\,\frac{d}{dt}\big(2\pi e^{-|t|}\big) = -2\pi i\,\operatorname{sgn}(t)\,e^{-|t|},$$

where $d/dt[e^{-|t|}] = -\operatorname{sgn}(t)e^{-|t|}$ holds in L^1_{loc} (hence distributionally) since $e^{-|t|}$ is absolutely continuous. Therefore

$$\omega_\theta(t) = \frac{1}{2\pi}\,e^{-it\theta}\,(-2\pi i\,\operatorname{sgn} t)\,e^{-|t|} = e^{-it\theta}\,(-i\,\operatorname{sgn} t)\,e^{-|t|},$$

which is (25), with $|\omega_\theta(t)| = e^{-|t|} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

Reduction of the estimating equation

Substituting ω_θ into (12),

$$\int_{-\infty}^{+\infty} e^{-it\theta}(-i\,\operatorname{sgn} t)\,e^{-|t|}\big[\hat\varphi_n(t) - e^{it\theta-|t|}\big]\,dt = 0.$$

The model-CF term is $\int_{-\infty}^{+\infty}(-i\operatorname{sgn} t)e^{-2|t|}\,dt$, which vanishes by oddness:

$$\int_{-\infty}^{+\infty}(-i\operatorname{sgn} t)\,e^{-2|t|}\,dt = -i\left(\int_{-\infty}^{0}(-1)e^{2t}\,dt + \int_0^{\infty} e^{-2t}\,dt\right) = -i\big(-\tfrac12 + \tfrac12\big) = 0.$$

The remaining ECF term, after collecting the $t<0$ and $t>0$ contributions, is the imaginary part in (26).

Fisher information

$$I(\theta) = \int_{-\infty}^{\infty}\frac{[\partial_\theta f(x;\theta)]^2}{f(x;\theta)}\,dx = \frac{4}{\pi}\int_{-\infty}^{\infty}\frac{x^2}{(1+x^2)^3}\,dx = \frac{4}{\pi}\cdot\frac{\pi}{8} = \frac12,$$

where $\int_0^\infty x^2/(1+x^2)^3\,dx = B(3/2,3/2)/2 = \pi/16$ via $x=\tan u$.

Regularity conditions

We verify conditions (C1)–(C4) for the weight $\omega_\theta(t) = (-i\,\operatorname{sgn} t)\,e^{-|t|}e^{-it\theta}$ of the Cauchy location family.

(C1): $|\omega_\theta(t)| = e^{-|t|}$, which is in $L^1 \cap L^2$.

(C2): The Cauchy characteristic function is injective in θ since $e^{i(\theta-\theta')t} = e^{-|t|[e^{|t|}-1]}\cdot c(t)$ for $\theta\neq\theta'$ cannot hold a.e.

(C3): Oddness of $\operatorname{sgn}(t)e^{-2|t|}$ gives $\int\omega_\theta(t)e^{it\theta-|t|}\,dt = 0$. The derivative identity follows from $-\partial_\theta\int\omega_\theta\varphi_\theta\,dt = \int|\omega_\theta(t)|^2\,dt/\nu$ for the score weight; more directly, differentiating (12) at the true θ with respect to θ yields $-I_\omega(\theta)$ where $I_\omega = \int|\omega_\theta(t)|^2/s_\theta(t)\,\nu(dt)$ is the information content of the weight; for the score weight this equals $I(\theta) = 1/2 > 0$.

(C4): Compute $\psi(x;\theta) = \int_0^\infty e^{-t}\sin(t(x-\theta))\,dt$ by the standard Laplace transform $\int_0^\infty e^{-t}\sin(at)\,dt = a/(1+a^2)$, so $\psi(x;\theta) = (x-\theta)/[1+(x-\theta)^2]$. This is the Cauchy score (up to the factor $\pi/2$), which is bounded, continuously differentiable in θ, and satisfies $\mathbb{E}_\theta[\psi(X;\theta)] = 0$. van der Vaart (1998, Thm. 5.9) gives consistency and Thm. 5.21 asymptotic normality with variance $\mathbb{E}[\psi^2]/(\mathbb{E}[\partial_\theta\psi])^2 = I(\theta)^{-1} = 2$. □

4 Supplementary Note: Proof of the score-weight equivalence

We prove that, for the score weight $\omega_\theta(t) = \lim_{\varepsilon\downarrow 0}\frac{1}{2\pi}\int_S s(x;\theta)\,e^{-itx-\varepsilon x^2}\,dx$,

$$\int_{-\infty}^{+\infty}\omega_\theta(t)\left[\hat\varphi_n(t)-\varphi_\theta(t)\right]dt = \frac{1}{n}\sum_{k=1}^n s(X_k;\theta).$$

The point is to integrate over t *before* attempting to write $\omega_\theta(t)$ in closed form, so that only the ordinary, full-line identity $\frac{1}{2\pi}\int_{\mathbb{R}} e^{-it(x-y)}\,dt = \delta(x-y)$—never a one-sided transform of 1 or x—is used. For the ECF term, $\hat\varphi_n(t) = n^{-1}\sum_k e^{itX_k}$ is a *finite* sum, so exchanging it with the t-integral is immediate (linearity, no limiting argument needed):

$$\begin{aligned}\int_{\mathbb{R}}\omega_\theta(t)\hat\varphi_n(t)\,dt &= \frac{1}{n}\sum_{k=1}^n\int_{\mathbb{R}}\omega_\theta(t)\,e^{itX_k}\,dt\\ &= \frac{1}{n}\sum_{k=1}^n\int_S s(x;\theta)\Big[\tfrac{1}{2\pi}\int_{\mathbb{R}}e^{-it(x-X_k)}\,dt\Big]dx = \frac{1}{n}\sum_{k=1}^n s(X_k;\theta),\end{aligned}$$

the inner bracket being $\delta(x-X_k)$ and the sifting valid since $X_k \in S$ almost surely. For the model term, dominated convergence under $\mathbb{E}_\theta|s(X;\theta)| < \infty$ justifies the same exchange, giving $\int_{\mathbb{R}}\omega_\theta(t)\varphi_\theta(t)\,dt = \int_S s(x;\theta)f(x;\theta)\,dx = \mathbb{E}_\theta[s(X;\theta)]$. Subtracting proves the identity; $\mathbb{E}_\theta[s(X;\theta)] = 0$ by the standard score identity, giving the stated form. □

A fully elementary check with no distributional language at all is available by regularising with $e^{-\varepsilon x}$ instead and evaluating both t-integrals by contour integration for $\varepsilon > 0$, then letting $\varepsilon \to 0^+$; this is carried out explicitly for the exponential score in Corollary 4.9 and re-verified numerically in the supplementary code.

Singular terms in the one-sided transform

A literal one-sided Fourier transform of the constant and linear functions on $[0,\infty)$ is not a multiple of $\delta(t)$ and $\delta'(t)$: regularising $\int_0^\infty e^{-itx-\varepsilon x}\,dx = 1/(\varepsilon+it)$ and letting $\varepsilon\to 0^+$ gives, by the Sokhotski–Plemelj formula, $\pi\delta(t) - i\,\mathrm{PV}(1/t)$, and differentiating introduces a further finite-part $1/t^2$ term. If one writes $\omega_\theta(t)$ down explicitly as a distribution in t first, these extra singular terms are present and must be tracked. Proposition 4.3 shows they never need to be tracked, because the estimating equation

uses ω_θ only inside the outer t-integral against $\hat{\varphi}_n(t) - \varphi_\theta(t)$; performing the x- and t-integrations in the order used in the proof—data point (or density) first, delta-sifting in t last—collapses the whole construction to the ordinary score equation, the extra terms contributing nothing because they are never formed.

Contour-integration verification for the exponential model

Both terms of Proposition 4.3 may be verified for the exponential PHM by exact contour integration of the ε-regularised integrand. Writing

$$\omega_\theta^\varepsilon(t) = \frac{1}{2\pi}\Big[\frac{1}{\theta}\cdot\frac{1}{\varepsilon + it} - (\beta+1)\cdot\frac{1}{(\varepsilon+it)^2}\Big]$$

for the regularised weight and closing the contour in the upper half-plane (where the only pole of $\omega_\theta^\varepsilon$ lies, at $t = i\varepsilon$),

$$\int_{\mathbb{R}} \omega_\theta^\varepsilon(t)\, e^{itZ_k}\, dt = \Big(\tfrac{1}{\theta} - (\beta+1)Z_k\Big)e^{-\varepsilon Z_k} \xrightarrow[\varepsilon\to 0^+]{} \tfrac{1}{\theta} - (\beta+1)Z_k,$$

$$\int_{\mathbb{R}} \omega_\theta^\varepsilon(t)\, \varphi_\theta(t)\, dt = \frac{\theta_*(\varepsilon - (\beta+1)\theta + \theta_*)}{\theta(\varepsilon+\theta_*)^2} \xrightarrow[\varepsilon\to 0^+]{} \frac{\theta_* - (\beta+1)\theta}{\theta\,\theta_*} = 0,$$

confirming (18) for this model term by term, with no principal-value or finite-part contribution surviving.