Robust Rank-Based IV Duration Inference in Stata: The aaaft Command and an Application to Health Data

Marcel Voia
(with Leonardo Francisco Sanchez Aragon and Kusum Menon)
STATA Conference, Ottawa, 2025
Based on: **Robust Duration Inference and Bias in Allocating Critical Hospital Resources**, by Anand Acharya, Lynda Khalaf, Kusum Menon, Marcel Voia, Myra Yazbeck and David Wensley

October 3, 2025

Outline

Motivation

- Resource allocation in healthcare often uses severity scores (e.g., PRISM, PIM) to prioritize care.
- Conventional duration models (AFT, Cox) can be biased by endogeneity, weak instruments, and non-normal errors.
- **Goal:** Introduce a robust, simulation-based rank IV estimator for treatment effects in duration models, implemented in Stata as `aaaft`.

Empirical Question

- Do severity scores perpetuate bias in triage?
- How robust is the proposed estimator to endogeneity and weak instruments?
- What is the effect size of severity on ICU duration?

Regional Health Utilization Model

Population Structure

- $g = 1, \ldots, G$ geographic regions.
- $i = 1, \ldots, N$ children per region.
- Each child has unobserved health ability $\tilde{\nu}_{ig} \in \mathbb{R}$; higher $\tilde{\nu}_{ig}$ means lower health ability.

Severe health events:

- Probability of ICU admission: $\mathbb{P}(a_{ig} = 1|\tilde{\nu}_{ig})$
- Severity score on ICU entry: $s_{ig}(\tilde{\nu}_{ig})$
- Monotonicity: $s'_{ig}(\tilde{\nu}_{ig}) > 0$, $s''_{ig}(\tilde{\nu}_{ig}) \leq 0$

Accidental Events & Patient Demand

Accidental health events:

- Event probability is independent of ability: $\mathbb{P}(a_{ig} = 1 | \tilde{\nu}_{ig}) = \mathbb{P}(a_{ig} = 1)$
- Example: children as car accident passengers.

Patient demand

- ICU duration: $y_{ig} \equiv \Lambda(t_{ig})$
- Optimal care benefit:

$$\mathcal{U}(y_{ig} | s_{ig}, x_{ig}, \tilde{\nu}_{ig}) = -\frac{1}{2}(y_{ig} - s_{ig}(\tilde{\nu}_{ig}))^2 + (\tilde{\nu}_{ig} + x_{ig})y_{ig}$$

- Efficiency: Match care to severity/exposure and health ability.

Efficient Care & Physician Supply

Efficiency conditions

- Parsimonious feature: match intervention to need.
- Patient-level efficiency: $y_{ig} = s_{ig}(\tilde{\nu}_{ig})$
- Optimal care:

$$y^*_{ig} = s_{ig}(\tilde{\nu}_{ig}) + x_{ig} + \tilde{\nu}_{ig}$$

Physician supply

- Team chooses y_{ig} to maximize perceived benefit:

$$\tilde{\mathcal{U}}(y_{ig}|s_{ig}, x_{ig}, \tilde{\nu}_{ig}) = \mathcal{U}(y_{ig}|s_{ig}, x_{ig}, \tilde{\nu}_{ig}) + \lambda_g(\tilde{\nu}_{ig})y_{ig}$$

- $\lambda_g(\tilde{\nu}_{ig})$: hospital/regional differences.

Institutional Constraints & Optimal Care

Institutional constraints

- Resource constraints: $C_g(\tilde{\nu}_{ig})$
- Team solves for quantity:

$$y_{ig}^* = \arg\max_y \left(\tilde{\mathcal{U}}(y_{ig}|s_{ig}, x_{ig}, \tilde{\nu}_{ig}) - C_g(\tilde{\nu}_{ig})y_{ig}\right)$$

Final optimal care formula

- $$y_{ig}^* = s_{ig}(\tilde{\nu}_{ig}) + x_{ig} + \tilde{\nu}_{ig} + \lambda_g(\tilde{\nu}_{ig}) - C_g(\tilde{\nu}_{ig})$$
- Confounding arises through unobserved $\tilde{\nu}_{ig}$ in both demand and supply.

Econometric Specification & Identification

AFT Duration model

- Core relation:

$$\Lambda(t_{ig}) = \beta s_{ig} + \delta X_{ig} + \epsilon_{ig}$$

- Controls: X_{ig} includes hospital effects, observed patient controls.
- Non-i.i.d. errors ϵ_{ig}, multi-level heterogeneity.

Challenge: Endogeneity due to unobserved health ability.
Instrumental variable strategy

Identification: Trauma as an Instrument

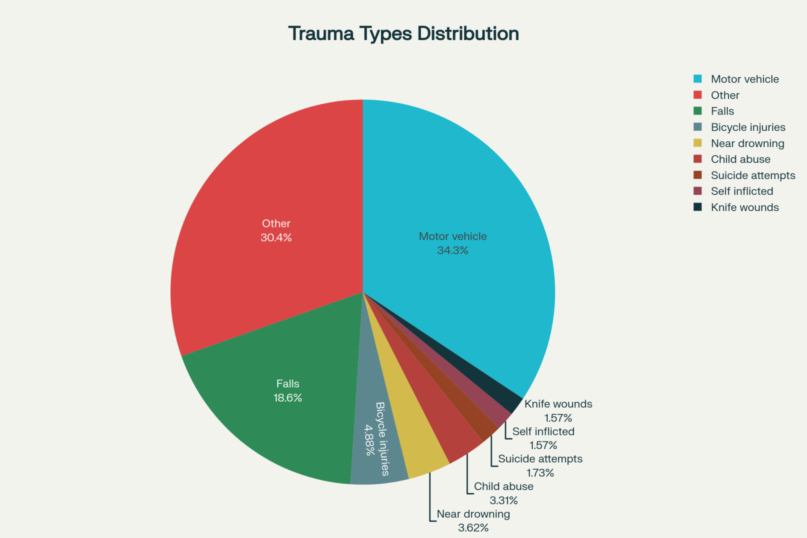
Solution: Use an *identification-robust* IV approach.

- Instrument: Trauma status $z_{ig} \in \{0, 1\}$ (accidental ICU admissions)
- Trauma events are presumed exogenous—admission due to trauma is "as if random" with respect to underlying health ability.
- **No-selection assumption:**

$$\mathbb{P}(a_{ig} = 1 | \nu_{ig}, \sigma_g \eta_{ig}) = \mathbb{P}(a_{ig} = 1)$$

- This ensures trauma instrument is not correlated with unobserved confounders.

Trauma type injuries



Rank-Based Quadratic Inference: Matrix Formulation

Inference via ranks:

- Quadratic inference function:

$$\mathcal{G}(\beta_0) = c(\beta_0)' \, p_z \, c(\beta_0)$$

- $p_z = z(z'z)^{-1}z'$ is the projection matrix of the instrument (trauma status).
- $c(\beta_0)$ is a rank-preserving score based on residuals.
- Inverting $\mathcal{G}(\beta_0)$ yields:
  - Hodges-Lehmann-Sen estimator for β
  - Size-controlled confidence set for severity coefficient

Distribution-Free Rank Test & Scoring
Rank scores:

- Quantile score: $c^{(i)} = F_o^{-1}\left(\frac{(i)}{n+1}\right)$
- Expected value score: $c^{*(i)} = E_{F_o}[V^{(i)}]$
- Empirical/simulated quantile scores are used for exact, distribution-free confidence sets for β.

aaaft_rils: Rank-based IV Inference

Title

`aaaft_rils` – Rank-based inference for treatment effects under Logistic, Gumbel, and Empirical distributions

Syntax

```
aaaft_rils depvar treatmentvar instrumentvar covariates
```

where

- `depvar`: outcome variable (e.g., log length of stay: `lnlosh`)
- `treatmentvar`: predicted risk score (e.g., `PRISM`, `PIM`)
- `instrumentvar`: instrumental variable (e.g., `Trauma`)
- `covariates`: one or more control variables (clinical and demographic)

Description and Transformations

Description

`aaaft_rils` estimates treatment effects by simulation-based rank inference, robust to nonlinearity and free from classic distributional assumptions.

The procedure tests a range of β values and computes a rank-based quadratic inference function (QIF) with three transformations:

- Logistic (logit) transformation of residual ranks
- Gumbel (Extreme Value Type I) transformation
- Empirical (nonparametric transformation based on $-\ln(1-p)$)

Procedure and Outputs

How it works

1. Mean-center all covariates and instrument
2. Project residuals orthogonally to controls across a grid of β
3. Apply rank transformation to each residual vector
4. Compute RAR statistic for each transformation
5. Select β minimizing RAR as point estimate
6. Simulate quantile scores (1000 reps) to build confidence intervals

Outputs:

- Estimates and confidence intervals for β (Logistic, Gumbel, Empirical)
- 95th percentiles of statistics
- Results stored in `r()`

Mata Integration and Requirements

- **Key Functions:**
  - `rankinf()`: Computes aligned residuals, rank scores, RAR statistics, selects β minimizing each transformation.
  - `leastsquares()`: Simulates quantile scores, computes test statistics, derives confidence intervals.
- **Required Package:** `moremata` (for `mm_ranks()`)
- **Install:** `ssc install moremata`

Monte Carlo Simulation

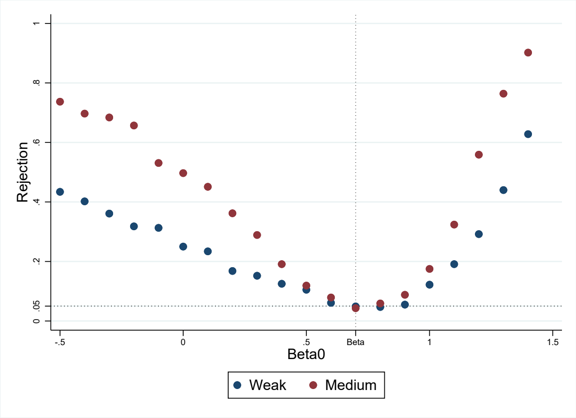
To evaluate the estimator's performance in the duration context, we draw a sample of $i = 200$ individuals across $g = 4$ hospitals where $n_1 \neq .. \neq n_g$ with the data generating process,

$$y_{ig} = \beta s_{ig} + \delta X_{ig} + \nu_{ig} + \eta_{ig}, \tag{1}$$

$$s_{ig} = \pi Z_{ig} + \sqrt{(1-\rho^2)} u_{ig} + \rho(\eta_{ig} + \nu_{ig}), \tag{2}$$

where ν_{ig} is a log transformation drawn from the Gamma distribution. The baseline duration, η_{ig} is generated with the extreme value quantile function, $\mu_g - \sigma_g ln(-ln(p))$, with hospital specific location μ_g and scale σ_g parameters. u_{ig} is Normal. The instrument and other covariates are drawn as binary or counts in similar proportions to the data. Below are the rejection rates with $\pi = .4$ and $\pi = .8$ with $\rho = .75$.

Figure: Monte Carlo Simulation of Rejection Rates for using a weak/medium instrument, moderate endogeneity and mild censoring



Note: This figure is a graph of rejection rates for our Monte-Carlos simulations with weak/medium instrument $\pi = .4, .8$, moderate endogeneity $\rho = .75$, and mild censoring.

Empirical Application: Health Data

- Data: 10,044 pediatric ICU admissions from 6 Canadian hospitals.
- Outcome: Log length of stay (`lnlosh`).
- Treatment: Severity scores (`PRISM`, `PIM`).
- Instrument: Trauma status.
- Covariates: Hospital dummies, age, chronic conditions, previous ICU, cardiac diagnosis.

Summary Statistics

Description (Variable)	Mean	Std. Dev.	Min	Max
Hospitals				
Hospital 1	0.142	0.349	0	1
Hospital 2	0.052	0.222	0	1
Hospital 3	0.333	0.471	0	1
Hospital 4	0.090	0.286	0	1
Hospital 5	0.293	0.455	0	1
Hospital 6	0.090	0.286	0	1
Age Categories				
Neonates $<$ 1 month old	.109	.312	0	1
Infants 1-12 months old	.2583	.438	0	1
Children 12-144 months old	.403	.491	0	1
Adolescents $>$ 144 months old	.228	.419	0	1
Length of stay in hospital (days)	117.99	340.79	1	17,541
Log-transformed length of stay	3.92	1.18	0	9.77
Risk Scores				
Pediatric Risk of Mortality score **(PRISM)**	5.10	5.66	0	49
Pediatric Index of Mortality (logit) **(PIM)**	-4.53	1.52	-8.41	5.30
Condition at Admission				
Trauma admission	0.066	0.247	0	1
Chronic conditions	0.48	0.74	0	1
Previous ICU admission	0.074	0.262	0	1
Cardiac diagnosis	0.183	0.387	0	1

Non-Trauma vs Trauma Durations

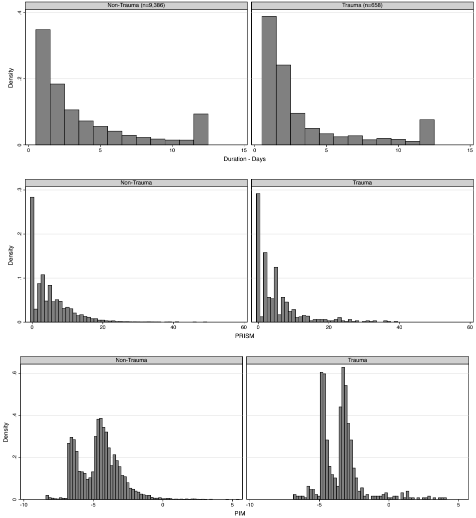
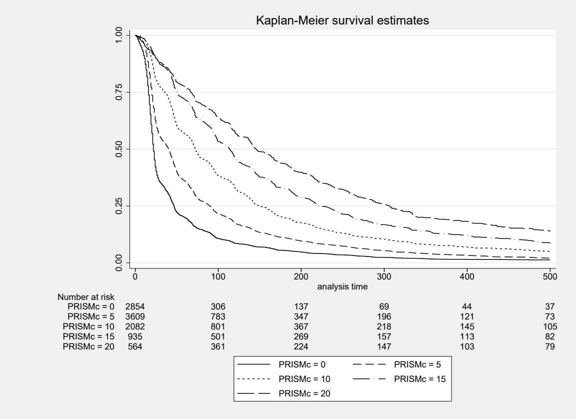
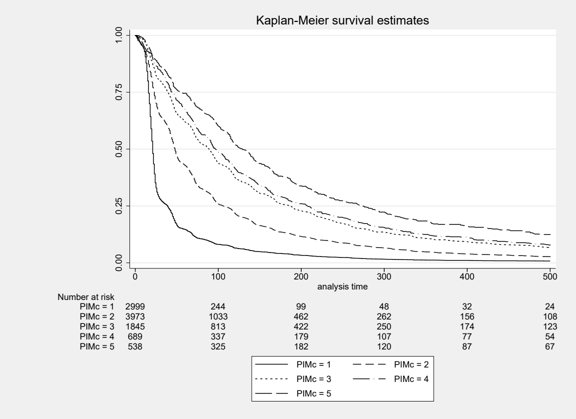


Figure: Stratified Kaplan-Meier function - PRISM



Note: The Kaplan-Meier functions monotonically shift outward from the origin for increasing PRISM groupings. Proportional shifts on the time (horizontal) scale is consistent with an accelerated failure time metric. Numbers below the figure detail the respective risk sets for each PRISM grouping (0=0, 1–5=5, 6–10=10, 11–15=15, and $>$16=20).

Figure: Stratified Kaplan-Meier function - PIM



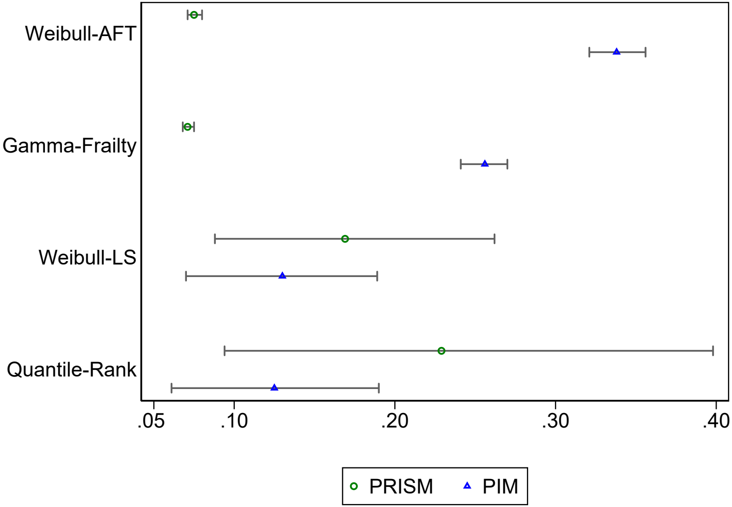
Note: The Kaplan-Meier functions monotonically shift outward from the origin for increasing PIM groupings. Proportional shifts on the time (horizontal) scale is consistent with an accelerated failure time metric. Numbers below the figure detail the respective risk sets for each PIM grouping (0–.5=1, .5–2.5=2, 2.5–5=3, 5–10=4, and $>10=5$).

Duration analysis of severity scores PRISM and PIM

Analysis	PRISM	PIM
Reference analysis		
Weibull – Likelihood	(0.071, **0.075**, 0.080)	(0.321, **0.338**, 0.356)
Gamma – frailty	(0.068, **0.071**, 0.075)	(0.241, **0.256**, 0.270)
Identification Robust IV		
Weibull – Least Squares	(0.088, **0.169**, 0.262)	(0.070, **0.130**, 0.189)
Empirical Quantile – Rank	(0.094, **0.229**, 0.398)	(0.061, **0.125**, 0.190)

Note: This table reports the estimates of β for four accelerated failure time ICU length of stay analysis of equation (10) – PRISM column two, and equation (11) – PIM column three. Outer values are 95% confidence sets and the center value in bold-blue is either the point estimate or Hodges-Lehman-Sen estimator.

95% confidence sets for PRISM and PIM



Results: Key Findings

- PRISM (physiology-based) and PIM (includes treatment/diagnosis) yield different effect sizes for ICU duration.
- Rank IV estimator robust to weak instruments, non-normal errors, and multi-level heterogeneity.
- Including diagnoses/treatments in scores can perpetuate bias; robust inference helps identify and correct for this.

Practical Takeaways

- `aaaft` provides robust, simulation-based inference for duration models with endogenous treatments.
- Easy integration with health data and standard Stata workflows.
- Use rank IV methods to address bias in resource allocation studies, especially when instruments are weak or data is non-normal.