Two-Stage Estimation Pipeline

STAGE 1: Feature Engineering (Phase 2)

Raw Transaction Data

SQLite Database (15M transactions)

- tx_id, found_at, mined_at
- waittime, min_respend_blocks
- mempool_size, mempool_tx_count

1. Create Epochs

30-min windows

2. Compute ρ_t

Mempool Congestion

3. Compute Time Cost

Impatience Proxy

4. Train W_hat Model

Random Forest

5. Compute Quantiles

F_tq (Tail Dist)

Phase 2 Features

W_hat, ρ_t, time_cost, F_tq

Merge Phase 2 Features

STAGE 2: Fee Prediction (Phase 3)

Enhanced Dataset

Raw Data + Phase 2 Features

- Exchange Flags (From/To)
- Blockspace Utilization
- Transaction Value (V_it)

1. Structural Features

Log transforms, interactions

2. Compute Riemann Sum

Competition term

3. Train Fee Models

OLS, Quantile, Gamma GLM, Spline

Fee Predictions & Analysis

Structural Coefficients, Diagnostics