Online Resource 1

Supplementary material for the article

Routing Frictions and Executable Liquidity in

Fragmented Markets

Digital Finance

Wen-Ting Wang

Institute of Statistics, National Chung Hsing University, Taichung, Taiwan

Corresponding author. Email: egpivo@gmail.com

1 Exact Multi-Family Measurement

The headline opportunity tape contains 13,768 family-transactions from 29 prespecified token-pair families and 71 sibling pools: 7,176 transactions in 15 Ethereum families and 6,592 transactions in 14 Base families. Family selection, deterministic hash inclusion probabilities, and transaction weights were frozen before gross gain, net gain, realized route, convergence, or markout outcomes were constructed. The WETH–USDC benchmark motivated and validated early stages of the design but does not supply the 29-family headline estimands.

The execution engine reconstructs transaction-ordered pool pre-states using contract integer arithmetic, exact fee rounding, and complete initialized-tick histories. It permits partial-tick execution and repeated tick crossings. When exact point identification fails, the observation enters deterministic lower and upper bounds.

The opportunity tape contains 12,817 point-identified observations, 917 economic no-sibling-overlap zeros, and 34 bounded overlap failures. Exact or explicitly bounded lower and upper outputs are aggregated without silently dropping deterministic failures. Independent allocation checks preserve the gross-complementarity classification, including states with more than two executable pools.

2 Route Census and Nested Route Universes

Route classification does not infer an economic objective from Swap-log count alone. It combines decoded calldata, receipt Swap and Transfer logs, compact call-tree traces, pool-call sequence, token-flow graphs, and transfer reconciliation. The final objective classes distinguish clean sibling splits, aggregator execution, arbitrage or backruns, liquidation or protocol operations, and complex multi-objective transactions.

All 250 observed integrated-route rows—177 on Ethereum and 73 on Base—were reviewed using normalized receipt, trace, and token-flow evidence. The review found no unresolved label and made no manual label change. This establishes the narrowest public-execution class supported by the recorded evidence; it does not recover private trader intent.

Population route rates use the full 13,768-transaction tape. Actual multi-pool integration is the broad extensive-margin object. Broad clean sibling integration is nested within that object, and strict verified integration is nested within broad clean integration. Route efficiency uses a different conditional universe: 170 transfer-reconciled, point-identified actual integrated routes with feasible net gain above the frozen denominator floor. The clean-route rate is therefore not used as the denominator for the all-objective efficiency result.

3 Dependence-Adjusted Inference

The primary procedure resamples families within chain and circular contiguous eight-week blocks within sampled family. It preserves all transactions and inverse-inclusion weights and propagates gas-coefficient, ETH/USD, gas-price, Base calldata/L1, and deterministic execution-bound uncertainty. Each chain and the equal-chain estimand have 5,000 valid draws. Four- and thirteen-week blocks, one- to three-month blocks, and the stationary calendar-week bootstrap are sensitivities. Transactions and calendar months are not independent headline observations.

Population route-rate and eligible-efficiency intervals retain family, calendar-date, exact-block, and family-block identifiers. Sparse objective cells follow the frozen small-family rules: five to nine families require wild-family inference, and cells with fewer than five families remain descriptive.

4 Failed Clean-Route Signs and Pooled Markouts

Dependence adjustment materially reduces the effective support of the clean-route outcome tests. Ethereum clean-route raw-dispersion changes are -11.51 basis points at one block and -7.44 basis points at five blocks, but their six-family wild intervals cross zero. Base point estimates are positive, but only three families contribute, so they remain descriptive. Pooled clean-route family-relative markout point estimates are negative at one and five blocks; their nine-family wild intervals also cross zero.

Pooled all-objective family-relative markouts are negative at one and five blocks under the two-way family/calendar-date intervals. This is a separate descriptive realized-outcome fact. It does not replace the failed clean-route estimand and does not identify a causal market-quality effect.

5 Cross-Chain Routing-Cost and Exact-State Audit

The support design was fixed before interpreting the counterfactual matrix. The initial exact-stratum screen retained only eight Ethereum and 74 Base rows from two families per chain. Its effective sample sizes were 6.28 and 15.34, and its maximum normalized weights were 20.95% and 14.21%. That screen fails the headline support gate.

The permitted repair is an outcome-blind propensity-overlap population using pair type, economic direction, quarter, log USD order size, log active virtual liquidity, sibling dominance, executable sibling count, pre-state volatility, and missingness indicators. It retains all 29 families. The largest absolute standardized difference is 0.0075; effective sample sizes are 908.56 for Ethereum and 308.91 for Base; maximum normalized weights are 1.14% and 1.79%. Hard trimming at propensity scores 0.02 and 0.98 was rejected because it worsened balance. Those bounds remain a central-region diagnostic.

The cost schedule is the only transported component. Each recipient transaction keeps its order, exact pre-state outcome, gross-gain bound, and weight. The donor schedule is matched by calendar date, with nearest-date matching allowed only within the same week. The observed diagonal cells reproduce the reported gross and net endpoints. Table S1 reports the weighting sensitivities.

Table S1: Cross-chain weighting sensitivities

Weighting	States	Cost schedule	Gross lower	Net lower
Equal family	Base	Base	19.62%	5.96%
Equal family	Base	Ethereum	19.62%	0.28%
Equal family	Ethereum	Base	24.88%	15.75%
Equal family	Ethereum	Ethereum	24.88%	1.90%
Transaction design	Base	Base	20.66%	9.28%
Transaction design	Base	Ethereum	20.66%	0.52%
Transaction design	Ethereum	Base	34.75%	23.17%
Transaction design	Ethereum	Ethereum	34.75%	2.29%
Volume weighted	Base	Base	57.58%	43.94%
Volume weighted	Base	Ethereum	57.58%	9.07%
Volume weighted	Ethereum	Base	63.40%	53.27%
Volume weighted	Ethereum	Ethereum	63.40%	16.38%

Point estimates use the same frozen overlap population. Weighting changes the target estimand; it does not create independent observations.

Primary uncertainty resamples families within chain and circular contiguous eight-week blocks within sampled family. Each of 5,000 draws shares donor gas, ETH/USD, and Base calldata/L1 shocks across all four cells. Four- and 13-week blocks, one-, two-, and three-month blocks, and a stationary calendar-week bootstrap are sensitivities. Transactions and calendar months are never treated as independent headline observations.

Execution validation. Transaction-ordered histories cover all 71 pools and reproduce the gross classification of all 13,768 transactions. A separate transported-cell check preserves gross classification and point-identification status.

Exact Q and D transport. Q replaces USD input through outcome-blind weighted-rank transport within pair type, economic direction, and quarter before exact re-execution. D transports the coherent donor venue state: prices, active liquidity, initialized-tick books, fee tiers, tick spacing, sibling count, and dominance. Donor states are matched within the prespecified support on liquidity rank and volatility. The support retains 3,478 Ethereum and 3,497 Base target transactions across all 29 families; its maximum absolute standardized difference is 0.0214.

Sixty-one transported executions remain point unidentified and enter the reported lower and upper bounds rather than being discarded.

Exact decomposition. For target chain t, donor chain $d \neq t$, and transported components $\mathcal{M} = \{K, Q, D\}$, let $\mathcal{P}_t$ denote the coherent transport-support population. The prespecified operator $T_S^{t\leftarrow d}$ transports the subset $S \subseteq \mathcal{M}$ from the donor while holding the remaining components at target values. The coalition value is

$$v_{t\leftarrow d}(S) = \underline{\theta}^N \left(T_S^{t\leftarrow d}(\mathcal{P}_t)\right) . \tag{S1}$$

The Shapley contribution of component $m \in \mathcal{M}$ is

$$\phi_{m,t\leftarrow d} = \sum_{S\subseteq\mathcal{M}\setminus\{m\}} \frac{|S|!\,(3-|S|-1)!}{3!} \left[v_{t\leftarrow d}(S\cup\{m\}) - v_{t\leftarrow d}(S)\right] . \tag{S2}$$

The contributions satisfy the exact adding-up identity

$$\sum_{m \in \mathcal{M}} \phi_{m,t\leftarrow d} = v_{t\leftarrow d}(\mathcal{M}) - v_{t\leftarrow d}(\varnothing). \tag{S3}$$

This is an accounting decomposition, not a structural or causal effect of a transported component.

6 Remaining Interpretation Boundaries

The data do not provide a complete external efficient-price benchmark. Raw Ethereum–Base differences are not chain treatment effects. Positive modeled net gain is not realized trader surplus, and the analysis does not identify welfare. The residual gap between feasible opportunity and observed routing can contain search, command constraints, bundling, rebates, urgency, private routing value, MEV objectives, heterogeneity, and model misspecification.

7 Data Availability

The data used in this study were derived from publicly available Ethereum and Base blockchain records.