Phase-Conditioned Constraint Coherence in a Real-World Mobility System

An EST-Compliant Tier-2 Evaluation on NYC TLC (2019–2023)

Qien Huang*

Abstract

Real-world socio-technical systems are rarely stationary. This poses a fundamental challenge for empirical evaluation of structural indicators that assume estimator coherence across time.

We apply the Force–Information–Time (FIT) framework with Estimator Selection Theory (EST) to New York City Taxi & Limousine Commission (TLC) Yellow Taxi trip data from 2019–2023. Our objective is to test whether a cost-family of constraint estimators exhibits sufficient coherence to support interpretation of Information/Constraint regime signals (P11).

We find that pooled evaluation across the full 2019–2023 period fails the EST coherence gate ($\rho = 0.543 < 0.6$), while all per-year windows pass, as do preregistered pre-COVID and post-COVID macro windows. We formalize this outcome as scope-limited validity (`OK_PER_YEAR`, `OK_PER_WINDOW`), while explicitly preserving the negative pooled result. Under EST discipline, pooled coherence failure becomes a diagnostic signal of regime heterogeneity, and meaningful structural interpretation requires preregistered phase conditioning.

Keywords: estimator selection theory; coherence gates; non-stationarity; Simpson's paradox; regime shifts; urban mobility; FIT

1 Introduction

Many empirical analyses of complex systems implicitly assume that structural indicators are stable under aggregation. In non-stationary systems—such as urban mobility across the COVID-19 shock—this assumption can fail silently: one can compute an estimator, report a value, and never notice that the value averages over incompatible regimes.

The FIT framework proposes that regime changes can be detected via interactions between **Information** and **Constraint**, but only when estimators form a coherent family under an explicit scope. Estimator Selection Theory (EST) makes this requirement operational by enforcing preregistered coherence gates and explicit failure labeling.

This note asks a narrow but fundamental question:

> *Under what temporal scopes does a constraint estimator family remain coherent in a real-world, regime-shifting system?*

We treat coherence failure as a first-class outcome, not an error to be patched.

*ORCID: 0009-0003-7731-4294. Repository: `https://github.com/qienhuang/F-I-T/`.

2 Data and System Boundary

2.1 Dataset

We use the NYC Taxi & Limousine Commission (TLC) Yellow Taxi Trip Records, distributed as monthly Parquet files [2]. We analyze January 2019 through December 2023: 60 monthly files containing 214,117,057 trips, aggregated into 1,826 daily system states.

2.2 Aggregation

Raw trips are aggregated into daily system states, computing spatial distributions (pickup zone counts) and operational metrics (duration, distance, fare). No trip-level causal inference is attempted; we work entirely at the aggregate level.

2.3 Boundary Declaration

We include only Yellow taxi operations within NYC. We exclude ride-hailing platforms (e.g. Uber, Lyft), any claims about policy intent, and demographic attribution. The interpretation domain is structural dynamics only; we are not doing demand forecasting.

3 Estimator Specification (Pre-registered)

All estimators, thresholds, and evaluation rules were preregistered and frozen prior to analysis. This is not a post-hoc search for “what works”.

3.1 Information estimator

For the Information proxy, we use pickup entropy: the Shannon entropy of the pickup-zone distribution [3], denoted `I_entropy_pu`. Higher entropy indicates more spatially dispersed pickups; lower entropy indicates concentration.

3.2 Constraint estimator family

This note focuses on the cost-family used in v1.4–v1.6, which contains two members:

Congestion measures friction in the system as travel time per unit distance:

$$C_{\text{congestion}} = \log(1 + \text{minutes\_per\_mile}).$$

Price pressure captures fare-side constraint:

$$C_{\text{price\_pressure}} = \log(1 + \text{fare\_per\_mile}).$$

Other candidate constraint proxies exist (scarcity, spatial concentration), but are not required to state the core Tier-2 finding reported here.

3.3 Coherence gate

The EST coherence gate uses Spearman rank correlation [5] with a threshold of $\rho \geq 0.6$. The rule is simple: if coherence fails, interpretation is forbidden under that scope.

4 Evaluation Protocol

We evaluate coherence under three explicitly declared scopes:

4.1 Pooled scope (2019–2023)

All data treated as a single regime.

4.2 Year-windowed scope (v1.5)

Independent coherence tests for each calendar year.

4.3 Macro-windowed scope (v1.6)

Two preregistered macro regimes: pre-COVID and post-COVID, declared as structural regime candidates based on domain knowledge rather than statistical convenience.

5 Results

5.1 Pooled evaluation

Under pooled aggregation, we observe Spearman $\rho = 0.543$, which falls below the 0.6 threshold. The outcome is **FAIL**, and we label it `ESTIMATOR_UNSTABLE` (pooled). No P11-level interpretation is permitted when we treat 2019–2023 as a single regime.

5.2 Year-windowed evaluation

When we evaluate each year independently, all individual years pass the coherence gate. The outcome is **PASS per year**, labeled `OK_PER_YEAR`. The estimators are internally consistent within a given year; it is cross-year aggregation that breaks coherence.

5.3 Macro-windowed evaluation

The preregistered pre/post-COVID windows yield:

Window	Spearman ρ	Result
Pre-COVID	0.928	PASS
Post-COVID	0.601	PASS
Pooled	0.543	FAIL

The label is `OK_PER_WINDOW`. Each macro regime is internally coherent; cross-regime aggregation is invalid.

6 Interpretation

6.1 Why pooled coherence fails

The 2019–2023 period spans distinct macro regimes with different constraint structures. Aggregating across these regimes induces a level shift analogous to Simpson's paradox [4]: within-regime correlations are positive, but the cross-regime mixture dilutes the pooled signal. This is consistent with structural breaks / level shifts in the underlying process [1]. Under EST, pooled coherence failure is therefore diagnostic, not pathological.

6.2 What scope-limited success means

The labels `OK_PER_YEAR` and `OK_PER_WINDOW` mean that constraint estimators form a coherent ordinal family within a regime, but the same estimators do not form a single family across regimes. Interpretation is therefore phase-conditional. This is not a post-hoc rescue: the windows are preregistered, and the pooled failure remains on record.

6.3 Implications for FIT P11

These results support a refined reading of Proposition P11: Information/Constraint regime signals are interpretable within phase-consistent windows, and coherence failure under aggregation can itself signal regime heterogeneity. If an I/C analysis fails coherence over a long span, the first question should be whether the system has undergone structural breaks that make pooled analysis invalid.

7 Non-claims and limits

We do not claim causal explanation of COVID impacts. We do not claim predictive regime forecasting. We do not claim universal validity of the estimator family. All claims are strictly scope-bound.

8 Reproducibility

The repo-ready entry point is `experiments/real_world/nyc_tlc_tier2p1/`. The results summary is `RESULTS.md`. Prereg files include `EST_PREREG_v1.5.yaml` (year-windowed) and `EST_PREREG_v1.6.yaml` (pre/post-COVID windows). Repo-safe run archives live under `results_runs/`.

9 Conclusion

The NYC TLC system does not admit a single coherent constraint family over 2019–2023. It does admit coherent families within declared temporal regimes. Under EST, phase conditioning is essential for structural analysis of non-stationary data, and pooled coherence failure is information rather than noise.

References

[1] Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22, 2003. doi: 10.1002/jae.659.

[2] NYC Taxi & Limousine Commission. Tlc trip record data. `https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page`. Accessed 2026-01-29.

[3] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[4] E. H. Simpson. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2):238–241, 1951. doi: 10.1111/j.2517-6161.1951.tb00088.x.

[5] Charles Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904.