

Forecasting NASDAQ Closing-Auction Volume on Real Market Data

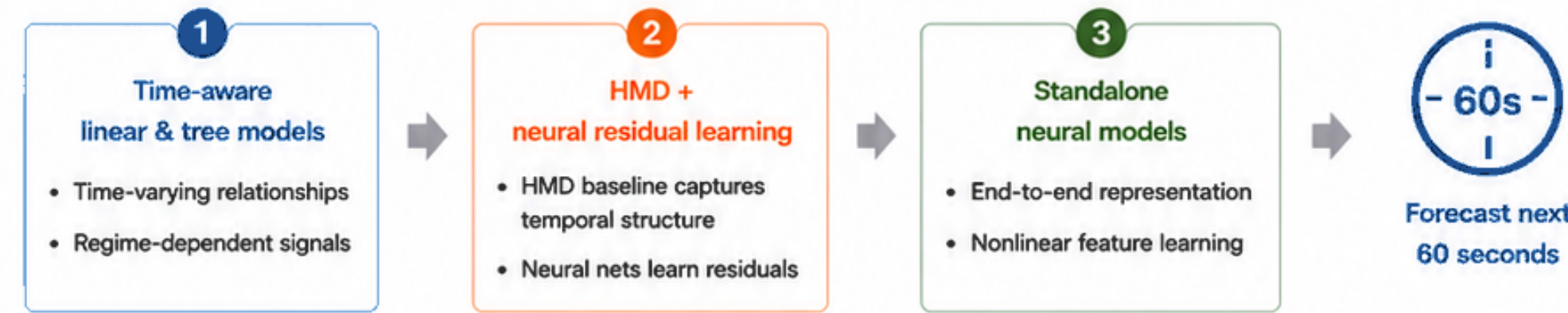
Brian Wang Winkle Fang Jinlin Chen
Columbia University ▪ AI Applications in Finance (IEOR E4743)

PROBLEM & MARKET STRUCTURE

Research Question

- How to forecast total auction volume in the NASDAQ closing cross 60 seconds ahead, using NOII auction-state data from real markets?
- We use NASDAQ's **Net Order Imbalance Indicator (NOII)** data, published every 10 seconds in the final 10 minutes before the 16:00 close.
- The prediction target is **total auction interest (matched size + imbalance size)** shifted forward by 6 buckets (60 seconds).

Research Strategy We first characterize the temporal structure of real closing-auction data and identify which features carry predictive information at different stages. We then approach the same forecasting problem from three directions:

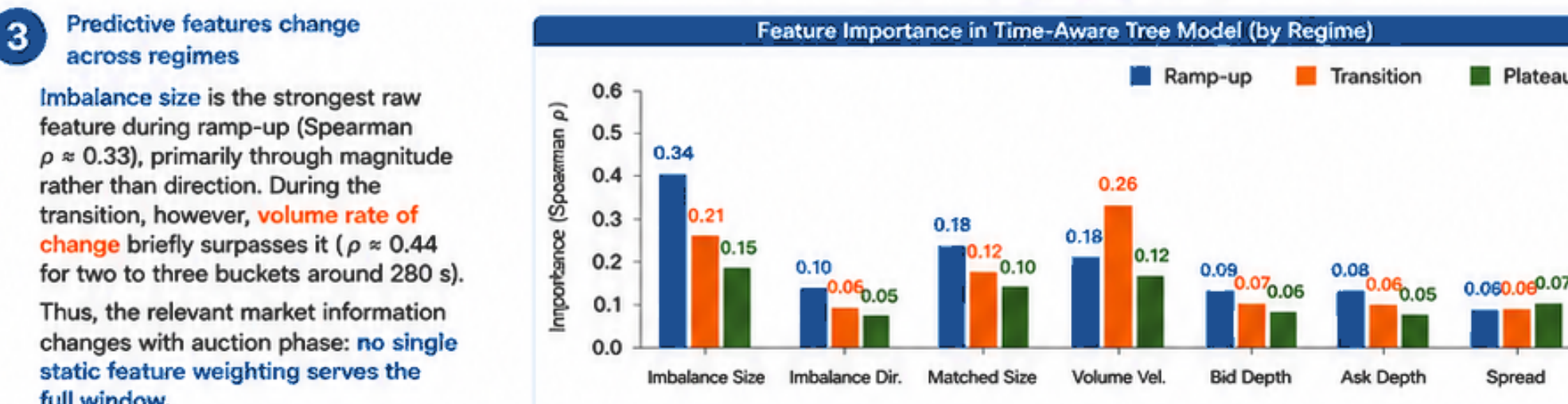
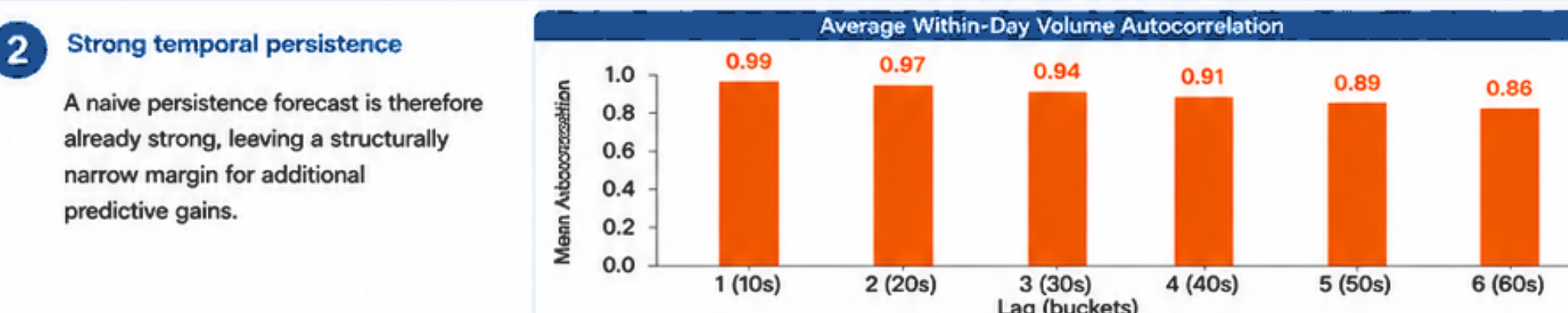
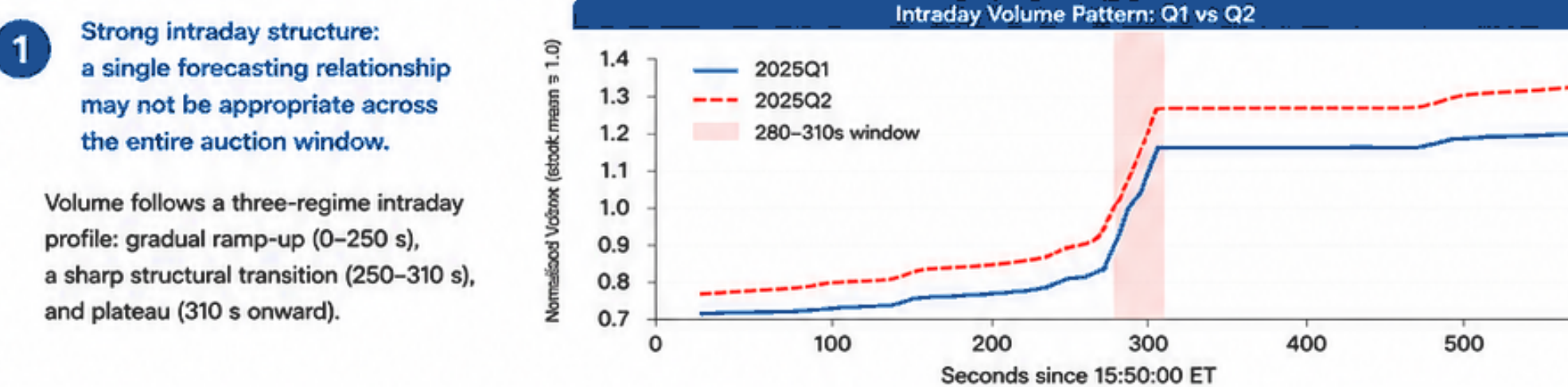


From Kaggle Benchmark to Real Market Data

Optiver Kaggle 2023 Benchmark	Our Real Market Dataset
~200 anonymized stocks	100 NASDAQ-100 constituents
480 trading days	122 sessions (Jan. 6 to Jul. 2, 2025)
Two-stage HMD + neural pipeline found HMD IC = 0.824	658,800 NOII snapshots
	HMD reaches IC = 0.944

★ This provides the real-market setting in which the three modeling approaches are evaluated.

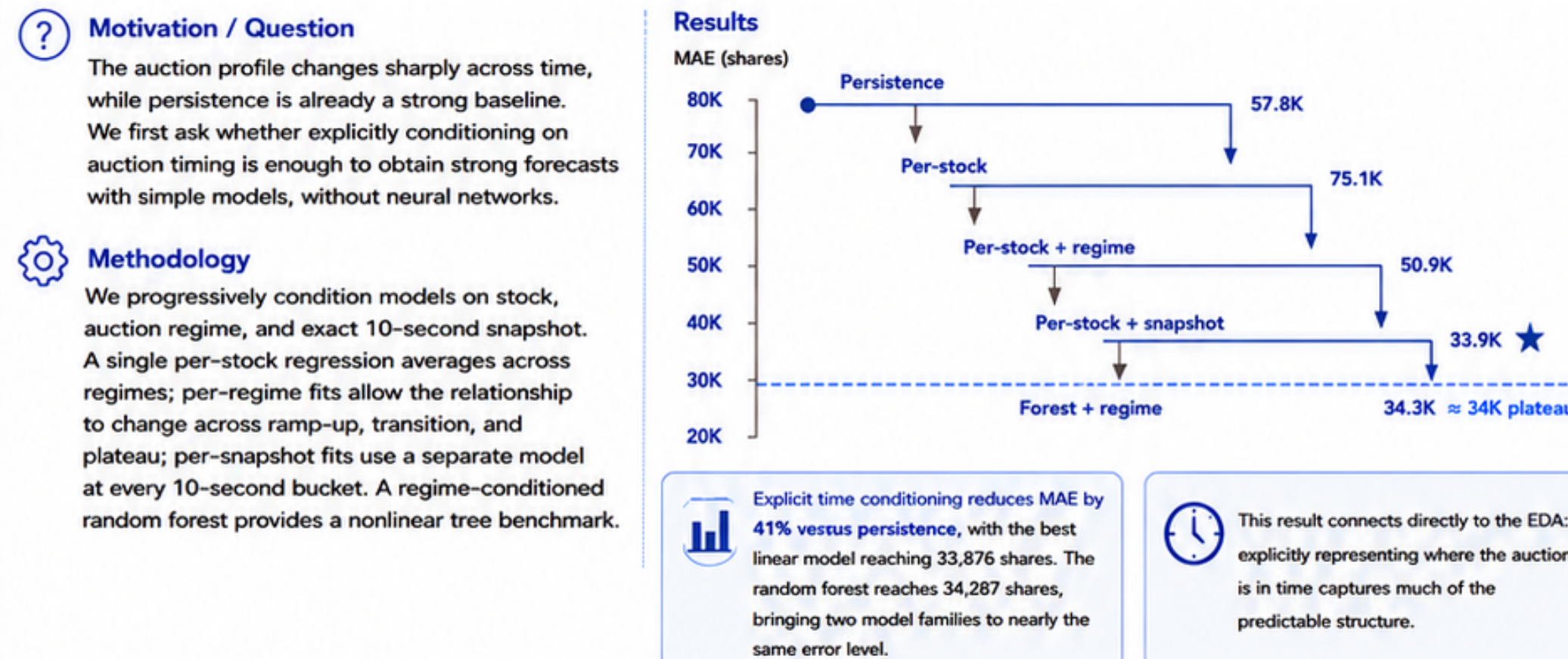
What Does the Auction Structure Tell Us?



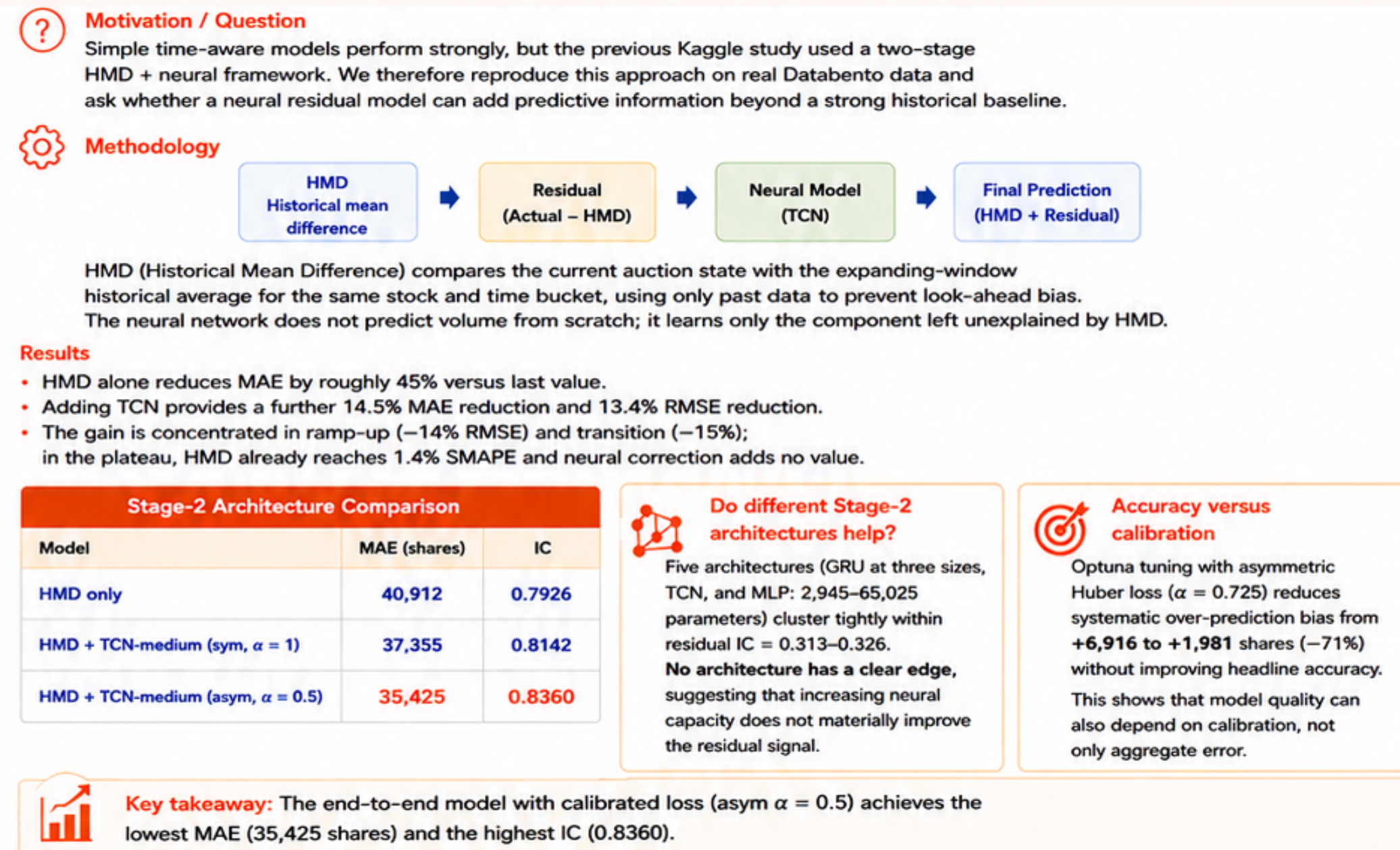
Taken together: auction predictability depends not only on the current order-book state, but also on where the auction is in its intraday evolution.

MODELING

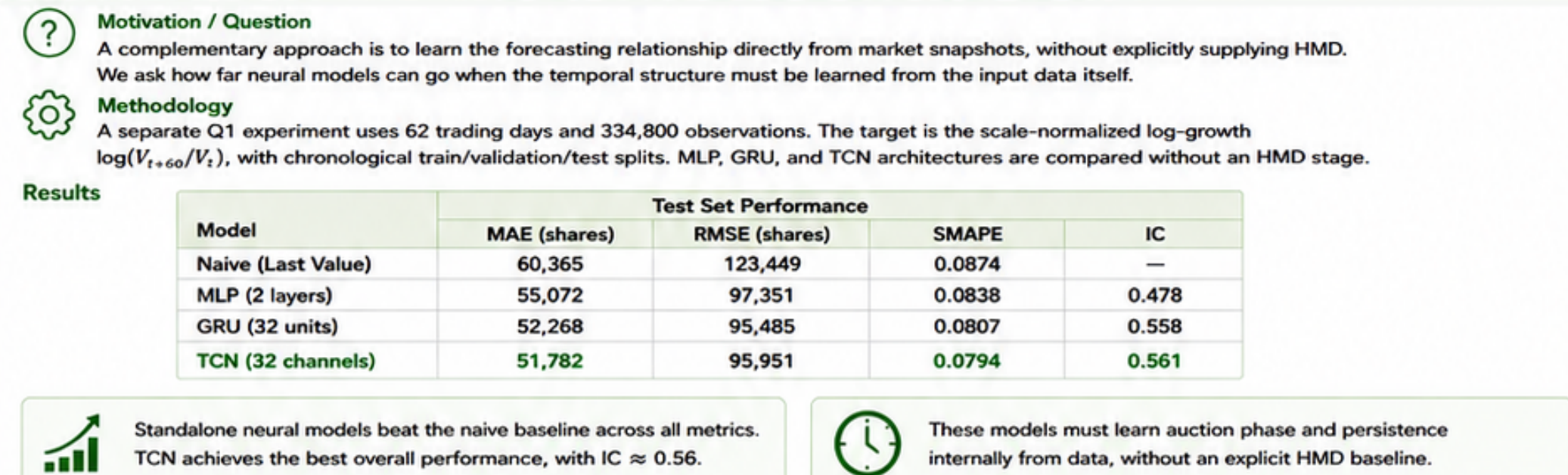
Approach 1: Time-Aware Linear & Tree Models



Approach 2: HMD + Neural Residual Learning



Approach 3: Standalone Neural Models



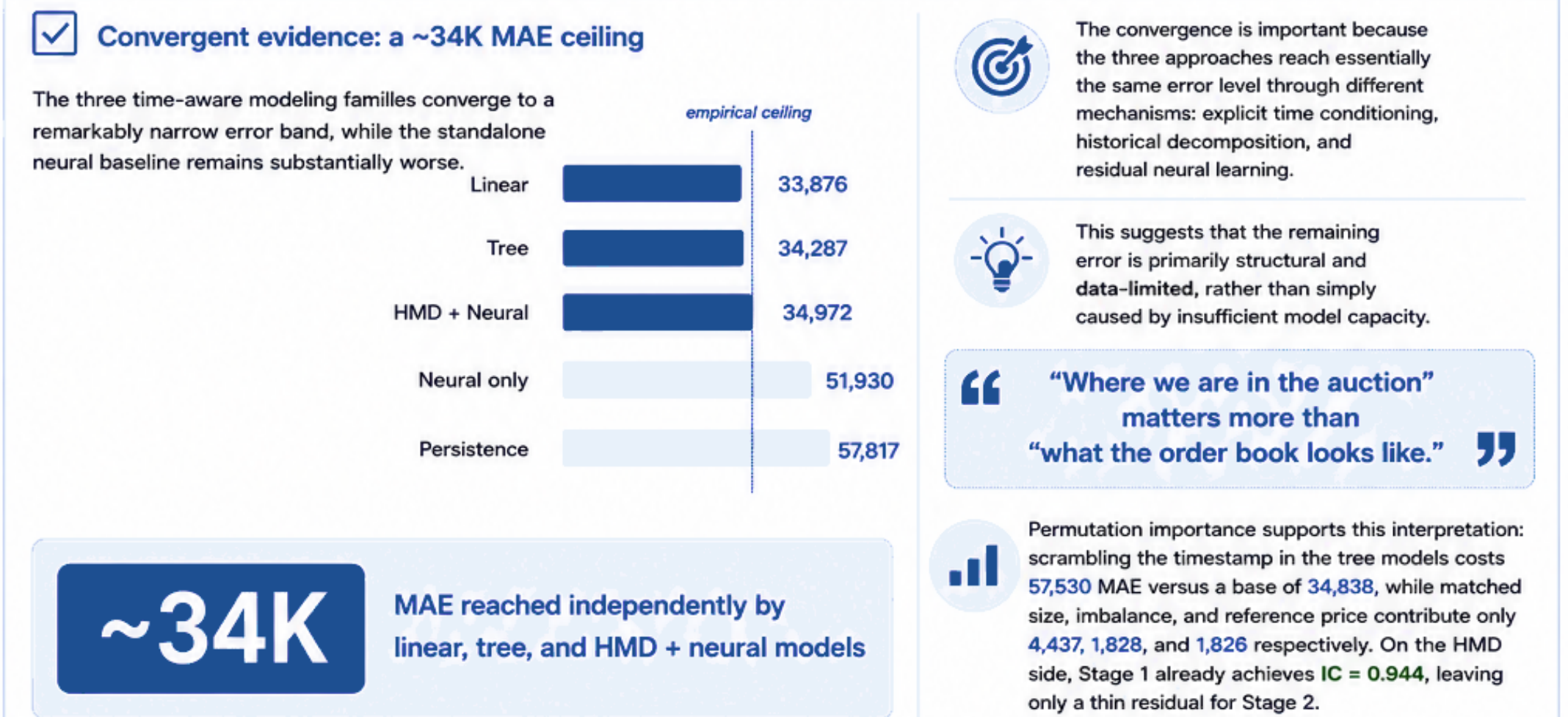
From Three Approaches to One Picture

The three approaches provide complementary evidence. Simple models perform strongly when time is represented explicitly; HMD captures most of the predictable historical structure and neural residual learning provides a smaller additional gain; standalone neural models recover substantial signal but remain below the time-conditioned approaches.

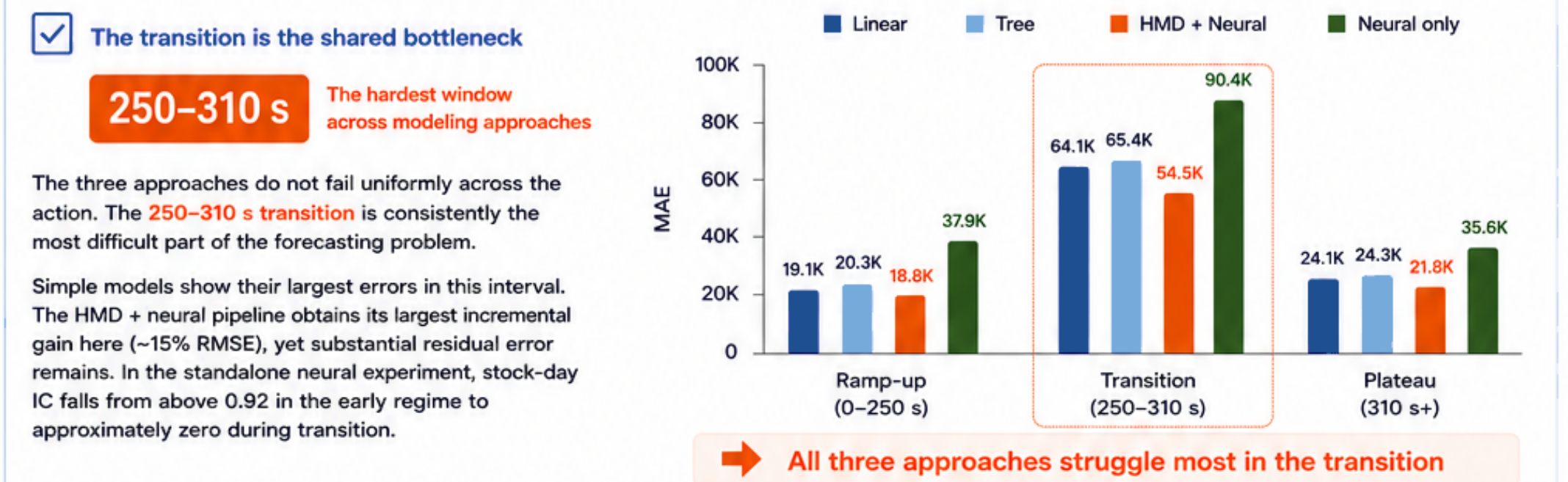
We therefore compare the approaches directly: where does performance converge, and where does the remaining error come from?

RESULTS & INTERPRETATION

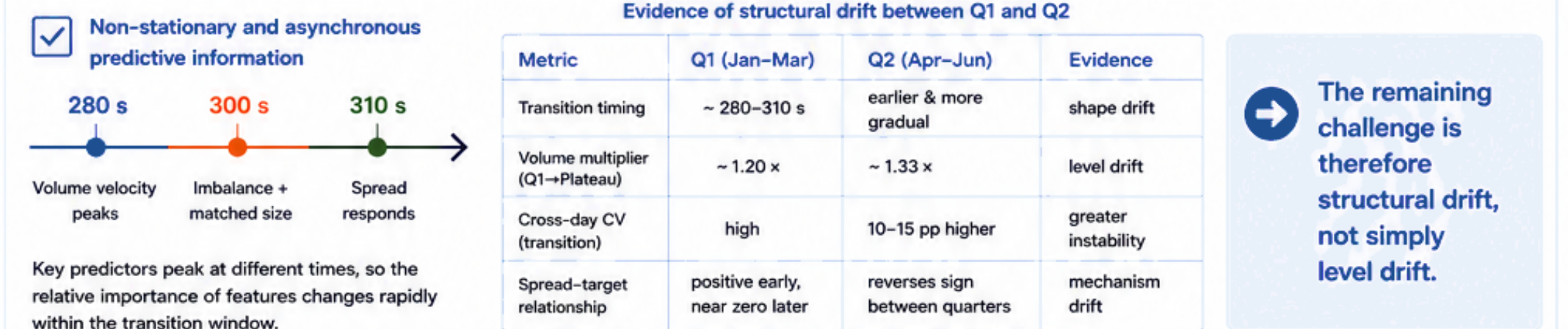
What Do the Three Approaches Agree On?



Where Does the Remaining Error Come From?



Why Is the Transition Hard?



What We Learned & Where to Go Next?

