Toward Financial World Modeling

Humzah Merchant, Alec Guthrie, Simon Mahns, Randall Balestriero, Bradford Levy

CHICAGO BOOTH THE UNIVERSITY OF CHICAGO BOOTH SCHOOL OF BUSINESS

BROWN

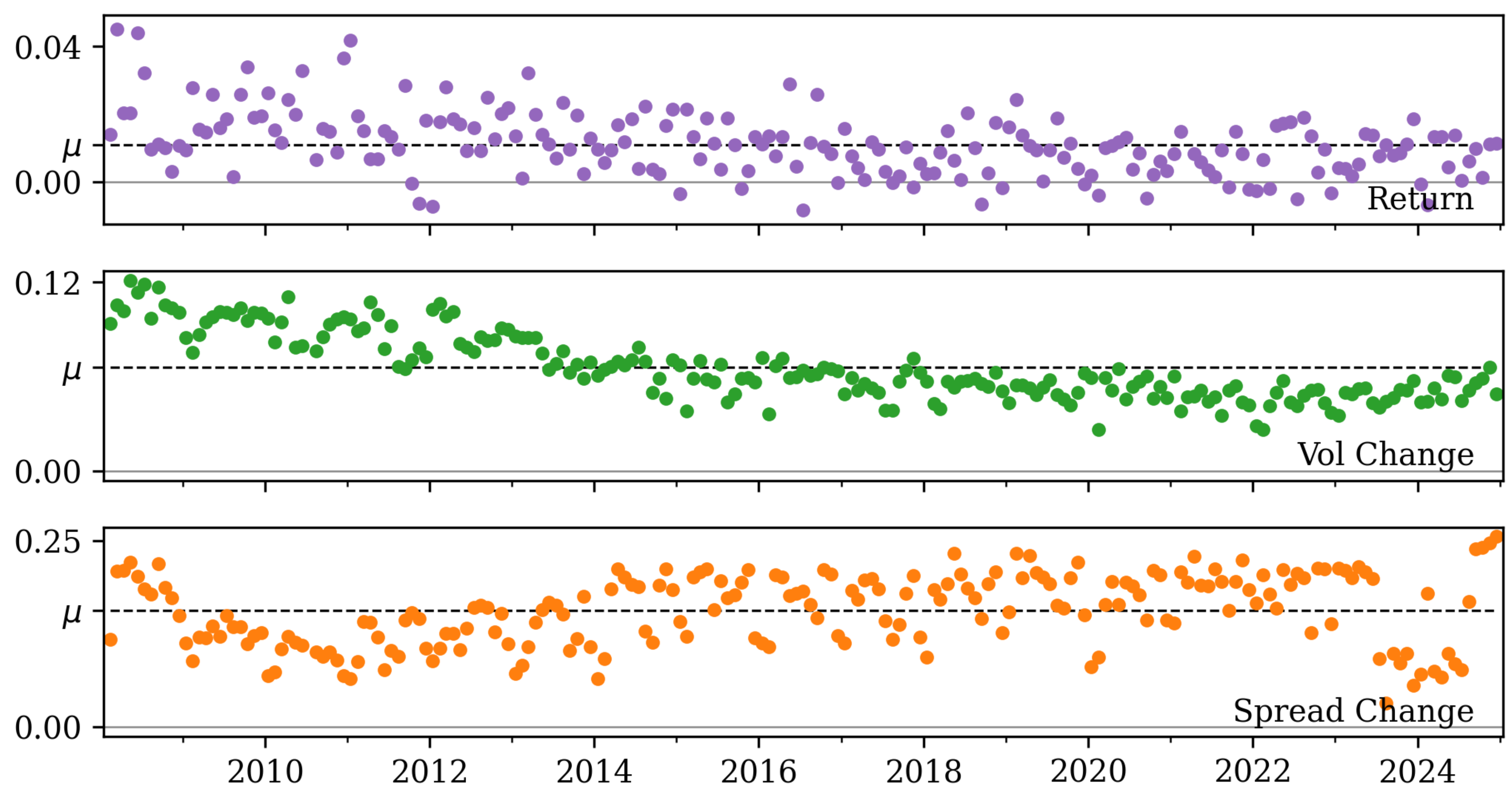
TL;DR: We introduce **Market-1T,** a best-in-class academic dataset for financial markets, and rigorously evaluate financial representation learning across predictive ability and latent organization.

Why Financial World Modeling?

- Financial markets play a vital role in allocating resources in an economy
- A financial world model can enable participants—firms and stakeholders—to better understand the consequences of their actions
- Financial markets data is typically very noisy, thereby providing a difficult testing ground

Market-1T

The literature on financial ML is fragmented between datasets which span different countries, markets, and time periods. In addition, many are private, constructed with survivorship bias, and/or lower frequency than ours (1 Hz.)

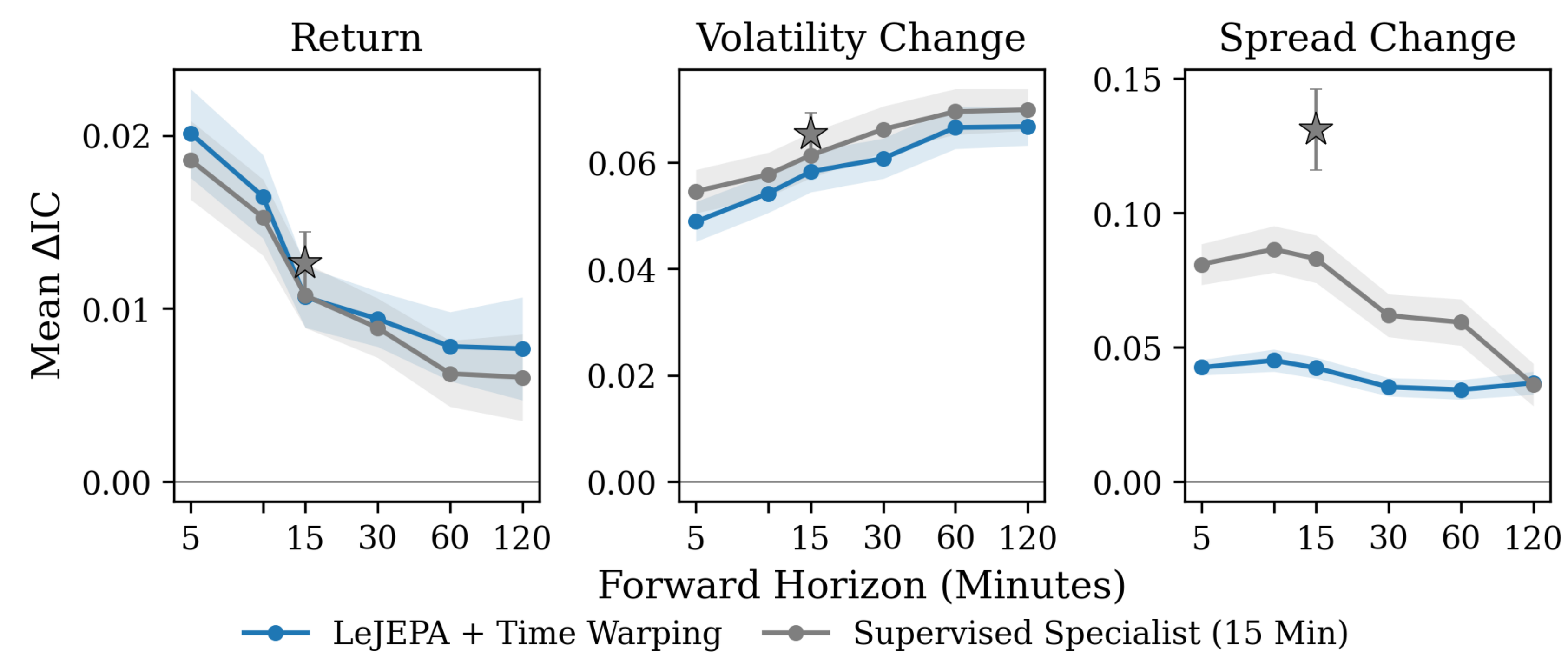


In partnership with *Massive*, we release a dataset covering top-of-the-book and trading data for all US equities from 2008-2025, leading to nearly 1 trillion aggregates.

For our evaluation, we randomly sample 32 months and evaluate on the following month.

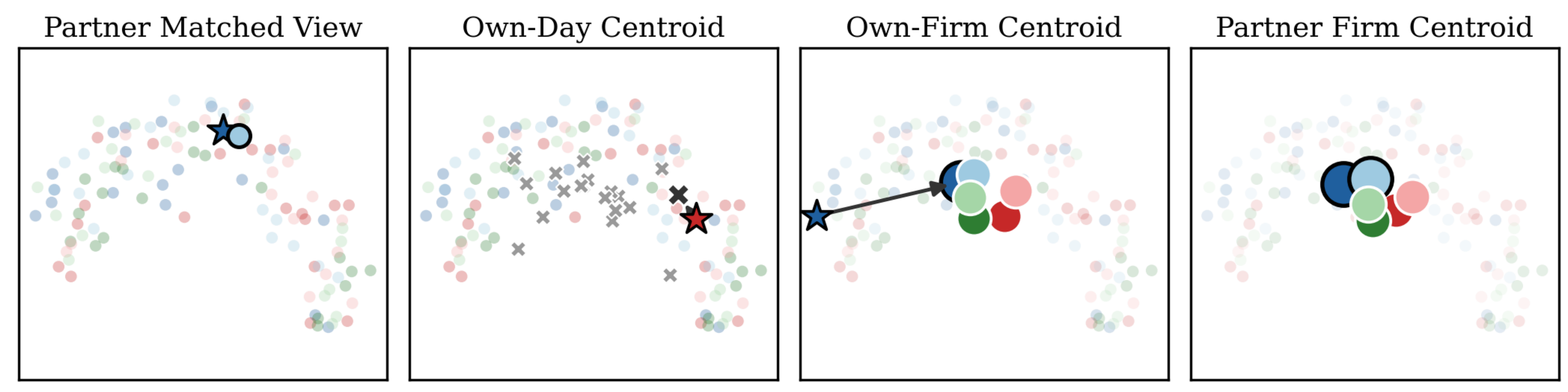
Predictive Ability

Good representations should preserve information that is useful across downstream tasks. We focus primarily on 15-minute returns information coefficient.

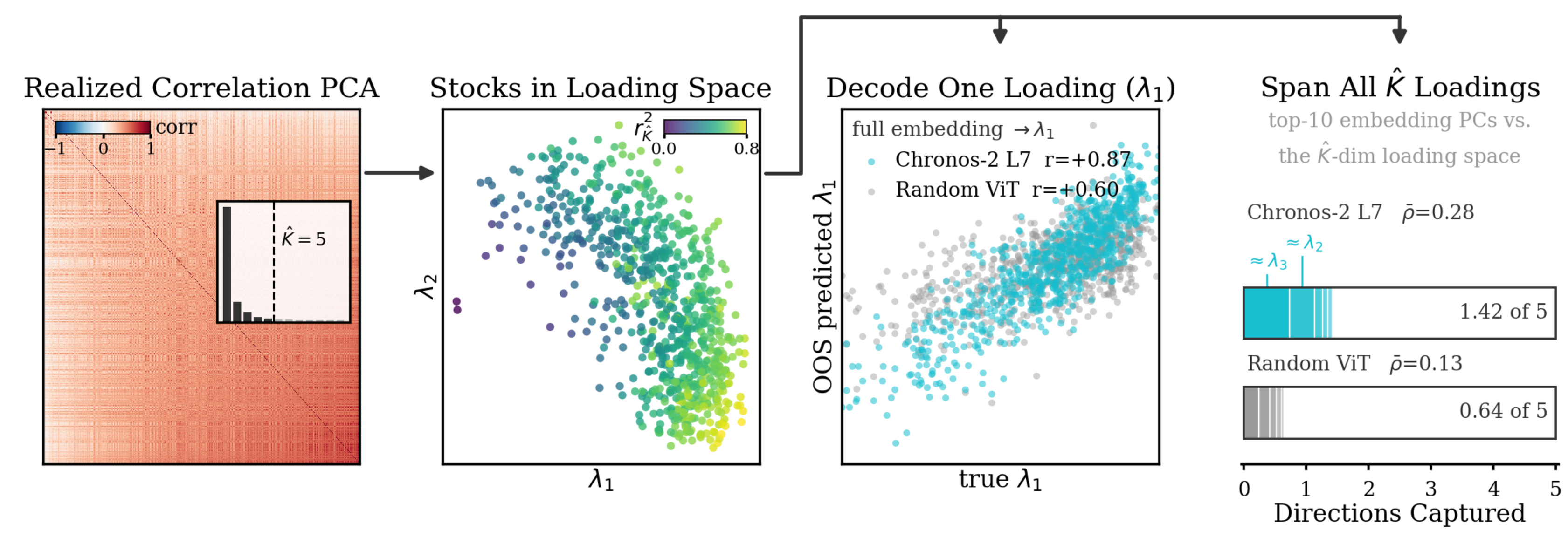


Latent Structure

We evaluate the latent structure based on the organization across time, individual assets, and similar assets ('partner firms'),



and how well it captures and prioritizes the factor structure of returns.



Summary of Results

We rank the average performance for all methods from 1 to 18 and compare the relationships across tasks (left) as well as the average performance between groups across models (right.)