Challenges with Financial Data for ML

Low Signal-to-Noise

- High randomness
- Weak patterns
- Efficient markets

Non-Stationarity

- Regime changes
- Structural breaks
- Evolving relationships

Limited Data

- Short history
- Rare events
- Survivorship bias

Look-Ahead Bias

- Data timestamps
- Restatements
- Delayed reporting

Multicollinearity

- Correlated features
- Factor exposure
- Redundant signals

Class Imbalance

- Rare events (crashes)
- Skewed returns
- Asymmetric outcomes

These challenges make financial ML harder than typical ML applications