Anomaly Detection of Time Series Data

Atsushi Motohashi, Rajesh Rao, Jingwen Zhang
Columbia University - Prof. Ali Hirsa; Wellington Management - Dr. Yossi Cohen, Anubha Thapliyal

Background

Phase 4 of an ongoing project on anomaly detection in credit card transaction data

Context and Prior Work

- Business cycle-driven spending patterns create distinct market environments, requiring regime analysis to capture underlying economic shifts.
- Prior iterations selected firm-level macroeconomic variables for regime detection using the Robust Rolling Regime Detection (R2-RD) method, and classical statistical techniques for anomalies.

Initial Challenges

- Macroeconomic variable selection chose variables which lacked causal interpretability.
- Univariate design lacked industry and firm context.
- Regime determination was limited with opaque frameworks for anomaly detection.

Objective & Contributions

Objective: Design an interpretable and practical anomaly detection model using firm-level data.

Contributions

1. Economically guided features and regimes: Replace systematic regression screening with curated features from FRED that align with inputs used in the Leading Economic Index (LEI) and validated against National Bureau of Economic Research (NBER).
2. Add industry and firm features from exploratory analysis, including sector ETF returns and firm-level volatility to improve model inputs.
3. Add a Random Forest regressor model to compare against classical statistical techniques which leverage majority voting for anomalies.

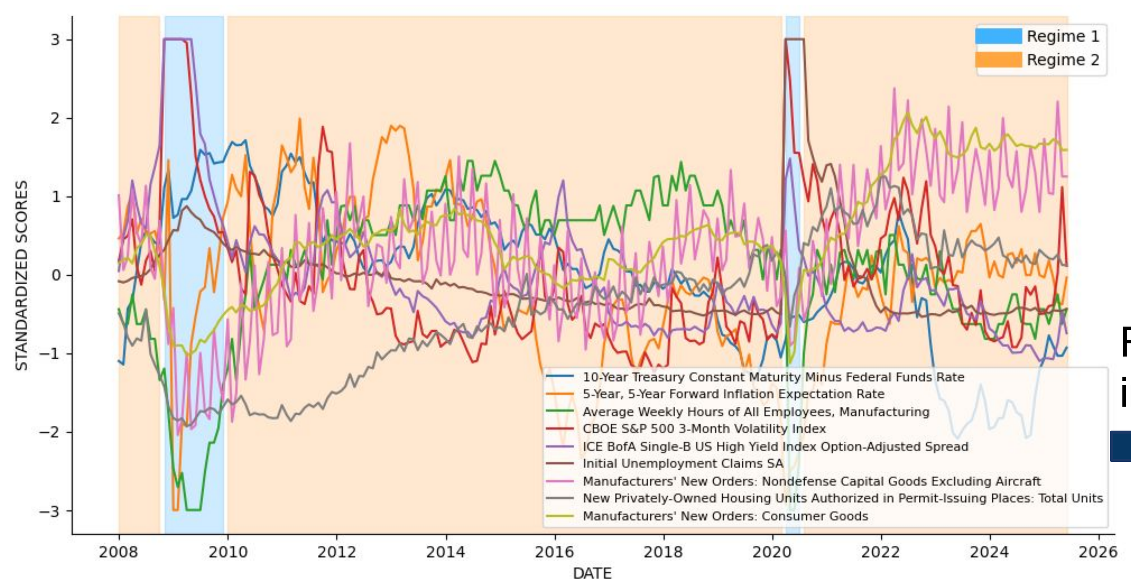
Data Source

Macroeconomic Indicators: Federal Reserve Economic Data (FRED), Dec 2007 - Jun 2025.
Corporate Transaction Data: Credit card transaction records provided by Wellington, covering 42 companies across 8 sectors, Jan 2016 – Dec 2022.
Market & Sector Data: SPDR sector ETFs and firm-level daily adjusted stock prices from YFinance.

Modeling

Business Cycle Regime Detection (R2-RD)

- Applied to FRED-based LEI indicators, the method successfully detected the 2008 financial crisis and the COVID-19 recession.
- The results are broadly consistent with the NBER business cycle dates.



Regime information

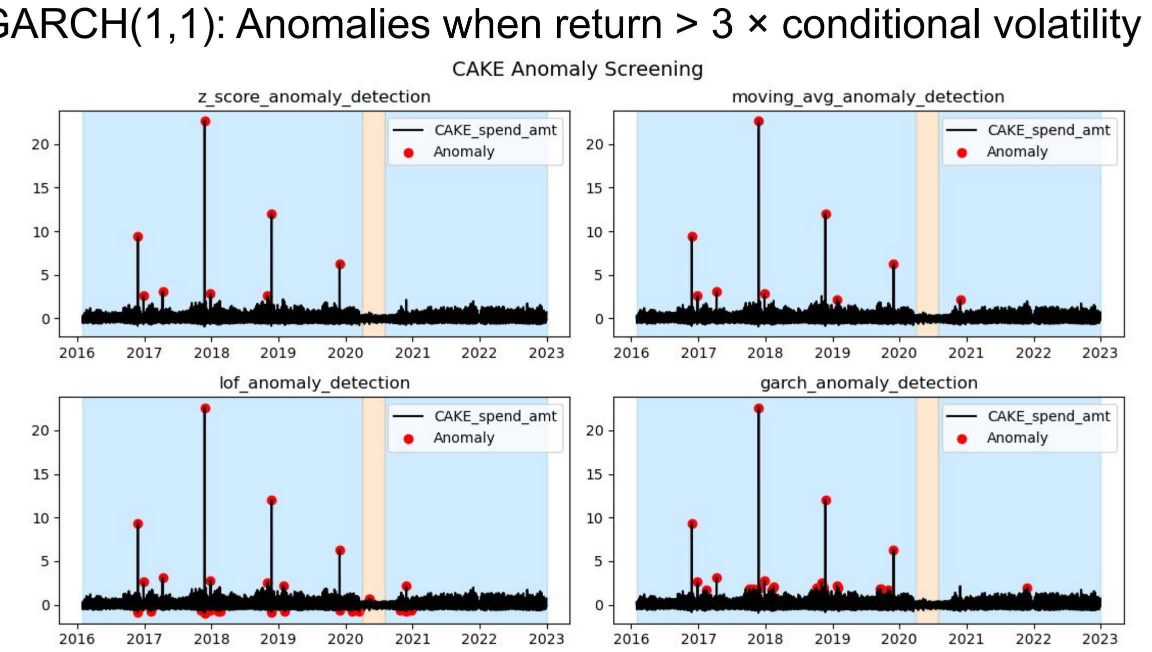
Methodological Enhancements

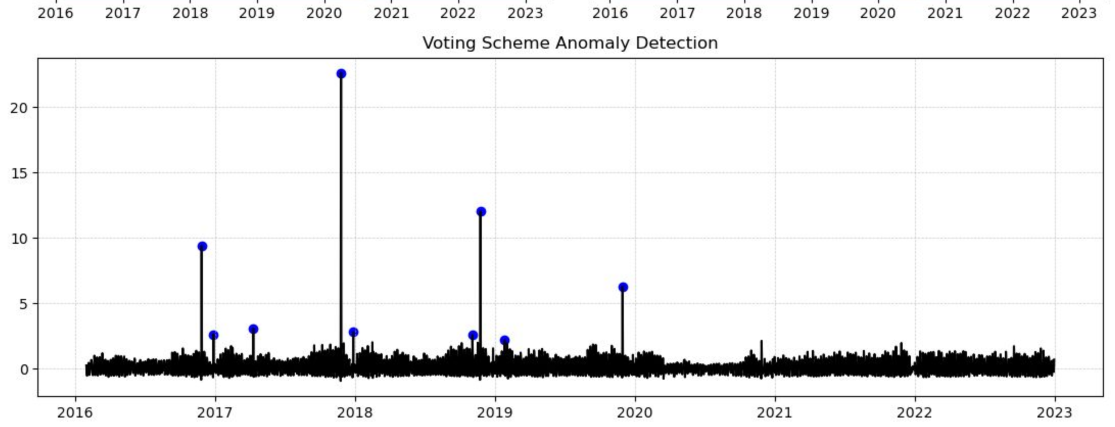
- **Cost Function**: Balances differences in means (Euclidean) and variances (Frobenius).
- **Stable Initialization**: Reuses parameters when the number of regimes remains unchanged, improving reproducibility and reducing randomness.
- **Expanding Window Training**: Trains incrementally up to 85% of the sample, helping prevent overfitting and avoiding unnecessary regime splits.

Statistical Approach to Anomaly Detection

"***Statistical Anomaly***" if flagged by ≥ 3 of 4 methods (Voting scheme)

1. Z-score: Flags extreme deviations ($|Z| > 3$)
2. Moving Average: Deviations from rolling mean (30-day window, 3σ)
3. Local Outlier Factor (LOF): Detects sparse-density outliers
4. GARCH(1,1): Anomalies when return > 3 × conditional volatility





Machine Learning Approach

"***ML Anomaly***" if the actual spending value falls outside the Random Forest prediction interval (mean ± 3σ across trees).

Features (4 Categories)

1. Firm-level spending trends (internal)
2. Company stock volatility (external)
3. Sector ETF returns (external)
4. Macro & Seasonal factors (external)

Random Forest Implementation Steps

- Merge internal + external features
- Define target = % change in spending
- Train (75%) and Test (25%) split for model
- Aggregate tree predictions → mean & σ
- Flag anomalies if outside mean ± 3σ

Hyperparameter Tuning

- Grid Search: tree depth, min split, min leaf

Feature	Importance
CAKE_spend_amt_14ewma	26.2562
CAKE_spend_amt_14ma	20.6082
CAKE_daily_vol	16.7631
CAKE_spend_amt_30ma	14.3265
Consumer Discretionary_daily_rtrn	10.5149
month	5.5567
CAKE_month_vol	5.49
holiday	0.2433
regime	0.2412

Exploratory Data Analysis & Feature Engineering

Macro-level

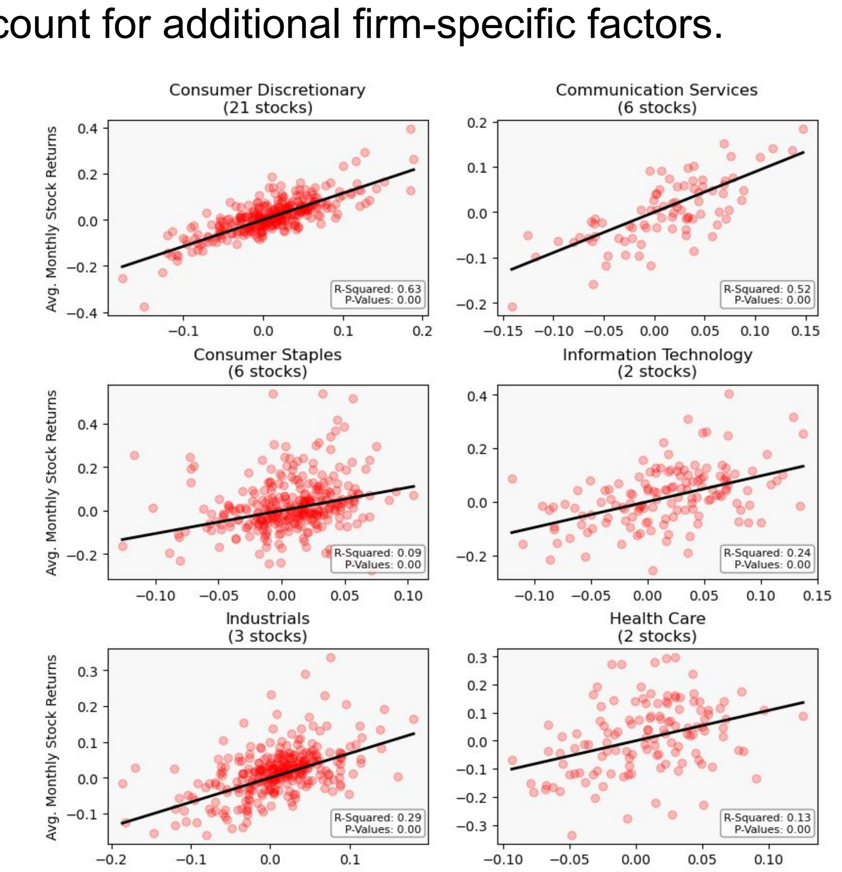
- **Variable selection**: Based on the Conference Board LEI, with missing components substituted with FRED proxies.

No.	Official LEI Component	Substituted Indicator Used	FRED Code
1	Average weekly hours in manufacturing	Same	AWHAEMAN
2	Average weekly initial claims for unemployment insurance	Same	ICSA
3	Manufacturers' new orders for consumer goods and materials	Same	ACOGNO
4	ISM® Index of New Orders	*Not available via FRED*	–
5	Manufacturers' new orders for nondefense capital goods excluding aircraft	Same	UNXANO
6	Building permits for new private housing units	Same	PERMIT
7	S&P 500® Index of Stock Prices	CBOE 3-month implied volatility index (proxy for S&P 500 expectations)	VXVCLS
8	Leading Credit Index™	ICE BofAML High Yield Option-Adjusted Spread	BAMLH0A2HYB
9	Interest rate spread (10-year Treasury minus Fed Funds Rate)	10-Year Treasury constant maturity minus Fed Funds Rate	T10YFFM
10	Average consumer expectations for business conditions	5-Year forward inflation expectation rate	T5YIFRM

- **Data processing**: Higher-frequency series were aggregated to business month-end averages, missing values were forward filled to account for reporting lags, and each series was standardized with Winsorization (±3σ).
→ **Inputs for regime detection.**

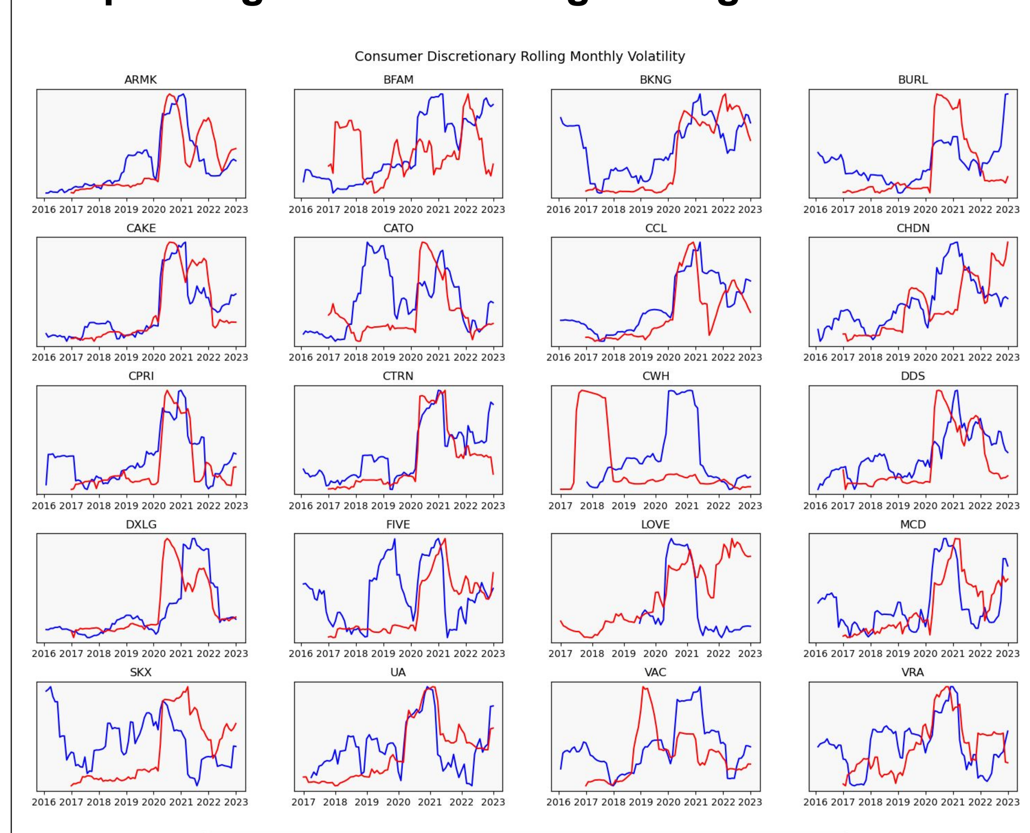
Sector-level

- **Sector ETFs**: The returns capture sector-wide trends → Selected features: **Sector ETF returns**
- Outliers (e.g., Consumer Staples) highlight the need to account for additional firm-specific factors.



Firm-level

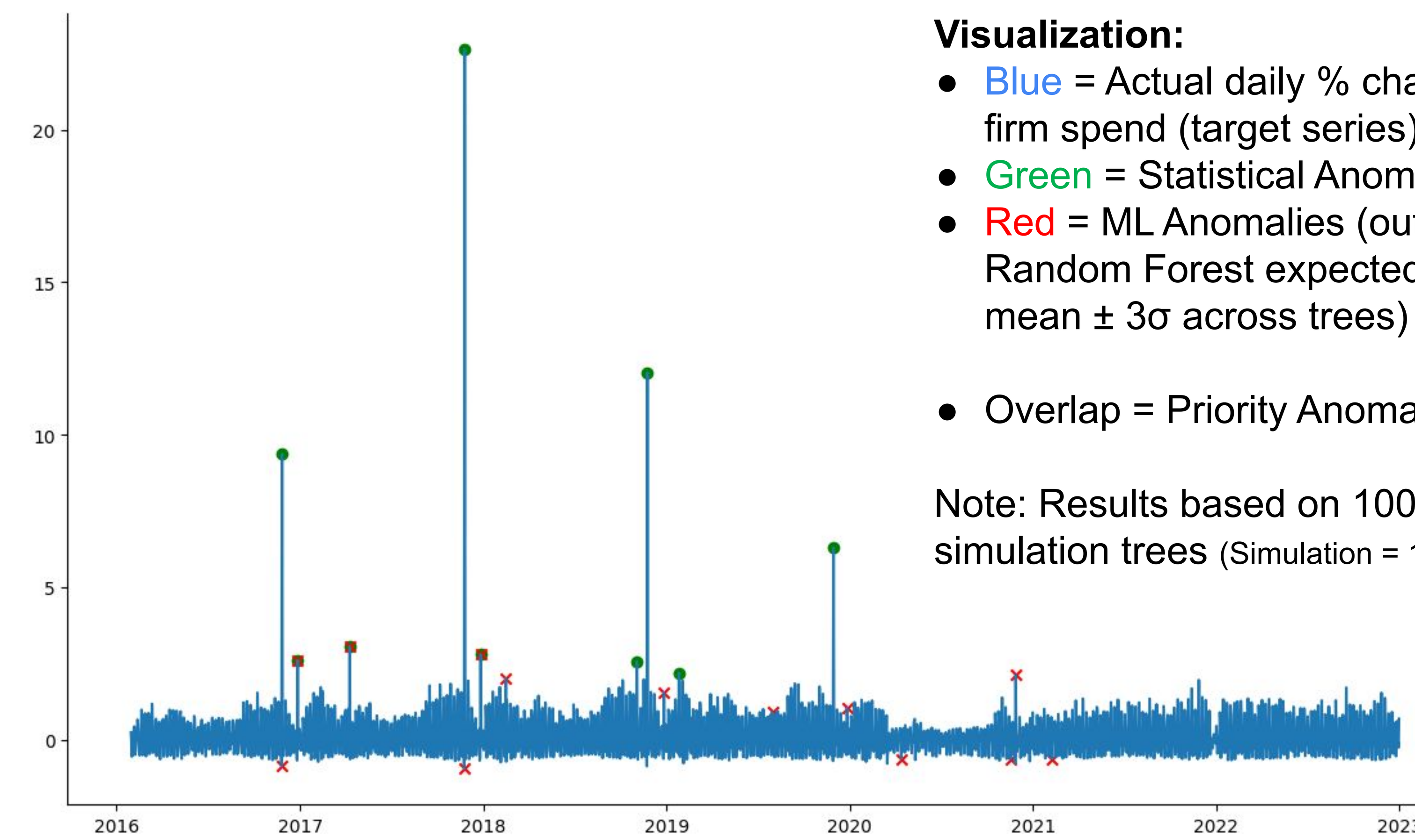
- **Stock Volatility**: Spending volatility and stock return volatility show correlated relationships. → Selected features: **Stock return volatility**
- **Spending Amount Moving Average**: trends



Analysis of Results

Case Study: CAKE (Cheesecake Factory Inc.)

- **Priority Anomaly = Statistical Anomaly and ML Anomaly**
- Overlapping anomalies were detected by both methods. These are considered as priority anomalies for analyst review.



Visualization:

- Blue = Actual daily % change in firm spend (target series)
- Green = Statistical Anomalies
- Red = ML Anomalies (outside Random Forest expected range: mean ± 3σ across trees)
- Overlap = Priority Anomalies

Note: Results based on 1000 simulation trees (Simulation = 1000)

Extension to Multiple Companies (Simulation = 100)

	z_score	moving_avg	lof	garch	voting	random_Forest	percent_match	ratio
CAKE	8.0	9.0	27.0	23.0	9.0	17.0	0.333333	0.529412
BURL	13.0	39.0	27.0	24.0	23.0	22.0	0.304348	1.045455
DDS	8.0	42.0	27.0	9.0	8.0	8.0	0.250000	1.000000
SKX	15.0	21.0	27.0	26.0	18.0	24.0	0.222222	0.750000
DXLG	10.0	27.0	27.0	11.0	9.0	15.0	0.222222	0.600000
IAC	31.0	13.0	27.0	36.0	12.0	23.0	0.222222	0.521739
ROL	35.0	71.0	27.0	59.0	23.0	45.0	0.142857	0.511111
SFM	9.0	23.0	27.0	9.0	9.0	5.0	0.111111	1.800000
CABO	53.0	67.0	27.0	71.0	42.0	15.0	0.100000	2.800000
CTRN	20.0	25.0	27.0	29.0	21.0	24.0	0.095238	0.875000
EA	36.0	36.0	27.0	39.0	24.0	19.0	0.090909	1.263158
VAC	29.0	50.0	27.0	45.0	23.0	10.0	0.047619	2.300000
CHDN	36.0	76.0	27.0	45.0	23.0	21.0	0.045455	1.095238
BFAM	28.0	91.0	27.0	51.0	23.0	22.0	0.043478	1.045455
CWH	23.0	37.0	27.0	32.0	23.0	3.0	0.000000	7.666667
LMND	28.0	48.0	27.0	34.0	18.0	5.0	0.000000	3.600000
SSTK	46.0	47.0	27.0	54.0	36.0	17.0	0.000000	2.117647
ZD	64.0	68.0	27.0	64.0	57.0	27.0	0.000000	2.111111
UA	8.0	33.0	27.0	9.0	8.0	4.0	0.000000	2.000000
CCL	46.0	29.0	27.0	42.0	19.0	10.0	0.000000	1.900000
MED	48.0	72.0	27.0	59.0	30.0	33.0	0.000000	0.909091

Scope: Applied the same model to all firms, using the defined inputs

Agreement Metric:

- Measured overlap rate between RF and statistical anomaly flags.
- Exclude large count gaps; ratio > 0.5. Ratio = vote count / RF count

Findings: Agreement varies by ticker; some firms show strong alignment, others do not.

Interpretation:

- Low agreement can mean RF is capturing nonlinear/interaction effects or generating false positives from the data.
- Without ground truth, neither method is definitively superior.

Actionable Rule:

- Prioritize anomalies confirmed by both methods (RF ∩ Statistical).
- Treat single-method signals as secondary monitoring cues

Limitation & Future Work

- **Divergence in results**: Statistical ensemble and ML approaches often differ.
- **Promising model**: LSTM is well-suited for time series but not yet applied.
- **Data preprocessing**: Unadjusted credit-card transaction percentage changes used; seasonally adjusted series created negative values spend/count amounts.
- **Key challenge**: Lack of labeled data or "true" set to determine an anomaly.
→ **Possible solutions**: 1. Build labels from historical anomalies (M&A, earnings revisions). 2. Apply text mining to investor-relations (IR) documents.

References & Acknowledgements

- Braei, M., & Wagner, S. (2020). Anomaly detection in univariate time-series: A survey on the state-of-the-art. arXiv preprint arXiv:2004.00433. https://doi.org/10.48550/arXiv.2004.00433
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41 (3), 1–58. https://doi.org/10.1145/1541880.1541882
- Hirsa, A., Xu, S., & Malhotra, S. (2024). Robust rolling regime detection (R2-RD): A data-driven perspective of financial markets. SSRN. https://doi.org/10.2139/ssrn.4729435
- Schmidl, S., Wenig, P., & Papenbrock, T. (2022). Anomaly detection in time series: A comprehensive evaluation. Proceedings of the VLDB Endowment, 15 (9), 1779–1797. https://doi.org/10.14778/3538598.3538602
- Zhong, Z., Fan, Q., Zhang, J., Ma, M., Zhang, S., Sun, Y., Lin, Q., Zhang, Y., & Pei, D. (2023). A survey of time series anomaly detection methods in the AIOps domain. arXiv preprint. https://arxiv.org/abs/2308.00393

Acknowledgements: We are deeply grateful to Wellington Management for offering us this invaluable opportunity. We also appreciate Professor Ali Hirsa for his supervision, and Advisors Yossi Cohen and Anubha Thapliyal for their valuable feedback.