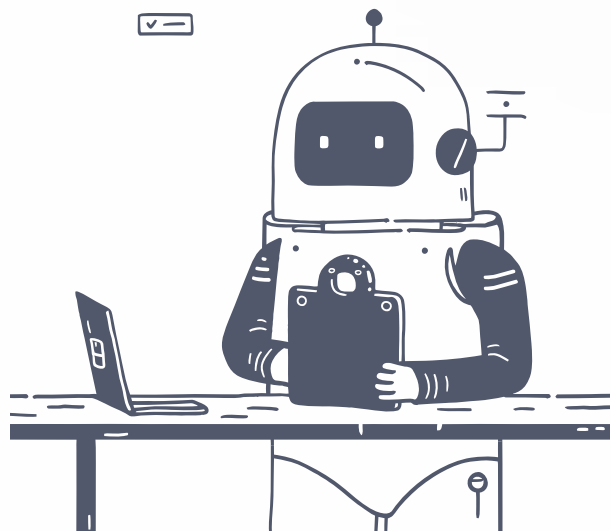
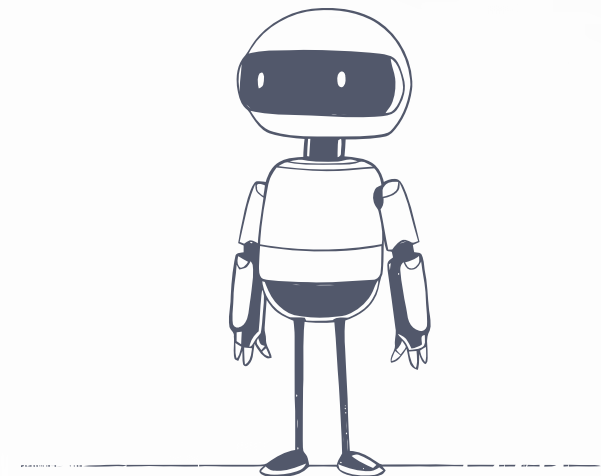


Market Anomaly Demo



Andrea Clark

December 3rd, 2025



Multi-Source Market Anomaly Detection Engine

An agentic system (powered by IBM Agentics) that detects market anomalies by seamlessly integrating structured market data, analyst expectations, and advanced LLM-powered reasoning to provide actionable insights.



WRDS Data Ingestion

- CRSP delivers daily market metrics including returns and volume.
- Compustat provides fundamental data on revenues and margins.
- IBES supplies analyst forecasts and consensus estimates.



Text2SQL

Natural-language financial querying powered by schema-aware SQL generation, enabling intuitive data exploration without manual query construction.



Anomaly Detection

Independent signal channels aggregate into unified anomaly scores with component-level explanations and clear z-score interpretability.

Data Architecture: WRDS Integration Pipeline

CRSP Market Data

Daily returns, volatility metrics, trading volume, and shares outstanding power statistical anomaly detection through return shocks and volatility analysis.

Compustat Fundamentals

Quarterly revenue reports, profit margins, leverage ratios, and cash flow statements identify fundamental drifts and accounting irregularities.

IBES Analyst Data

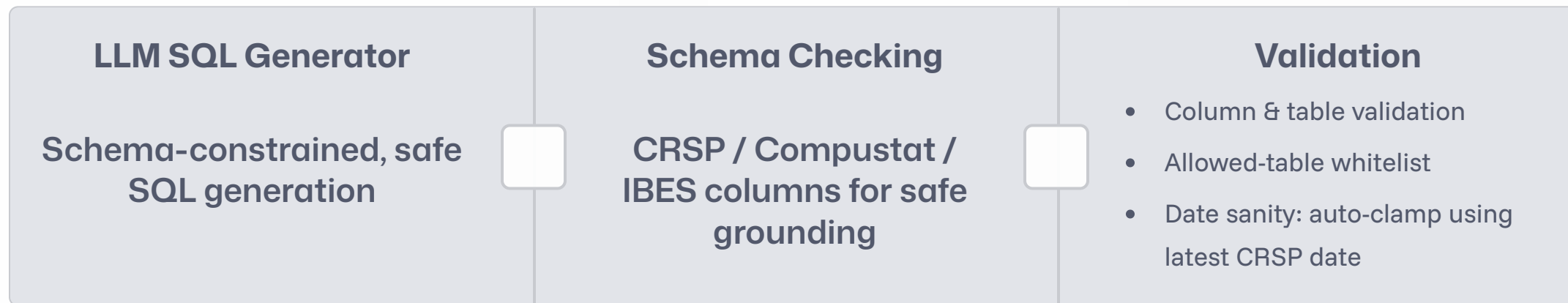
EPS forecasts, revision tracking, and surprise metrics capture sentiment deviations and expectation misalignments across the market.

Processing Pipeline

Data flows from WRDS through quality control ingestors into local SQLite and Parquet storage, feeding automated feature generation.

Text2SQL: Natural-Language Querying

Allows analysts to ask intuitive questions like *"Which companies beat earning estimates last quarter?"* to query the rich market database without having to specify any SQL code.



Anomaly Detection & Scoring Framework



CRSP Statistical

Return shocks, volatility jumps, liquidity drops analyzed through z-score interpretation



Compustat Fundamentals

Margin changes, leverage shifts, accrual anomalies versus historical baselines



IBES Sentiment

Forecast revisions, surprise dispersion, consensus divergence tracking

Z-Score Interpretation

- $|z| \leq 0.5$: Normal range
- $|z| \approx 1.5$: Moderate anomaly (~33%)
- $|z| \approx 2.0$: Elevated concern (~50%)
- $|z| \geq 3.0$: Strong anomaly (~100%)

Aggregation Model

Weighted combination of independent channels produces final anomaly score with full transparency and component-level breakdown.

$$A_{final} = w_1 A_{CRSP} + w_2 A_{Fund} + w_3 A_{Sent}$$

Compustat & IBES Anomaly Probabilities

Compustat

- Build rolling baseline for each metric (revenue, margins, leverage):
 μ = historical mean/median, σ = volatility/MAD
- Compute standardized deviation: $z = \frac{X_{\text{current}} - \mu}{\sigma}$
- Convert to anomaly probability: $P_{\text{Fund}} = 1 - \Phi(|z|)$
- Aggregate across metrics using max or weighted average

IBES

- Compute EPS surprise: $z_{\text{surprise}} = \frac{EPS_{\text{actual}} - EPS_{\text{consensus}}}{\sigma_{\text{surprise}}}$
- Compute forecast revision anomaly: $z_{\text{rev}} = \frac{\Delta_{\text{rev}}}{\sigma_{\text{rev}}}$
- Convert each into anomaly probabilities:
 $P_{\text{surprise}} = 1 - \Phi(|z_{\text{surprise}}|), \quad P_{\text{rev}} = 1 - \Phi(|z_{\text{rev}}|)$
- Combine: $P_{\text{IBES}} = \max(P_{\text{surprise}}, P_{\text{rev}})$

How Are the Weights Computed?

The weights in our final anomaly score formula are dynamically determined to reflect the informativeness of each signal channel, ensuring a data-driven approach.

Signal Variance Normalization (Default)

Each component's weight is proportional to how informative (i.e., how variable) its anomaly signal is across the universe of securities.

$$w_i = \frac{Var(A_i)}{\sum_j Var(A_j)}$$

Adaptive Intuition

- If CRSP anomalies vary significantly, CRSP gets more weight.
- If fundamentals are stable but IBES surprises spike, IBES gets more weight.
- Weights adapt automatically over time to market conditions.

Future Directions



Integrate External News & Sentiment

- Add news-based anomaly signals (e.g. Yahoo Finance, Google News, RSS feeds)
- Incorporate Google Trends/search-volume spikes as early indicators
- Use NLP models for tone, uncertainty, and narrative shifts
- Align news timestamps with CRSP/IBES events for causal consistency



Bayesian Market Priors

- Build an adaptive Bayesian anomaly model
- System learns prior distribution over typical market behavior
- High volatility → weaker prior (more flexible)
- Stable markets → stronger prior (shrinkage toward historical norms)
- Adaptive anomaly probabilities calibrated to market regimes



Improved Feature Engineering

- Add cross-sectional features: peer deviation, sector-normalized anomalies
- Integrate macroeconomic indicators (VIX, rates, PMI) for market context
- Explore temporal models (RNNs, Transformers) for sequence-aware anomaly scoring