An Agentic LLM Framework for Multi-Source Financial Market Anomaly Detection

Andrea Clark-Sevilla
Computer Science
Columbia University
New York, USA
aoc2111@columbia.edu

***Abstract*—This paper presents a multi-agent system for detecting market anomalies through comprehensive financial analysis. The system translates natural language queries into integrated analyses across multiple WRDS datasets (CRSP, Compustat, IBES, Capital IQ) supplemented with Google Trends and Google News data. We introduce a Bayesian confidence weighting approach that recognizes data quality variations across companies, ensuring that predictions for well-covered firms leverage comprehensive market signals while smaller companies rely on fundamental and event-based indicators. The system employs specialized agents for natural language parsing, Text2SQL generation, anomaly detection, and confidence assessment, enabling sophisticated market analysis through intuitive user queries.**

***Index Terms*—market anomaly detection, multi-agent systems, financial data integration, Bayesian weighting, Text2SQL**

I. Introduction

Detecting anomalies in financial markets is essential for risk management, investment decision-making, and regulatory oversight. Traditionally, this task has relied on quantitative methods ranging from statistical techniques to more complex machine learning methods. However, these approaches typically require substantial technical expertise, struggle to integrate heterogeneous data sources, and often fail to provide interpretable explanations for their findings. This work presents a novel agentic pipeline that enables users to specify anomaly detection criteria in natural language and automatically generates comprehensive market reports by synthesizing information across multiple financial databases and macroeconomic indicators.

II. Related Work

A. Traditional Approaches

Early approaches to financial anomaly detection relied heavily on statistical methods. Z-score analysis, which flags observations exceeding a specified number of standard deviations from the mean, remains a common baseline for identifying unusual price movements or trading volumes [1]. More sophisticated statistical approaches include Mahalanobis distance for multivariate outlier detection and GARCH-family models for capturing time-varying volatility dynamics [2], [3]. While computationally efficient, these methods typically assume specific distributional properties that financial data often violate and struggle to incorporate contextual information about corporate events or macroeconomic conditions.

B. Machine Learning Approaches

The advent of machine learning brought more flexible anomaly detection tools. Unsupervised methods such as Isolation Forest [4] have proven effective at identifying financial anomalies by exploiting the principle that anomalous points are easier to isolate in feature space. Clustering approaches like DBSCAN [5] can discover unusual patterns without requiring labeled data, while autoencoders and variational autoencoders learn compressed representations of normal market behavior and flag observations with high reconstruction error as potential anomalies [6], [7]. Recent work has extended these approaches to deep learning architectures, including Transformer-based models designed specifically for limit order book data and LSTM networks for sequential anomaly detection in time series. [9] Despite their sophistication, these methods remain challenging to deploy in production. They require heavy feature engineering, handle multimodal financial data poorly, and often produce signals that non-technical users cannot easily interpret. Validation is typically manual and domain-intensive, making it hard to distinguish meaningful market events from statistical noise.

C. LLM-Based Approaches

Simultaneously, multi-agent LLM frameworks have emerged as a powerful paradigm for complex financial tasks. The TradingAgents framework [11] simulates a professional trading firm with specialized agents for fundamental analysis, sentiment analysis, technical analysis, and risk management, demonstrating superior performance in cumulative returns and Sharpe ratio compared to traditional strategies. Park (2024) introduced an LLM-based multi-agent framework specifically designed for anomaly detection in financial markets, where agents collaborate on tasks including data conversion, expert analysis via web research, and institutional knowledge utilization [12]. These frameworks leverage the collective intelligence of specialized agents to navigate the complexity of financial markets more effectively than single-model approaches [11], [13].

This work introduces a novel agentic system that overcomes limitations of prior approaches by combining natural-language querying with automated integration of multi-source financial data. Users can request indicators such as unusual volatility or outsized returns in plain language, and the system constructs the corresponding cross-database queries and produces a comprehensive markdown report over a specified horizon (typically 90 days). The report includes intuitive statistical diagnostics, notably Z-score–adjusted anomaly probabilities for each signal, providing an interpretable summary of market behavior.

III. Method

A. System Architecture

This work develops a multi-agent system for market anomaly detection that translates natural-language requests into structured financial analyses. The system is implemented using the IBM Agentics framework [15], which enables agents to communicate through Pydantic-validated data models. Our architecture consists of three primary agent types:

1) **Natural Language Parser**: Extracts temporal parameters (e.g., 90-day windows), company identifiers, industry sectors, and anomaly indicators from user queries.
2) **Text2SQL Generator**: Produces database-specific SQL queries for each WRDS data source. This agent leverages Agentics' transduction mechanism to map natural-language intent into executable SQL while grounding the query in the actual schema of each database.
3) **Anomaly Detection Agent**: Retrieves the requested data, computes anomaly metrics across heterogeneous data sources, and synthesizes a composite anomaly score, integrating Bayesian confidence weighting.

B. Agentics Transduction for Text-to-SQL Generation

A core feature of our system is the use of *transduction*, the mechanism in Agentics by which natural-language inputs are transformed into structured programs—in this case, SQL queries. The key steps are:

1) **Schema Grounding**: We automatically generate structured schemas for each data source, specifying primary identifiers (`CUSIP`, `PERMNO`, `ticker`) and relevant fields for subsequent data assembly. These schemas are encoded as Pydantic models that serve dual purposes: validating Text2SQL agent outputs and providing metadata for a local SQLite index of all WRDS sources, which enables rapid schema lookup during query construction.
2) **Semantic Decomposition**: The agent converts the parsed intent (e.g., "90-day volatility", "highest momentum") into a structured intermediate representation reflecting required joins, filters, and temporal conditions.
3) **Constrained SQL Synthesis**: Using transduction, the agent generates SQL code that must satisfy:
   - the schema constraints (no hallucinated columns or tables),
   - the Pydantic output contract (query components must validate),
   - execution rules such as row limits, date clamping, and required WHERE clauses.
4) **Corrective Feedback Loop**: If the generated SQL fails validation or execution (e.g., no rows returned, invalid date ranges), the execution trace is fed back into the agent, which regenerates a corrected query. This loop is a defining property of Agentics-style transduction.

This approach ensures that natural-language instructions are reliably transformed into syntactically valid and schema-accurate SQL queries, while preventing hallucinations, a common issue in unconstrained LLM-based Text2SQL systems.

C. Data Sources and Integration

Our system integrates six complementary data sources, each providing distinct market perspectives:

WRDS Financial Databases:

- **CRSP**: Daily stock prices, returns, and trading volumes.
- **Compustat**: Balance sheet fundamentals, income statements, and cash flow data.
- **IBES**: Analyst earnings estimates, recommendations, and forecast revisions.
- **Capital IQ Key Developments**: Corporate events such as M&A activity, management turnover, and regulatory filings.

Contextual Data Sources:

- **Google Trends**: Search-volume signals related to companies or sectors.
- **Google News**: Recent media coverage and event relevance.

Each WRDS data source is accessed through its own Text2SQL agent with a dedicated schema map, with company identifiers harmonized across sources via `CUSIP`, `PERMNO`, and `ticker` symbol mappings. The Google-based contextual sources operate outside the Text2SQL pipeline, instead providing post-hoc anomaly inspection and validation. When the system flags a company with an elevated anomaly score, these sources contextualize the finding by identifying concurrent news events or attention spikes that may explain the detected deviation.

D. Anomaly Detection Framework

We define an anomaly as a deviation from expected company behavior within a specified temporal window (typically 90 days). For each data source, we compute a source-specific anomaly signal that is then combined into a single composite score.

For company i at time t, the overall anomaly score is:

$$A_{i,t} = \sum_{j=1}^{J} w_{ij} \cdot \phi_j(X_{i,t}^{(j)}), \tag{1}$$

where $X_{i,t}^{(j)}$ is the retrieved feature vector from source j, $\phi_j(\cdot)$ is the standardized anomaly measure for that source, and w_{ij} is its confidence weight.

a) CRSP (Market Activity): For CRSP, we compute a volatility-based anomaly using a standardized z-score:

$$\phi_{\text{CRSP}} = \frac{\sigma_{i,t}^{\text{window}} - \mu_i^{\text{hist}}}{\sigma_i^{\text{hist}}},$$

which captures how unusual the recent volatility is relative to the historical distribution. To provide interpretability, the z-score is then mapped to a probability-like scale (e.g., $|z| \leq 0.5 \to 0$, $|z| \approx 1.5 \to 0.33$, $|z| \approx 2.0 \to 0.5$, $|z| \geq 3 \to 1$).

b) CIQ Key Developments (Event Signals): CIQ provides event-level information such as management changes, M&A announcements, restructurings, or regulatory actions. From these events, we compute an anomaly score based on event density and recency:

$$\phi_{\text{CIQ}} = P(\text{abnormal event activity} \mid \text{event frequency in window}),$$

where the probability is derived from empirical baselines for typical event rates. Periods with unusually dense or high-impact events (e.g., multiple executive departures) produce higher scores.

c) IBES and Compustat (Fundamentals and Analyst Signals): For IBES and Compustat, which provide structured and event-driven signals, we compute a probability score directly:

$$\phi_{\text{IBES}} = P(\text{negative revision} \mid \text{surprise magnitude}),$$

$$\phi_{\text{Compustat}} = P(\text{fundamental shift} \mid \Delta\text{ratios}),$$

where probabilities are derived from empirical distributions of earnings surprises, estimate revisions, and ratio movements.

d) Adaptive Use of Available Data: Coverage varies significantly across firms. If only CRSP and CIQ are available, or if a company lacks IBES/Compustat data, the weighting mechanism automatically normalizes over the available sources:

$$w_{i,j} = \frac{\text{Precision}(i,j)}{\sum_{k \in \mathcal{A}(i)} \text{Precision}(i,k)},$$

where $\mathcal{A}(i)$ is the set of data sources actually available for firm i.

This ensures that the CRSP signal does not lead to over-inflated anomaly scores when only numerical signal is present, as is often the case with small-coverage firms. Conversely, for well-covered firms (e.g., large caps), the weights naturally shift toward multi-source confirmation.

This design ensures that the anomaly score is both interpretable and robust, adapting to the level of information available for each company.

Anomaly dimensions considered for the final composite score include the following:

- **Price-based (CRSP)**: Volatility spikes, abnormal returns, and deviations from sector-adjusted benchmarks.
- **Fundamental-based (Compustat)**: Earnings surprises, unexpected shifts in leverage, liquidity ratios, and other financial statement indicators.
- **Analyst-based (IBES)**: Recommendation downgrades, dispersion in analyst estimates, and negative earnings forecast revisions.
- **Event-based (CIQ Key Developments)**: Unusual density or clustering of corporate announcements such as executive departures, M&A activity, strategic initiatives, or regulatory actions.

E. Bayesian Confidence Weighting

Data availability and quality vary substantially across firms. To adjust for this heterogeneity, we assign confidence weights:

$$w_{ij} = \frac{\text{Precision}(i,j)}{\sum_{k=1}^{J} \text{Precision}(i,k)} \tag{2}$$

$\text{Precision}(i,j)$ reflects:

- **Data completeness**: Fraction of available fields in source j.
- **Trading activity**: Market liquidity and return frequency (CRSP).
- **Analyst coverage**: Number of analysts tracking firm i (IBES).
- **Event frequency**: Recency and density of reported corporate developments (CIQ).

Rather than a separate agent, these weights are applied directly within the anomaly computation pipeline.

F. Query Processing Workflow

The system operates through the following steps:

1) **Query Parsing**: User provides a natural-language specification (e.g., "Identify technology-sector firms with abnormal volatility and negative analyst sentiment in the last 90 days").
2) **Schema Mapping**: Agentics retrieves the relevant WRDS table schemas and constructs schema-aware Pydantic contracts for SQL generation.
3) **Parallel Data Retrieval**: Text2SQL agents issue validated SQL queries concurrently across all required databases.
4) **Anomaly Computation**: Source-specific anomaly measures are computed and aggregated using Bayesian confidence weights.
5) **Report Synthesis**: Results are compiled into a structured markdown report with visualizations and data source attributions.

All generated SQL queries undergo validation checks—schema consistency, date-range feasibility, and automatic row limits—to ensure robustness and computational efficiency.

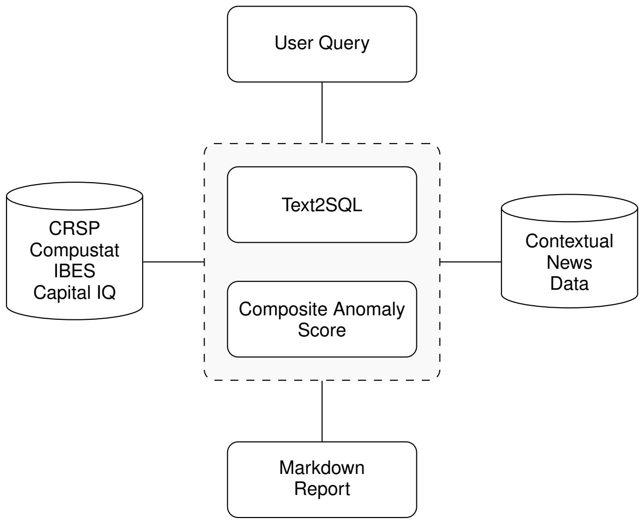

Fig. 1. Architecture of the Agentics-based market anomaly detection system

IV. Evaluation & Metrics

This section describes the end-to-end pipeline employed in our demonstration of the proposed agentic system. The workflow consists of three sequential components: (1) natural-language financial retrieval from the WRDS data via Text2SQL, (2) anomaly scoring based on the aforementioned data sources, and (3) contextual validation through news interrogation.

A. Natural-Language Query Interpretation & SQL Generation

We evaluated the system's ability to translate user queries into executable WRDS-compatible SQL statements. The following natural-language query was used: "**Which companies beat earning estimates last quarter?**"

The Text2SQL agent identifies the relevant IBES summary tables, extracts schema metadata from the local WRDS SQLite instance, and generates a syntactically valid SQL query without manual parameter specification. The resulting SQL statement is shown in **Listing 1** below:

```
SELECT ibes_eps_summary.ticker, ibes_eps_summary.
    actual_eps, ibes_eps_summary.consensus_estimate
FROM ibes_eps_summary
WHERE ibes_eps_summary.actual_eps > ibes_eps_summary
    .consensus_estimate
AND ibes_eps_summary.fiscal_period_end
BETWEEN DATE('2022-09-30', '-3 months') AND DATE('
    2022-09-30');
```

Listing 1. Earnings Estimate Query

The query output is returned as a structured dataframe and used as the candidate set for anomaly assessment.

B. Multi-Source Financial Anomaly Scoring

To illustrate downstream anomaly scoring, a ticker from the aforementioned SQL output, `002J`, was used, and a subsequent ticker with greater data coverage, `HEI`, was also evaluated.The anomaly score synthesizes deviations across multiple financial dimensions (market activity, fundamentals, analyst estimates, and corporate events), with Bayesian weighting applied to account for differences in data coverage and reliability.

1) Case Study: Ticker `002J` (IBES-only Coverage): Ticker `002J` corresponds to an IBES synthetic identifier with incomplete mappings to CRSP and Compustat. The system computed a *moderate* anomaly score of 0.49. This result reflects the following components:

- **Market Activity (CRSP):** A volatility z-score of -0.58, indicating statistically normal price behavior relative to historical data.
- **Analyst Estimates (IBES EPS):** Elevated uncertainty due to higher dispersion in forecast estimates, contributing meaningfully to the anomaly score.
- **Bayesian Weighting:** Reduced weight assigned to CRSP signals due to the absence of `PERMNO` and `GVKEY` mappings, resulting in analyst-based signals receiving greater influence.

This example demonstrates how the scoring framework adapts when certain data streams are unavailable, preventing over-reliance on incomplete information.

2) Case Study: Ticker `HEI` (Fully Mapped Across Databases): Ticker `HEI` represents a firm with complete mappings across CRSP, Compustat, IBES, and CIQ. The system produced a *high* anomaly score of 0.79, driven primarily by the market signal:

- **Market Activity (CRSP):** A volatility spike of approximately 2.5 standard deviations above the firm's historical distribution.
- **Analyst Estimates (IBES EPS):** Moderately elevated, contributing to the composite score but not dominating.
- **Cross-Database Reinforcement:** Fundamental and event-level data supported the elevated anomaly classification.

A system-generated anomaly summary for `HEI` is displayed in Fig. 2.

These examples highlight the framework's ability to integrate heterogeneous financial signals while adjusting for differences in data completeness, enabling nuanced anomaly assessments under varying information conditions.

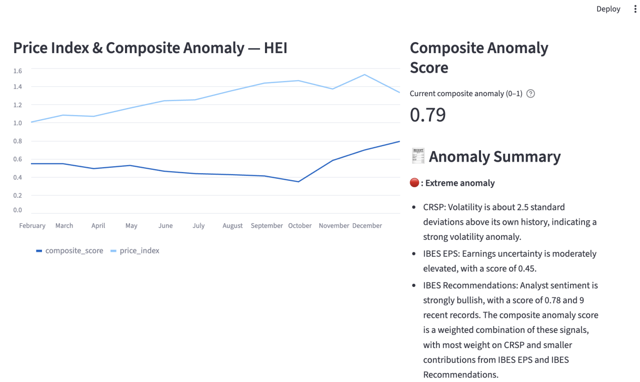

Fig. 2. System-generated anomaly summary for ticker `HEI`. (Placeholder image.)

V. Conclusion

This work presents an agentic system that unifies natural-language interaction, multi-source financial data integration, and interpretable anomaly detection within a modular multi-agent framework. By combining a Text2SQL agent that translates user prompts into structured WRDS queries with specialized detectors for market, fundamental, analyst, and event-based signals, the system provides an accessible yet rigorous interface for financial anomaly assessment. The approach advances existing methodologies by enabling plain-English specification of analysis goals, dynamically weighting heterogeneous data streams through a Bayesian confidence mechanism, and generating transparent natural-language reports that summarize and contextualize detected anomalies across multiple dimensions.

Several extensions offer promising directions for improving the robustness of our system. First, moving from Z-score–based heuristics to a fully Bayesian anomaly detection framework, that is, one that learns priors over each data stream and derives anomaly probabilities from posterior predictive distributions, would reduce sensitivity to heavy-tailed market behavior and eliminate the need for hard-coded anomaly thresholds. Our current implementation uses Z-score–based heuristics and fixed functional mappings to translate deviations into anomaly scores. While effective, these mappings embed implicit threshold assumptions and do not fully account for heavy-tailed market behavior. A fully Bayesian framework would offer a more flexible and principled alternative.

Second, expanding the system's capability to incorporate contextual and textual signals, such as Google News or regulatory filings, through embedding-based representations would yield a more holistic and multimodal understanding of atypical market activity. Finally, integrating temporal models and online learning mechanisms could enable real-time anomaly detection that adapts to evolving market regimes. Together, these avenues point toward a more comprehensive, probabilistic, and context-aware agentic architecture for financial market surveillance.

References

[1] D. C. Montgomery and G. C. Runger, "Applied Statistics and Probability for Engineers," 7th ed. New York, NY, USA: Wiley, 2018.

[2] P. C. Mahalanobis, "On the generalized distance in statistics," Proceedings of the National Institute of Sciences of India, vol. 2, pp. 49–55, 1936.

[3] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," Journal of Econometrics, vol. 31, no. 3, pp. 307–327, 1986.

[4] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), pp. 413–422, 2008.

[5] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD), pp. 226–231, 1996.

[6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, no. 6088, pp. 533–536, 1986.

[7] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in Proceedings of the 2nd International Conference on Learning Representations (ICLR), 2014.

[8] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," arXiv:1901.03407, 2019.

[9] S. Schmidl, P. Wenig, and T. Papenbrock, "Deep Learning for Time Series Anomaly Detection: A Survey," ACM Computing Surveys, vol. 57, no. 1, 2024.

[10] C. Zhang, Y. Mao, Y. Fan, Y. Mi, Y. Gao, L. Chen, D. Lou, and J. Lin, "FinSQL: Model-agnostic LLMs-based Text-to-SQL Framework for Financial Analysis," arXiv:2401.10506, 2024.

[11] Y. Xiao, E. Sun, D. Luo, and W. Wang, "TradingAgents: Multi-Agents LLM Financial Trading Framework," arXiv:2412.20138, 2024.

[12] T. Park, "Enhancing Anomaly Detection in Financial Markets with an LLM-based Multi-Agent Framework," arXiv:2403.19735, 2024.

[13] J. Li, Q. Zhang, Y. Yu, Q. Fu, and D. Ye, "More Agents Is All You Need," arXiv:2402.05120, 2024.

[14] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, and J. Wen, "A Survey on Large Language Model Based Autonomous Agents," Frontiers of Computer Science, vol. 18, no. 6, 2024.

[15] IBM Research, "Agentics: A Framework for Multi-Agent Systems with LLMs," GitHub repository, 2024. [Online]. Available: https://github.com/IBM/Agentics