SETSCI Conference Proceedings
25, 37–41, 2026

International Conference on Open Finance
September 24, 2026, Boston, MA, USA

https://doi.org/10.36287/setsci.25.7.001
2687-5527/ © 2026 The Authors. Published by SETSCI

Business-Aware Agentic Forecast Refinement for Enterprise Financial Planning

Hari Rajagopalan

Systems Architect, Stripe, LLC., Austin, TX, USA

Abstract – Enterprise financial forecasts are generated across multiple business dimensions such as products, regions, channels, customer cohorts, and segments. In a large enterprise, this can result in billions of populated intersections that finance teams cannot review manually before a planning deadline. This paper presents a proactive, human-supervised agentic AI workflow implemented in a financial-technology and payments planning environment. The process starts when a baseline forecast reaches a configured planning milestone. The agent reviews each granular trajectory, identifies relevant historical periods, considers business seasonality, selects an adjustment strategy, generates the correction, and validates the resulting forecast before proposing the adjustment to finance users. The implementation was evaluated in one quarterly forecast and one long-range planning cycle covering 11 dimensions, approximately 1.2 billion populated intersections, and more than 15 financial planners. Planners accepted 67.2% of the proposed adjustments. For the quarterly forecast, the organization's forecast-to-actual accuracy measure improved from approximately 96% to 98%. This workflow shifted planner effort from searching for anomalies to reviewing proposed corrections, applying business judgment, and collaborating on forward-looking assumptions while preserving finance authority over final decisions.

Keywords – Agentic AI; enterprise forecasting; FP&A; human-in-the-loop systems; anomaly detection; forecast adjustment; decision intelligence; financial technology

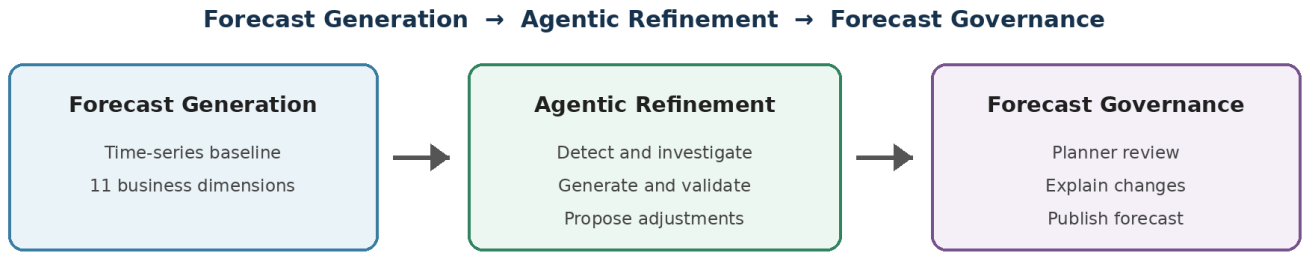


Fig. 1 High-level agentic forecast-refinement architecture

I. Introduction

Time-series forecasting provides finance teams with a consistent baseline for enterprise planning. This baseline is commonly generated across business dimensions such as products, regions, channels, customer cohorts, and segments. Grouped and hierarchical time series are useful because they allow the same business to be viewed through different levels of aggregation [1]. However, the number of populated intersections increases rapidly when several dimensions are combined. Finance teams can review whether the overall forecast aligns with company strategy, but it is not possible for them to inspect every granular trajectory before a planning deadline.

This creates an important gap in the forecasting process. The aggregate forecast can look reasonable while some of the deeper cuts contain noisy or unrealistic patterns. These patterns may be caused by unusual historical months, incomplete seasonality, or a temporary business event that is extrapolated into the future. The impact becomes larger when the planning horizon extends for several years. A noisy quarter in the historical data can influence multiple future periods even when the overall company forecast remains close to expectations.

Before this implementation, finance planners spent much of the first one to two weeks of a forecast cycle searching the baseline for these patterns. The adjustment process also required coordination across multiple teams. An adjustment made for a region could change a product or customer slice managed by another planner. This would require another correction and could result in repeated rounds of adjustment and reconciliation. Under deadline pressure, there was also a risk that teams would correct the overall number while leaving inaccurate behavior in the granular forecast.

This paper presents an industrial case study of an agentic refinement layer introduced into an existing enterprise forecasting process. The objective was not to replace the time-series model. The objective was to inspect the granular forecast automatically, prepare business-aware corrections, and provide finance teams with a controlled process to review the proposed changes. The paper makes three practical

contributions. First, it describes how proactive agents can inspect a forecast at enterprise scale without waiting for individual prompts. Second, it explains how historical trajectory, calendar events, and prior planner actions can be combined to select an adjustment strategy. Third, it reports the results from two live planning exercises while retaining human approval for every financial adjustment.

II. Background and Related Work

A. Multidimensional Enterprise Forecasting

Enterprise planning data is naturally grouped across several business attributes. A company may need the forecast by geography, product, customer type, channel, cohort, or combinations of these dimensions. Forecasting literature distinguishes between hierarchical structures, where lower levels roll into a defined parent, and grouped structures, where several alternative aggregation paths exist [1]. Forecast reconciliation methods are designed to produce coherent values across these structures [2]. The industrial problem examined in this paper is related but different. The baseline already exists within the planning environment. The remaining problem is to identify granular trajectories that are commercially implausible and prepare corrections without requiring planners to search every intersection.

Generic anomaly detection methods identify observations that differ from expected behavior [3]. In financial planning, the definition of expected behavior depends on the business being forecast. A large spike may indicate an error in one segment and a valid seasonal event in another. The same statistical pattern may therefore require different treatment depending on product characteristics, calendar timing, customer behavior, or organizational strategy. This makes contextual investigation important after a threshold identifies a trajectory requiring review.

B. Agentic AI and Human Review

Agentic AI systems combine language-model reasoning with tools, data access, and a sequence of actions intended to complete a defined task. Emerging financial-services applications show how multiple agent roles can support analysis, risk, and decision workflows [4], [5]. The important design question in financial planning is where the automation should stop. A deterministic rule can identify that a threshold was breached, but it cannot always determine which historical period is clean, whether a seasonal spike is expected, or whether a planner has forward-looking knowledge that is unavailable in the data.

Human-AI interaction research recommends that systems make their actions visible, support efficient correction, and allow users to dismiss or modify recommendations [6]. These principles are especially relevant to a forecast that may influence resource allocation and financial commitments. The implemented workflow therefore performs the repetitive inspection and calculation work while preserving planner control over the final forecast. The agent prepares an adjustment and explains the basis for the recommendation. The planner can then accept, modify, reject, or retain the existing forecast.

III. Business Problem and Design Objectives

The implementation supports enterprise revenue planning in financial technology and payments. A time-series process first forecasts business volume across 11 dimensions. Applicable rates are then used to calculate revenue or margin. The evaluated data contains approximately 1.2 billion populated intersections. The long-range planning horizon can extend for four years, which increases the impact of any unusual historical period that is carried forward by the baseline model.

The initiative was designed around five business objectives. The first objective was coverage. Every populated trajectory meeting the configured conditions should be available for inspection, even when finance users do not have time to review it manually. The second objective was proactive execution. The workflow should start when the latest baseline reaches a configured planning milestone instead of waiting for a planner to identify a problem and write a prompt. The third objective was business-aware correction. The agent should evaluate historical behavior and business events before deciding how the trajectory should be adjusted. The fourth objective was controlled writeback. Proposed changes should enter the planning environment as reviewable adjustments rather than silently changing the submitted forecast. The fifth objective was transparency. Finance users and central finance should be able to understand what changed and where the change occurred.

Table 1. Transformation of the forecast-refinement workflow

Stage	Workflow transformation
Inspection	Manual search of granular cuts → milestone-based agent scan
Investigation	Planner reconstructs history → agent reviews up to 24 months
Calculation	Manual propagation → agent-generated and validated trajectory
Decision	Manual entry → planner accepts, modifies, rejects, or retains baseline
Governance	Manual reconciliation → snapshot-based change explanation

IV. Agentic Forecast-Refinement Workflow

C. Milestone Trigger and Data Access

The process starts when a new baseline forecast reaches a configured planning milestone. The agentic layer accesses the latest forecast and historical actuals through a distributed SQL environment that includes Presto-based tables. Presto provides a distributed query layer across enterprise data sources and is suited to large analytical workloads [7]. The planning milestone gives the agent a defined version of the baseline to inspect and prevents the workflow from depending on ad hoc planner prompts.

Finance teams configure thresholds that identify trajectories requiring review. The threshold can reflect the expected behavior of a particular planning metric and can differ by business use case. A threshold breach creates an investigation item. It does not immediately write an adjustment. The agent first retrieves the available trajectory, historical periods, dimensional context, and prior planning information needed to determine whether the behavior is abnormal for that segment.

D. Historical Reference Selection

For each identified trajectory, the agent can review up to 24 months of historical data. It looks for relatively clean periods that can be used as reference points for the correction. This is important because simply selecting the most recent months can repeat the same noise that created the anomaly. The agent also

evaluates the direction and shape of the trajectory so that the selected reference period represents behavior that is more suitable for projection.

The historical investigation is performed separately for each relevant dimensional intersection. A product in one region may have a different growth pattern from the same product in another region. Similarly, a customer cohort may have a different seasonal pattern from the aggregate business. The purpose of the historical review is therefore to establish a relevant local reference rather than impose one company-wide growth assumption.

E. Business and Calendar Context

Business context is an important part of this process because unusual behavior does not always represent an incorrect forecast. In the payments industry, Black Friday and Cyber Monday can create large spikes in payment volume. Their calendar placement can change across years, with activity falling in November or extending into December. A generic statistical correction could smooth the spike and create a commercially incorrect forecast. The agent reviews the historical event pattern and the behavior of the relevant segment before recommending a change.

The agent can also use previous manual adjustments made by planners during the forecast cycle as additional context. These adjustments may show how finance has recently treated a region, product, or customer segment. This does not mean that the agent automatically repeats every manual adjustment. The information becomes one part of the context used to understand the trajectory. Forward-looking information that exists only with product, go-to-market, or finance teams is still handled during planner review.

F. Strategy Selection and Adjustment Propagation

After the reference periods and business context are reviewed, the agent determines which adjustment strategy is more suitable. In the implemented workflow, it can choose between month-over-month growth and year-over-year growth acceleration. Month-over-month growth is useful when the recent trajectory provides a stable basis for continuation. Year-over-year acceleration is useful when seasonality makes the relationship with the corresponding prior-year period more informative.

The agent also determines where the correction should begin. Once the first anomalous month is adjusted, the change is propagated through the remaining forecast. This prevents a one-month correction from creating an artificial discontinuity in the next period. Propagation is particularly important in the long-range planning exercise because the forecast may extend for four years. The agent calculates the revised path across the available horizon using the selected strategy and the chosen historical reference.

G. Validation and Planner Decision

The workflow includes a final validation after the proposed correction is generated. The agent rechecks the resulting trajectory and determines whether the revised values remain consistent with the historical and business context available to it. This step reduces the risk of producing an adjustment that fixes the first anomalous month but creates an implausible pattern later in the horizon. If the proposed path passes the final check, it is staged in the planning environment for finance review.

Finance users can accept the adjustment, modify it, reject it, or retain the existing baseline. This review is important when planners have forward-looking information that is not present in historical systems, such as a product launch, a commercial decision, a market constraint, or a change in go-to-market strategy. The agent completes the granular inspection and calculation work, while finance retains authority over the forecast that is ultimately submitted.



Fig. 2 Agent investigation, adjustment, and review sequence

H. Activity Explanation and Publication

A forecast activity explainer provides additional transparency into the planning process. It compares forecast snapshots and identifies what changed, where the change occurred, the size of the adjustment, and its impact on the consolidated forecast. During an active cycle, this information can be summarized for central finance as a daily digest of material movements and areas requiring attention. This is especially useful for globally distributed planning teams because it provides visibility into changes made across time zones.

Approved adjustments become part of the working forecast and can be re-evaluated as the cycle develops. Periodic snapshots may be published for internal reporting and management visibility. Once finance declares the cycle complete, the final adjusted forecast is published through the organization's established planning process. This creates a closed refinement loop while keeping the planning platform as the system of record.

V. Implementation and Evaluation

I. Evaluation Setting

The workflow was evaluated in two live enterprise planning exercises. The first was a quarterly forecast. The second was a long-range planning cycle with a horizon of up to four years. More than 15 financial planners participated across the two exercises. Both exercises used the existing baseline forecasting process, and the agentic layer operated as a refinement workflow on top of that baseline. The evaluation therefore reflects the combined operational process rather than a controlled replacement of one forecasting algorithm with another.

The number of identified anomalies differed by cycle. The initial cycle contained more anomalous behavior because the workflow was being applied to a baseline that had not

previously benefited from agent-supported refinement. As corrections were reviewed and incorporated, later iterations contained fewer recurring issues. Some proposed changes were intentionally ignored when finance users determined that the existing trajectory reflected information unavailable to the agent.

J. Planner Acceptance

Across the two exercises, finance users accepted 67.2% of the adjustments proposed by the agent. The remaining adjustments were rejected, modified, or replaced when planner knowledge supported a different trajectory. The acceptance rate demonstrates that the agent generated useful recommendations at scale. The rejected adjustments are also important because they show the value of retaining human judgment. In several cases, finance teams had real-world knowledge about organizational strategy, products, markets, or customer behavior that could not be inferred from historical data alone.

The acceptance measure should not be interpreted as a conventional model-accuracy metric. It represents the percentage of proposed operational adjustments accepted by planners across the combined cycles. The result captures both the usefulness of the generated recommendations and the role of planners in determining when the historical evidence should be overridden by forward-looking knowledge.

K. Forecast Accuracy

For the quarterly forecast, performance was evaluated after two to three months of actual results became available. The organization measures accuracy using the closeness between the forecast amount and the resulting actual amount. Based on this measure, accuracy improved from approximately 96% before refinement to approximately 98% after the agent-supported and human-approved adjustments. The improvement was more visible in granular cuts where noisy historical behavior had previously affected the baseline.

This result represents the outcome of the full planning process, which includes the existing time-series baseline, agent-generated adjustments, business context, and finance planner review. It should not be interpreted as a controlled benchmark of forecasting algorithms or as evidence that the same improvement will occur in every cycle. The implementation had two live evaluation exercises, and only the quarterly exercise had sufficient subsequent actuals available for the reported accuracy comparison.

L. Planner Workflow

The most important workflow improvement was the change in how planners used their time. Earlier, planners used the beginning of the cycle to search for missing links, misalignment, and noisy data patterns. The agent now performs this inspection proactively and provides proposed fixes for review. This allows planners to focus on evaluating the corrections and working with product and go-to-market teams to understand future cohorts and market assumptions.

These teams provide forward-looking information that is not available in the forecasting data. This is the area where finance professionals make their most valuable contribution to the process. Instead of spending the planning window finding every possible error, they can focus on whether the adjusted forecast reflects the organization's product strategy, go-to-market plans, operational priorities, and industry trends. The effort change was observed qualitatively; the implementation did not track a standardized number of refinement iterations or a complete planner-hour measure.

VI. Discussion

The case shows that the value of agentic AI in forecasting is not limited to generating another statistical forecast. The existing time-series model continues to provide the baseline. The agentic layer addresses a different problem: it reviews a large number of granular trajectories, determines which history is relevant, considers business context, selects a correction strategy, and prepares an adjustment for finance review. This extends the existing forecasting process without requiring the organization to replace the baseline model or remove planner authority.

The implementation also shows why anomaly identification and anomaly correction should be treated as connected but separate activities. A threshold is useful for identifying where attention is required. The correction depends on the trajectory, the historical reference period, the seasonality, the business event, and the period from which the change should be propagated. A normal script can execute a fixed calculation after a threshold is breached. The agentic workflow adds value by selecting among strategies and checking the resulting trajectory using the context available for that segment.

The 67.2% acceptance rate supports a balanced interpretation of automation. A majority of the proposed adjustments were useful to planners, which shows that proactive generation can reduce manual calculation at enterprise scale. At the same time, almost one-third were not accepted as proposed. These cases show that the human review is an intentional governance control. Product decisions, go-to-market plans, market expectations, and organizational priorities may be known to finance teams before they appear in historical data. A forecast-refinement system should make it easy for planners to apply this knowledge rather than treating disagreement as an error.

The activity explainer extends the same design into forecast governance. Multidimensional adjustments can create coordination problems because one planner's change may affect another planner's view. Snapshot comparison and daily summaries provide central finance with a way to monitor material movement without reconstructing every change manually. This is valuable near a forecast deadline, when the organization needs both speed and an explanation of how the consolidated forecast changed.

The approach can be extended beyond revenue forecasting. Demand, workforce, supply-chain, and other enterprise planning processes contain the same combination of high dimensionality, historical patterns, business exceptions, and human approval. The specific thresholds and adjustment strategies will differ by planning domain. The general workflow remains applicable: identify an exception, investigate the local history, apply business context, generate a correction, validate the result, and retain a human decision before publication.

VII. Limitations and Future Work

This study has several limitations. First, the evaluation covers two live planning exercises within one enterprise environment. The observed acceptance rate and accuracy improvement may not transfer directly to another organization, industry, or forecasting process. Second, the accuracy result

uses the organization's operational forecast-to-actual closeness measure. The study does not report a standardized statistical metric such as MAPE across all dimensional intersections. Third, the long-range planning exercise did not yet have sufficient actuals for a comparable outcome evaluation.

Fourth, the workflow depends on finance-configured thresholds and the quality of the historical and dimensional data available to the agent. Missing business context can still result in a recommendation that planners reject. Fifth, the study reports a qualitative shift in planner effort but does not provide a controlled time-and-motion analysis. The process is intentionally iterative, and a standardized number of refinement cycles was not tracked.

Future work will evaluate whether feeding accepted adjustments into subsequent baseline forecasts reduces recurring anomalies. Additional cycles will also make it possible to compare acceptance patterns by segment, strategy, and type of anomaly. Another area of work is to capture more planner rationale in a governed form so that recurring product, regional, and market judgments can improve future recommendations. Any expansion of autonomy should continue to preserve reviewability, adjustment history, and explicit finance approval.

VIII. CONCLUSION

This paper presents a business-aware agentic forecast-refinement workflow implemented in a large financial-technology and payments planning environment. The solution automatically inspects a baseline forecast across 11 dimensions and approximately 1.2 billion populated intersections. For each identified trajectory, the agent reviews up to 24 months of history, selects relevant reference periods, considers business and calendar context, chooses an adjustment strategy, propagates the correction through the planning horizon, and validates the revised trajectory before proposing an adjustment.

The workflow was used in one quarterly forecast and one long-range planning cycle involving more than 15 financial planners. Finance users accepted 67.2% of the proposed adjustments. For the quarterly exercise, the organization's forecast-to-actual accuracy measure improved from approximately 96% to 98% after agent-supported and human-approved refinement. The implementation also shifted planner attention from searching for granular errors toward reviewing proposed corrections and adding forward-looking knowledge from product plans, go-to-market strategies, organizational priorities, and industry trends.

The case demonstrates how agentic automation can extend an existing enterprise forecasting process while retaining human authority over financial decisions. The agent provides scale, consistency, and proactive investigation. Finance professionals provide the business judgment required to determine whether the proposed forecast is appropriate for the organization's future. The value comes from combining these capabilities in one controlled refinement process.

REFERENCES

[1] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne, Australia: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/

[2] S. L. Wickramasuriya, G. Athanasopoulos, and R. J. Hyndman, "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization," Journal of the American Statistical Association, vol. 114, no. 526, pp. 804–819, 2019.

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys, vol. 41, no. 3, article 15, 2009, doi:10.1145/1541880.1541882.

[4] I. Okpala, A. Golgoon, and A. R. Kannan, "Agentic AI systems applied to tasks in financial services: Modeling and model risk management crews," arXiv:2502.05439, 2025.

[5] M. Dimodugno and M. Mammadov, "AICrypto-Assistant: A multi-agent LLM platform for democratizing crypto-asset analysis," SETSCI Conference Proceedings, vol. 24, pp. 17–22, 2025, doi:10.36287/setsci.24.2.017.

[6] S. Amershi et al., "Guidelines for human-AI interaction," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, paper 3, pp. 1–13, 2019, doi:10.1145/3290605.3300233.

[7] R. Sethi et al., "Presto: SQL on everything," in 2019 IEEE 35th International Conference on Data Engineering, pp. 1802–1813, 2019, doi:10.1109/ICDE.2019.00196.