ROBUST ANOMALY DETECTION FOR PROCUREMENT PRICING UNDER DATA SPARSITY AND CONCEPT DRIFT

Martin Ndithi Ndeto

Abstract

Procurement pricing anomalies pose financial and governance risks for public and institutional organizations. Automated detection is challenging due to sparse historical observations, heterogeneous item categories, and non-stationary pricing dynamics driven by inflation and external supply shocks. Labeled fraud data is typically unavailable, limiting the applicability of supervised approaches. This paper examines a robust, semi-supervised approach to procurement price anomaly detection under these conditions. Category-conditioned baselines are estimated using robust statistics and adapted over time to account for distributional shift. Empirical analysis on real institutional procurement data indicates that static thresholds and non-robust estimators generate excessive false positives in sparse and volatile categories, while robust adaptive baselines improve stability and operational usability. The paper concludes by outlining limitations and research directions for governance-oriented anomaly detection systems.

1. Introduction

Procurement systems play a central role in institutional operations, managing transactions across diverse categories of goods and services. Pricing anomalies—arising from data entry errors, supplier opportunism, or procedural irregularities—can lead to financial loss and undermine trust in procurement processes. As procurement workflows become increasingly digitized, automated anomaly detection mechanisms are often relied upon to support audit and oversight activities.

Despite their importance, procurement environments pose structural challenges for anomaly detection. Historical data is frequently sparse for long-tail items, price distributions vary substantially across categories, and pricing behavior evolves over time due to inflation, currency fluctuations, and market disruptions. These characteristics violate the assumptions underlying many standard anomaly detection methods.

In practice, many procurement oversight systems rely on static rules or fixed statistical thresholds, such as deviations from historical averages. While straightforward to implement, such approaches are brittle under distributional shift and tend to produce high false-positive rates, imposing unsustainable burdens on human reviewers.

This work investigates whether robust and adaptive statistical methods can improve the stability and operational usability of procurement price anomaly detection systems under realistic constraints, emphasizing interpretability and deployment considerations over model complexity.

2. Related Work

2.1 Anomaly Detection

Anomaly detection has been widely studied across statistical, distance-based, and density-based paradigms, particularly in settings where labeled anomalies are scarce or unavailable.[1] Classical approaches often rely on distributional assumptions that are violated in sparse or heavy-tailed data regimes, leading to unstable behavior in practice.

2.2 Fraud Detection in Governance Contexts

Prior work on financial fraud detection highlights the importance of interpretability, false-positive control, and human-in-the-loop review in governance-oriented systems. However, much of this literature assumes access to labeled data or relatively stable distributions, conditions that are rarely met in institutional procurement environments.

2.3 Concept Drift

Non-stationarity and concept drift are well-documented challenges in deployed machine learning systems, motivating adaptive baselines and sliding-window approaches.[2] In procurement, drift arises both gradually (e.g., inflation) and abruptly (e.g., supply-chain disruptions), complicating the use of fixed historical references.

3. Problem Formulation

We consider a stream of procurement transactions, where each transaction is represented as:

$$x=(c,p,q,v,t)$$

with c denoting the item category, p the unit price, q the quantity, v the vendor, and t the transaction time.

For each category $\boldsymbol{c}$, historical prices form a time-indexed distribution $\boldsymbol{P_c(t)}$**.** The objective is to assign an anomaly score to each transaction based on its deviation from expected category-specific price behavior, subject to the following constraints:

- No labeled anomalies are available
- Category-level sample sizes vary widely
- Price distributions evolve over time

The goal is to flag potentially anomalous prices while maintaining low false-positive rates and preserving interpretability for downstream human review.

4. Data Description

4.1 Dataset Overview

The dataset consists of historical procurement transactions collected over multiple years from an institutional procurement system. Transactions span a wide range of item categories, including consumables, equipment, and services. Category frequencies exhibit a long-tailed distribution, with many categories appearing infrequently.

4.2 Data Challenges

Three dominant challenges characterize the data:

- Sparsity: Many categories have limited historical observations, making mean-based estimates unstable.
- Heterogeneity: Price scales vary by orders of magnitude across categories.
- Non-stationarity: Temporal analysis reveals gradual inflationary trends as well as abrupt shifts during external disruptions.

These characteristics motivate the use of robust and adaptive detection methods.

5. Methodology

5.1 Baseline Approaches

Two commonly used baselines are evaluated:

- Static thresholds: Fixed percentage deviations from historical averages.
- Z-score detection: Normalization using category-level means and standard deviations.

Both approaches assume stable distributions and are sensitive to extreme values.

5.2 Robust Estimation

To mitigate sensitivity to extreme values and data sparsity, category-level baselines are estimated using the median and median absolute deviation (MAD), which are more stable under heavy-tailed distributions.[3] Prices are normalized relative to these robust statistics rather than mean-based estimates.

5.3 Adaptive Thresholding

To account for concept drift, statistics are computed over rolling historical windows rather than fixed time horizons. Thresholds are periodically recalibrated, allowing the system to adapt to gradual distributional changes while retaining sensitivity to localized anomalies.

5.4 System Design

The detection pipeline operates in batch mode, periodically updating category-level statistics and scoring incoming transactions. Flagged anomalies are forwarded to domain experts for review, supporting a human-in-the-loop workflow.

6. Evaluation

6.1 Evaluation Strategy

In the absence of labeled anomalies, evaluation emphasizes comparative stability, false-positive behavior, and expert review rather than absolute detection accuracy. This aligns with the operational objectives of procurement oversight systems.

6.2 Metrics

Evaluation considers:

- Proportion of transactions flagged
- Variance of flag rates across categories
- Consistency of flagged cases under temporal shifts

6.3 Baseline Comparison

Robust adaptive methods are compared against static and non-robust baselines across both dense and sparse categories.

7. Results

Static thresholds and z-score methods produced high flag rates in volatile and sparse categories, overwhelming reviewers with false positives. Robust estimators reduced spurious flags by stabilizing category-level baselines. Adaptive thresholding further improved temporal consistency, maintaining stable flag rates despite distributional drift.

8. Discussion

8.1 Practical Implications

The results highlight the importance of robustness and adaptivity in governance-oriented anomaly detection systems. Controlling false positives is critical for sustaining human review workflows and maintaining institutional trust.

8.2 Limitations

This study abstracts away inter-vendor relationships and coordinated bidding behavior, limiting its ability to detect higher-order procurement anomalies. Additionally, expert review introduces subjectivity into evaluation.

8.3 Future Research Directions

Future work includes graph-based modeling of vendor interactions, Bayesian change-point detection, and uncertainty-aware decision support systems for procurement governance.

9. Conclusion

This paper examined a robust, semi-supervised approach to procurement price anomaly detection under data sparsity and concept drift. Empirical analysis indicates that adaptive robust baselines outperform static methods in terms of stability and operational usability. Overall, the results highlight the importance of domain-aware, interpretable anomaly detection methods for governance-oriented machine learning systems.

10. Ethics and Governance Considerations

Machine learning systems deployed in procurement oversight can have downstream reputational and organizational consequences. False positives must be managed carefully, and anomaly detection systems should remain transparent, auditable, and embedded within human decision-making processes. Prior work emphasizes the need to consider such sociotechnical impacts in institutional ML deployments.[4]

Status Note

This manuscript documents exploratory research derived from a real-world procurement system and is shared to provide context on problem formulation, methodology, and evaluation under practical constraints. The work has not yet been submitted for peer review and continues to evolve, including extensions to more expressive models and additional evaluation.

References

1. *Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys.*
2. *Gama, J., Žliobaitė, I., et al. (2014). A survey on concept drift adaptation. ACM Computing Surveys.*
3. *Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. Journal of the American Statistical Association.*
4. *Selbst, A. D., et al. (2019). Fairness and abstraction in sociotechnical systems. FAT*.*