

REGIME-AWARE REVENUE PREDICTION

Kelly Chen, Marc Last, Maddie Wilson – Oracle – Columbia University

A. Abstract

Retailers face the difficult task of predicting merchandise-class revenues, where consumer demand varies across price tiers and is strongly influenced by macroeconomic conditions. Traditional forecasting often assumes stability, leading to poor accuracy during regime shifts. We develop a **regime-informed machine learning pipeline** to forecast jewelry sales across **low-, mid-, and high-tier** segments. The framework integrates **three modules: feature selection (RFE with Random Forests), regime classification (rolling PCA plus regime-adapted Hidden Markov Models), and forecasting (Linear Regression, Random Forest, XGBoost)**. By embedding macroeconomic context, our approach improves accuracy, interpretability, and business relevance, offering a scalable tool for merchandise optimization under changing economic regimes.

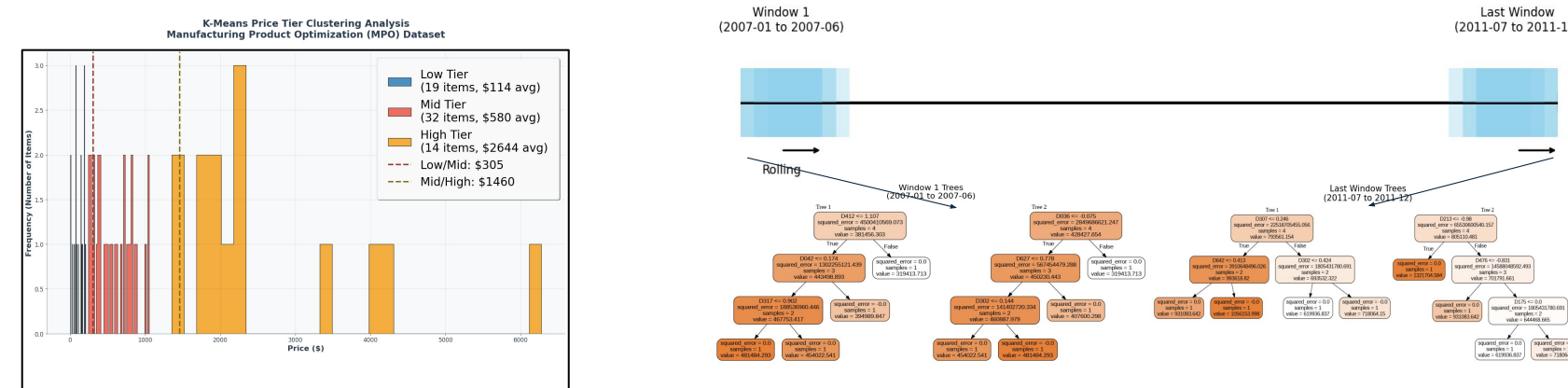


B. Module 1: Indicator Selection

Forecasting at the merchandise-class level requires filtering hundreds of macroeconomic indicators to find the true drivers of revenue. Using the **RFE with Random Forests** framework of Hirsra et al. (2022), we systematically reduce this universe to the 20 most predictive features per tier. Distinct indicator sets emerged for low-, mid-, and high-tier segments, revealing heterogeneous sensitivities and providing a scalable foundation for regime detection.

Approach:

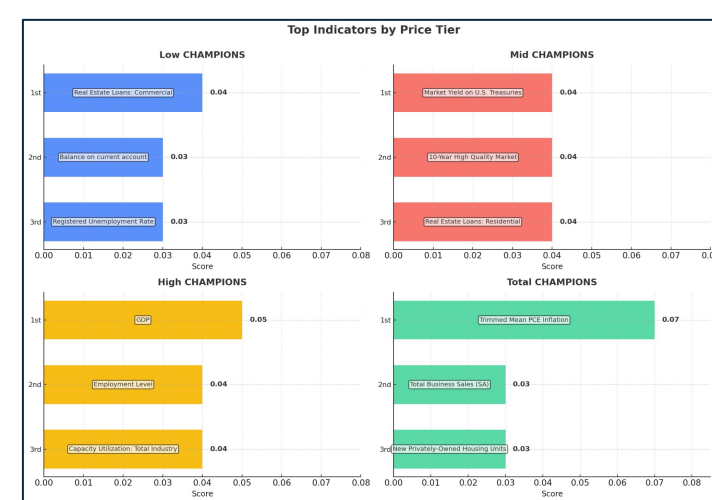
- Used **k-means clustering** on average unit price to create low-, mid-, and high-tier merchandise classes.
- Applied Random Forest + RFE to identify the top 20 indicators per tier.



- Introduced a 6-month rolling window to stabilize importance rankings and mitigate overfitting.

Outputs:

- Compact sets of **20 macroeconomic drivers** per tier.
- Indicator lists feed directly into **Module 2 (Regime Detection)**.



C. Module 2: Regime Detection

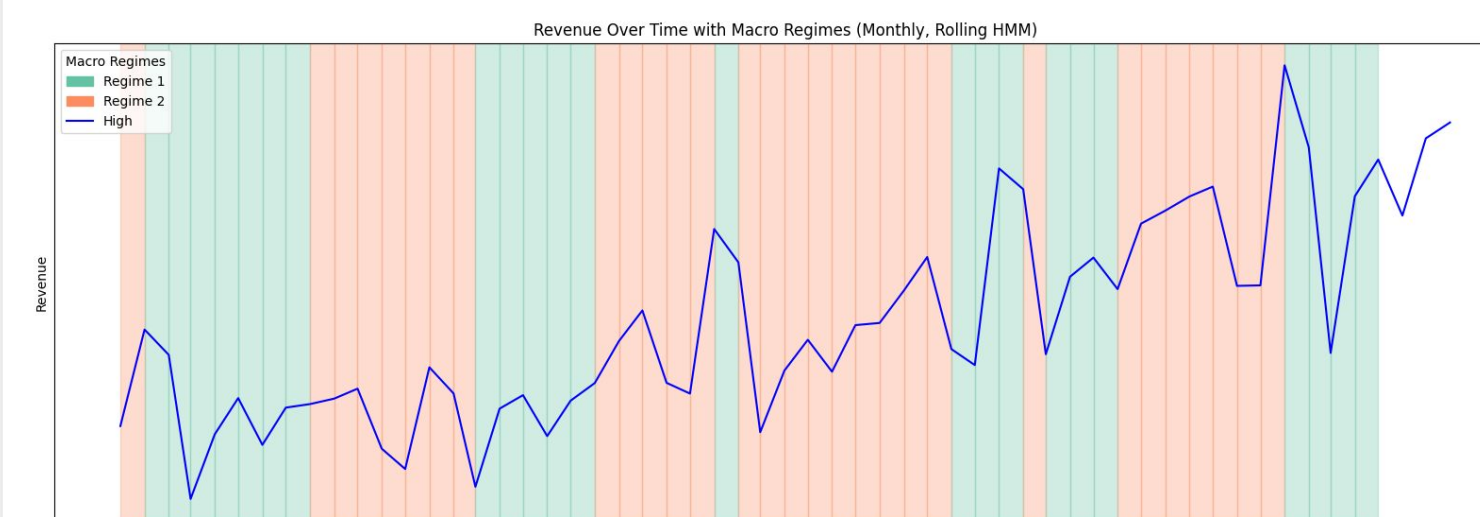
To forecast effectively in a volatile environment, we used a **Gaussian Hidden Markov Model (G-HMM)** to classify the historical macroeconomic landscape. This approach segmented economic history into **two interpretable regimes**: a Stable Growth regime and a Transitional/Volatile regime. This classification is a crucial input for our forecasting models, allowing them to adapt to different economic cycles.

Approach:

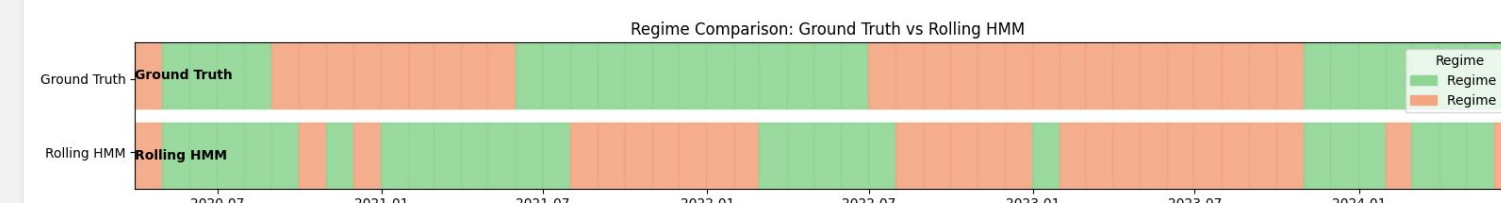
- Applied **Robust Rolling PCA (R2PCA)** to compress macro indicators into evolving principal components.
- Applied Gaussian Hidden Markov Model (G-HMM) to these PCs to identify unobserved latent states, or regimes, in the time series.

Outputs:

- Monthly regime labels** aligned with tier-level revenues.



- Transition probabilities** describing real-time shifts between regimes, passed to Module 3 for regime-aware forecasting.

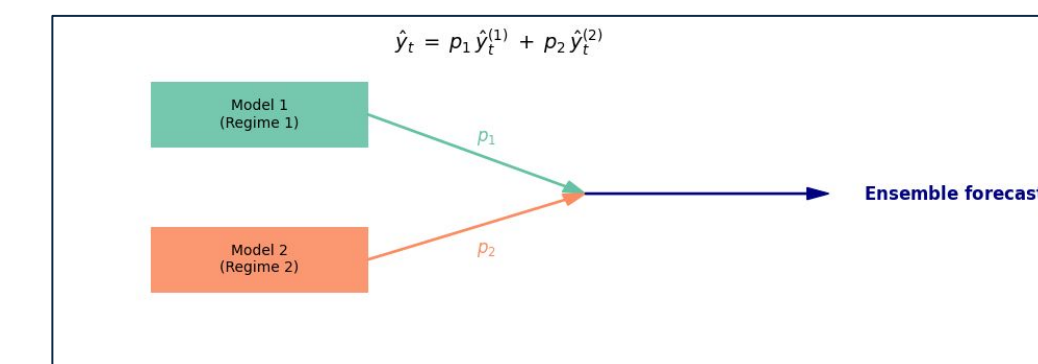


Key Findings & Innovations:

- Temporal Stability:** Our method ensures stable, aligned-with-real-world economic cycles regimes by avoiding the "flip-flopping" issue of traditional HMMs.
- Dynamic Adaptability:** Our framework adapts to market changes, identifying the emergence of new regimes not present in historical data, which is crucial for forward-looking forecasting.

D. Module 3: Revenue Prediction

At the item (SKU) level, forecasting models (**Linear Regression, Random Forest, XGBoost**) are trained **separately** within each regime, so parameters adapt to regime-specific dynamics. Forecasts are then combined using regime transition probabilities: outputs from each regime-specific model are **weighted by the likelihood** of that regime materializing next. This probabilistic ensemble adapts dynamically to macroeconomic shifts and improves accuracy around turning points.

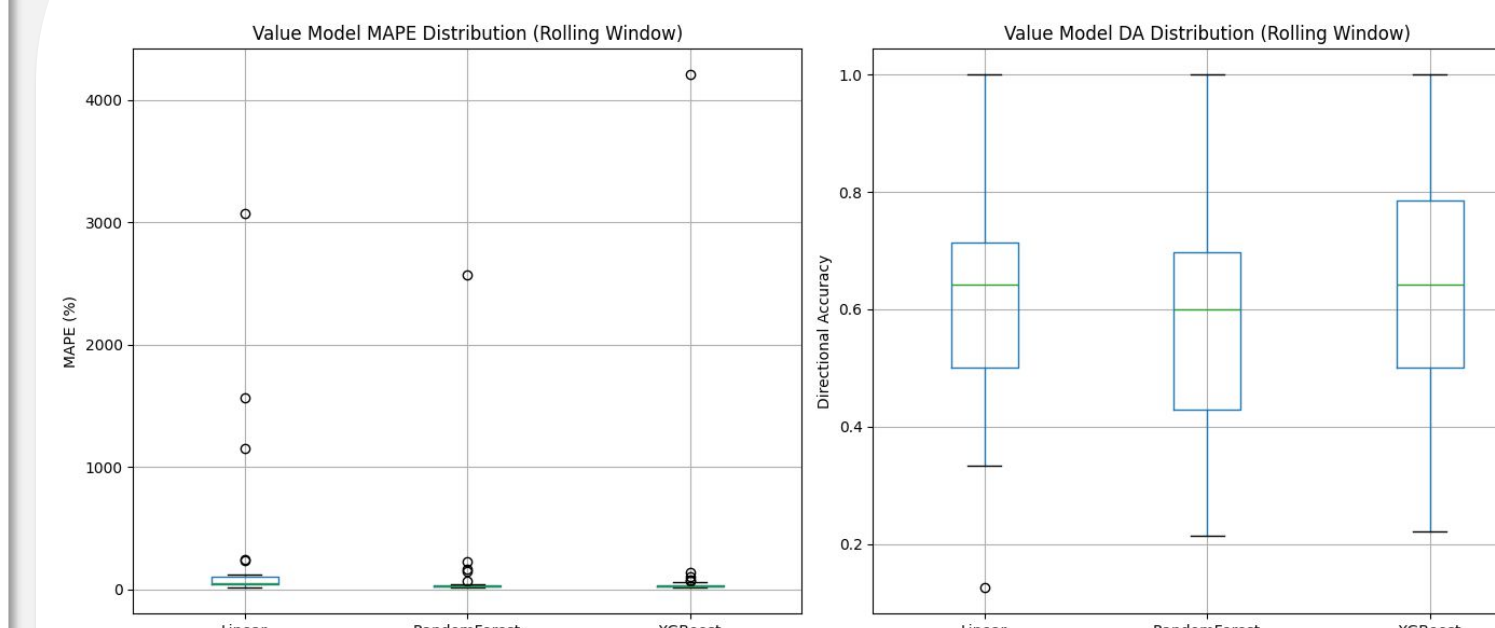


Outputs:

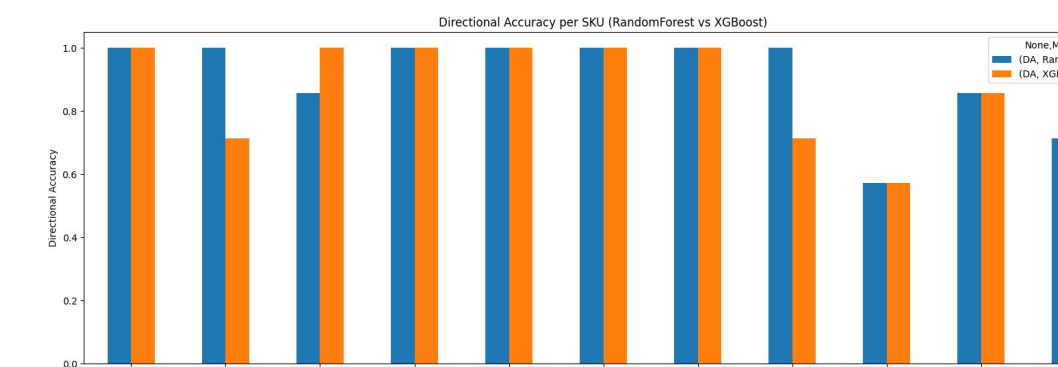
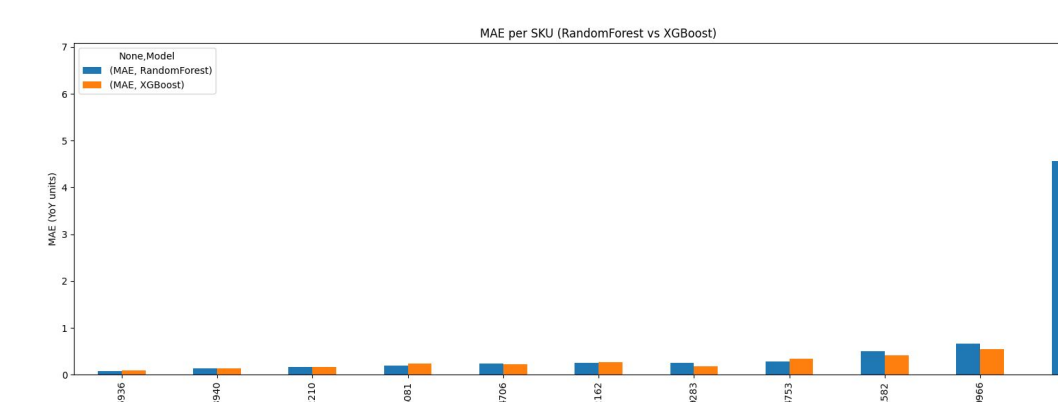
- SKU-level revenue forecasts aligned with current and forward regime probabilities.
- Improved **predictive accuracy** versus static baselines, especially during regime shifts.
- Actionable **guidance for merchandise planning** across low-, mid-, and high-tier segments

E. Results - Model Performance

Metrics for revenue prediction



YoY growth prediction on high tier per item



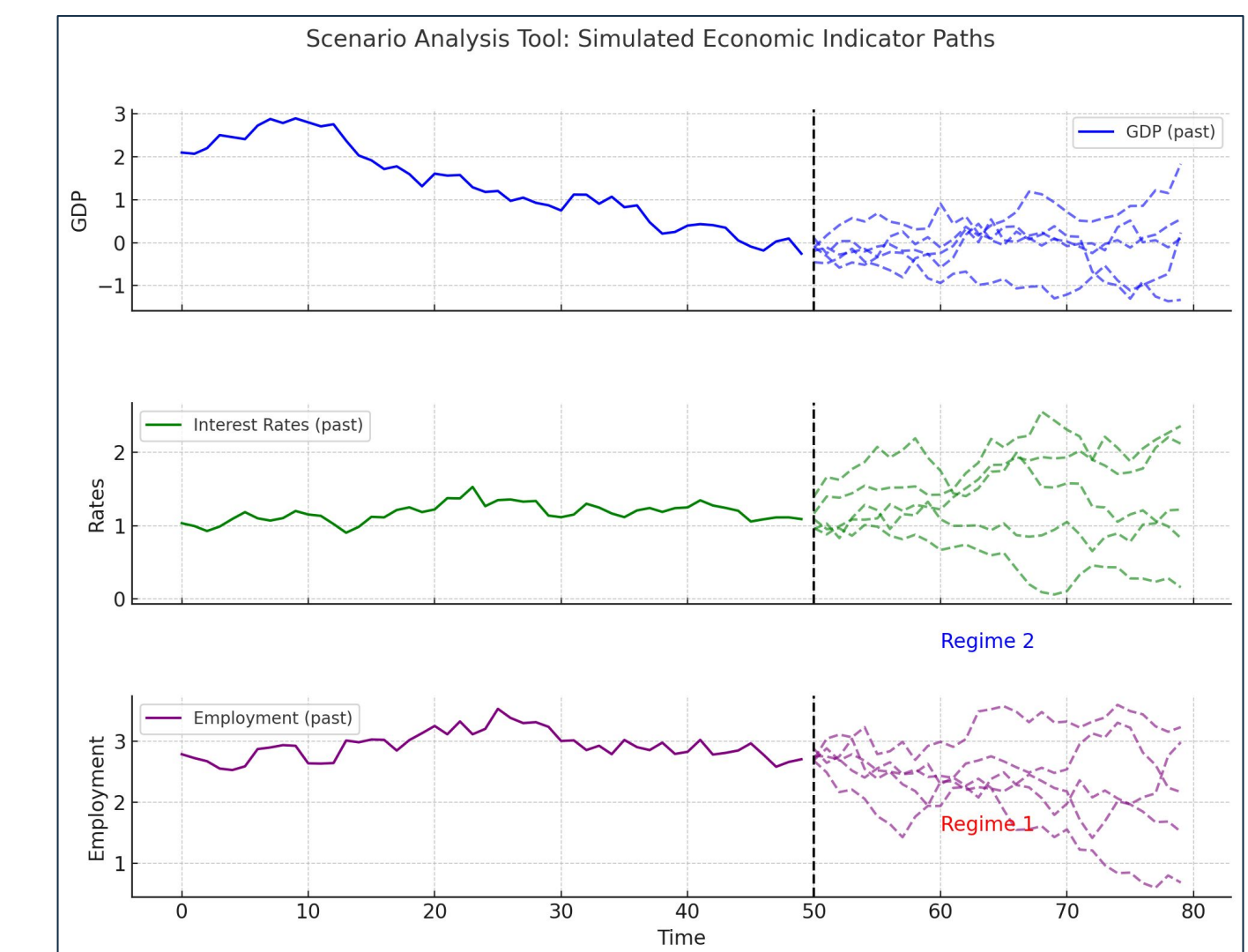
F. Conclusions / Future Work

Our regime-informed forecasting framework demonstrates how macroeconomic conditions shape revenue dynamics across merchandise tiers. By combining systematic **indicator selection, robust regime detection, and probabilistic forecasting**, the pipeline provides retailers with interpretable, actionable predictions at the item level.

Future Work

Looking ahead, we are extending the pipeline with **simulation of macroeconomic indicators**. We are exploring advanced methods like the **Robust Rolling Regime Detection (R2RD)** to ensure our model adapts to new data without look-ahead bias, as well as developing a bootstrap simulation method to preserve cross-correlations in the data.

This enhancement will allow us to study how hypothetical shifts in interest rates or inflation would alter regime classification. The result is a **scenario-planning tool** where retailers can test "what-if" conditions and anticipate how their business will evolve under different macroeconomic environments. This capability bridges rigorous econometric modeling with practical decision-making for revenue planning.



10. Acknowledgments

We are deeply grateful to **Bata Batjargal (Oracle)** for his guidance and mentorship, which shaped the direction of this project and made a lasting impact on its outcome. His perspectives were key to our progress. We also wish to thank **Professor Ali Hirsra** and **Teaching Assistant Miao Wang** for their steady support and thoughtful supervision.

11. References

- Hirsra, A., Xu, S., & Malhotra, S. (2024). Robust Rolling Regime Detection (R2-RD): A Data-Driven Perspective of Financial Markets.
- Karatas, T., & Hirsra, A. (2021). Two-Stage Sector Rotation Methodology Using Machine Learning and Deep Learning Techniques. arXiv