MacroTone: A Regime-Based Factor Allocation System Using Macroeconomic and NLP-Derived Signals

Vidit Pokharna, Devang Ajmera, Osho Sharma

Fall 2025

Summary

This project develops **MacroTone**, a fully reproducible, end-to-end regime-based portfolio allocation system that fuses macroeconomic indicators with Natural Language Processing (NLP)-derived sentiment from Federal Reserve communications. The objective of the system is to dynamically rotate among U.S. equity factor portfolios while maintaining a 15% annualized volatility target.

Using the most recent full pipeline execution, MacroTone achieves the following performance metrics:

- Annualized Return: 11.37%
- Annualized Volatility: 5.86%
- Sharpe Ratio (Excess): 1.87
- Maximum Drawdown: -4.92%

The full system integrates automated data ingestion from FRED, the Fama-French data library, and a custom FOMC web crawl. These inputs feed into a FinBERT sentiment scoring engine, an ensemble of Ridge and XGBoost forecasting models, a deterministic backtesting engine with transaction cost modeling, a comprehensive diagnostics suite, and a fully interactive Streamlit dashboard.

This macro and NLP fusion captures regime transitions that static factor allocations consistently fail to adapt to. By explicitly conditioning allocations on both economic structure and central bank communication, MacroTone produces more resilient portfolios during periods of policy and macroeconomic transition. Every statistic reported in this study is fully traceable through a transparent engineering pipeline that links FinBERT-derived tone scores (`data/interim/nlp_regime_scores.parquet`) to model forecasts (`data/processed/preds.parquet`) and finally to the live dashboard (`ui/app.py`).

1 Introduction and Motivation

Factor investing has demonstrated strong empirical performance over long horizons. However, a central limitation of traditional factor strategies is that they rely on static exposures that do not adapt to changing macroeconomic conditions. A large body of empirical evidence shows that factor premia vary substantially across regimes defined by inflation, growth, and financial conditions. As a result, static allocations routinely experience sharp drawdowns during regime transitions.

Factor investing has demonstrated strong empirical performance over long horizons. However, a central limitation of traditional factor strategies is that they rely on static exposures that do not adapt to changing macroeconomic conditions. A large body of empirical evidence shows that factor premia vary substantially across regimes defined by inflation, growth, and financial conditions. As a result, static allocations routinely experience sharp drawdowns during regime transitions.

MacroTone is designed to directly address this structural weakness by integrating three core components into a unified allocation system:

1. A macroeconomic regime detection system based on FRED time series
2. NLP-derived monetary policy tone extracted from FOMC communications, following the methodology of recent research on Dynamic Factor Allocation Leveraging Regime-Switching Signals[1] and analyzing the Fed's monetary policy statements[2]
3. Machine learning models that map detected regimes to forward factor returns
4. A volatility-targeted portfolio construction engine with realistic transaction costs
5. A fully automated and reproducible research and deployment pipeline

The entire project is reproducible using a single command: `PYTHONPATH=src uv run make quickstart`. This command orchestrates all data ingestion, NLP scoring, dataset construction, model training, backtesting, diagnostics generation, and dashboard launch under deterministic random seeds. This ensures that every experiment can be repeated exactly and audited from raw inputs to final performance outputs.

2 Literature Review and Conceptual Foundations

MacroTone is grounded in three complementary strands of academic literature that together motivate its hybrid macroeconomic and textual regime modeling framework.

2.1 Regime-Based Factor Allocation

The paper *Regime-Based Factor Allocation* demonstrates that the payoffs to equity factors are not stable over time, but instead change sign and magnitude across macroeconomic regimes defined by real growth, inflation, and liquidity conditions. The paper shows that static factor tilts systematically destroy risk-adjusted performance when regimes shift, because exposures that are rewarded in one regime can become severely penalized in another.

MacroTone directly implements this insight through the construction of a Macro Regime Index (MRI). The MRI is built using principal component analysis on CPI, GDP, Industrial Production, term spreads, and corporate credit indicators sourced from FRED. These components jointly represent the underlying macro structure of growth, inflation, and liquidity. The resulting MRI features

[1] `https://arxiv.org/abs/2410.14841`

[2] `https://www.banorte.com/cms/casadebolsabanorteixe/analisisyestrategia/analisiseconomico/mexico/20220727_NLP_Statements_Research_Note_Final.pdf`

enter both the predictive feature panel (`data/processed/panel.parquet`) and the scenario simulator in the UI, ensuring that portfolio allocations are explicitly conditioned on the detected macro regime at every point in time.

2.2 NLP of Central Bank Communications

The *NLP Statements Research Note (20220727)* establishes that the tone of Federal Reserve policy communications contains incremental predictive information for Treasury yields and equity returns beyond what is captured by contemporaneous macroeconomic releases. The underlying mechanism is forward guidance, where markets respond to verbal policy signals before economic data reflect those expectations.

MacroTone extends this finding by constructing a fully automated FOMC text ingestion and sentiment scoring pipeline. The crawler collects all available FOMC statements, minutes, and press conferences, producing a corpus of 683 documents spanning 1994 through 2025. Each document is segmented into 512-token chunks for FinBERT processing, and sentiment scores are aggregated to the monthly frequency. The resulting NLP regime features are stored in `data/interim/nlp_regime_scores.parquet` and act as a second, orthogonal regime axis alongside the MRI.

MacroTone adopts this hybrid modeling philosophy by feeding NLP-derived regime scores directly into both the Ridge and XGBoost forecasting models alongside structured macroeconomic features. This allows the ensemble to learn non-linear interaction effects between macro structure and policy language. This design elevates MacroTone from a conventional macro-factor model to a fully multi-modal financial learning system.

3 Data Engineering and Signal Construction

3.1 Macroeconomic Data

Macroeconomic indicators are sourced from the Federal Reserve Economic Data (FRED) system and cached locally to ensure deterministic reproducibility across runs. The dataset includes inflation measures (CPI and PCE), real activity indicators such as GDP growth and Industrial Production, yield curve term spreads, and corporate credit conditions. All raw series are standardized, lag-aligned, and transformed into stationary representations prior to dimensionality reduction.

To construct a compact representation of the macroeconomic environment, principal component analysis (PCA) is applied to the standardized macro feature set. The first three components are retained and interpreted as growth, liquidity, and inflation regimes. These components form the Macro Regime Index (MRI). The standardized macro inputs together with the extracted MRI components are saved to `data/interim/macro_features.parquet`.

These MRI components serve two critical roles in the system. First, they enter directly into the machine learning feature panel used for forecasting factor returns. Second, they are exposed as interactive controls within the Streamlit scenario simulator, allowing users to apply macroeconomic shocks and observe the resulting portfolio reallocations in real time.

3.2 Market Factor Data

Monthly factor returns are fetched automatically from the Fama-French data library. The pipeline retrieves the three core equity factors (Market Excess Return, SMB, and HML), the Momentum factor (UMD), and the risk-free rate (RF). The latest pipeline execution confirms the following dataset properties:

- 1,192 monthly observations
- Date range from July 1926 through October 2025
- Zero missing values across all factor series

The full factor dataset is stored at `data/raw/ff/ff_monthly.parquet`

These factor returns serve two distinct functions within MacroTone. First, the future realizations of the factor returns act as prediction targets during the supervised learning phase. Second, the Market factor and risk-free rate jointly provide benchmark excess return comparisons during the backtesting stage. After macroeconomic and NLP-derived features have been engineered, the pipeline executes `macrotones.features.dataset` to merge all signals and factor targets into the unified modeling dataset `data/processed/panel.parquet`.

This ensures that the learning system is trained on a fully synchronized macro, NLP, and market factor feature space.

3.3 FOMC NLP Corpus

The Federal Open Market Committee (FOMC) text corpus is assembled through a custom multi-source web crawler that scans official Federal Reserve domains for policy statements, minutes, and press conference transcripts. During the most recent refresh, the crawler discovered 489 candidate links and retained 18 newly valid documents after date validation. This brings the full corpus to 683 documents spanning the period from 1994 through 2025.

Each document is segmented into 512-token blocks to respect the input constraints of the FinBERT transformer model. Sentiment scores are computed for each block and aggregated to month-end dates to form a structured monthly tone matrix of dimension `249 months × 25 tone features`.

The resulting NLP regime dataset is saved as `data/interim/nlp_regime_scores.parquet`.

This tone matrix is explicitly aligned to month-end timestamps and merged with the MRI components during dataset construction to form the final 21-feature modeling panel. The NLP regime features also directly generate the regime-conditioned performance summaries exported to `data/processed/regime_summary.csv` and `data/processed/regime_boxplot.png`.

To improve computational efficiency, all FinBERT scores are cached. Unless new FOMC documents are detected during subsequent crawls, the NLP pipeline reuses the existing sentiment features and avoids recomputation. This caching mechanism materially reduces end-to-end runtime while preserving deterministic reproducibility.

Together, the macroeconomic, market factor, and NLP pipelines form the complete signal extraction layer of MacroTone. Each stage writes intermediate artifacts to disk, enabling full traceability from raw data through to final portfolio decisions.

4 Feature Panel and Dataset Assembly

All macroeconomic, NLP, and market factor signals are merged into a single unified modeling dataset saved as `data/processed/panel.parquet with shape (240 × 21)`.

This panel represents the core learning and backtesting substrate of the MacroTone system and is designed to align regime information, predictive signals, and market outcomes at the monthly frequency.

4.1 Panel Contents and Structure

The feature panel is composed of the following major signal blocks:

- **Macro Regime Index (MRI) components**: The first three PCA components extracted from standardized FRED macroeconomic series, interpreted as growth, inflation, and liquidity regimes.
- **NLP-derived tone features**: Monthly FinBERT sentiment aggregates derived from the FOMC corpus, capturing forward-looking policy tone across multiple semantic dimensions.
- **Lagged factor returns and volatility controls**: Historical realizations of Market, SMB, HML, and UMD returns and rolling volatility estimates that condition the predictive models on recent market dynamics.
- **Cash state indicator**: A binary control feature that enables the portfolio construction layer to allocate to cash under extreme volatility or elevated risk conditions.

The dependent variables for supervised learning are the forward one-period-ahead factor returns. These forward returns serve as the prediction targets used by the Ridge and XGBoost models during training.

4.2 Data Provenance and Assembly Process

The panel is constructed programmatically by executing:

```
uv run python -m macrotones.features.dataset
```

This dataset builder merges the following upstream artifacts:

- Standardized macro features and MRI components from `data/interim/macro_features.parquet`
- Monthly NLP sentiment regime scores from `data/interim/nlp_regime_scores.parquet`
- Fama-French factor returns and risk-free rates from `data/raw/ff/ff_monthly.parquet`

All inputs are aligned on month-end timestamps, forward-shifted appropriately for prediction targets, and filtered for complete overlap across all feature sources. This strict synchronization ensures that each observation represents a fully observed regime state paired with its corresponding future return outcome.

4.3 Downstream Usage and Consumers

The assembled feature panel feeds directly into every downstream modeling and analytics component of MacroTone:

- **Model Training**: `macrotones.models.train` consumes the panel and produces predictive factor return forecasts saved to `data/processed/preds.parquet`
- **Backtesting Engine**: `macrotones.backtest.engine` uses both the panel and the predictions to generate portfolio equity curves and turnover statistics, exporting `data/processed/backtest.parquet` and `data/processed/turnover_rolling.parquet`
- **Regime Analytics**: The panel underlies regime-conditioned performance summaries written to `data/processed/regime_summary.csv` and `data/processed/regime_boxplot.png`
- **User Interface Visualizations**: All KPI dashboards, regime attribution displays, and scenario simulator outputs in `ui/app.py` reference derived quantities that ultimately originate from this panel.

4.4 Quality Control and Regime Coverage Validation

To ensure stability and completeness across economic regimes, multiple diagnostics validate the feature panel. Regime-conditioned return distributions are inspected through `data/processed/regime_boxplot.png`.

This visualization confirms that each macro and NLP regime bucket contains sufficient observations to support statistical inference. Additional internal coverage checks ensure that no regime combination is sparsely populated due to misalignment or data gaps. Together, these diagnostics confirm that the modeling dataset achieves balanced representation across distinct macroeconomic and policy sentiment environments.

In summary, `panel.parquet` serves as the single source of truth for all predictive modeling, backtesting, regime diagnostics, and UI-driven scenario analysis within the MacroTone platform.

5 Modeling Architecture

Two predictive models are used to map macroeconomic and NLP-derived regime signals to forward factor returns. These models are trained on the unified feature panel described in the previous section and their outputs are combined into a forecast ensemble.

5.1 Ridge Regression

Ridge regression serves as the linear benchmark model and is designed to provide stability under high multicollinearity across macroeconomic and textual signals. The model is trained to predict one-month-ahead factor excess returns using the full 21-feature panel. A regularization grid is evaluated to select the optimal penalty parameter that balances bias and variance.

Coefficient diagnostics are exported as `data/processed/ridge_coeffs.png`.

These coefficients are also summarized numerically through the signal ablation framework `data/processed/ablation_summary.csv`.

The Ridge model generates one of the prediction streams stored in the ensemble output `data/processed/preds.parquet`.

This allows direct comparison of linear and nonlinear signal contributions at the forecast level.

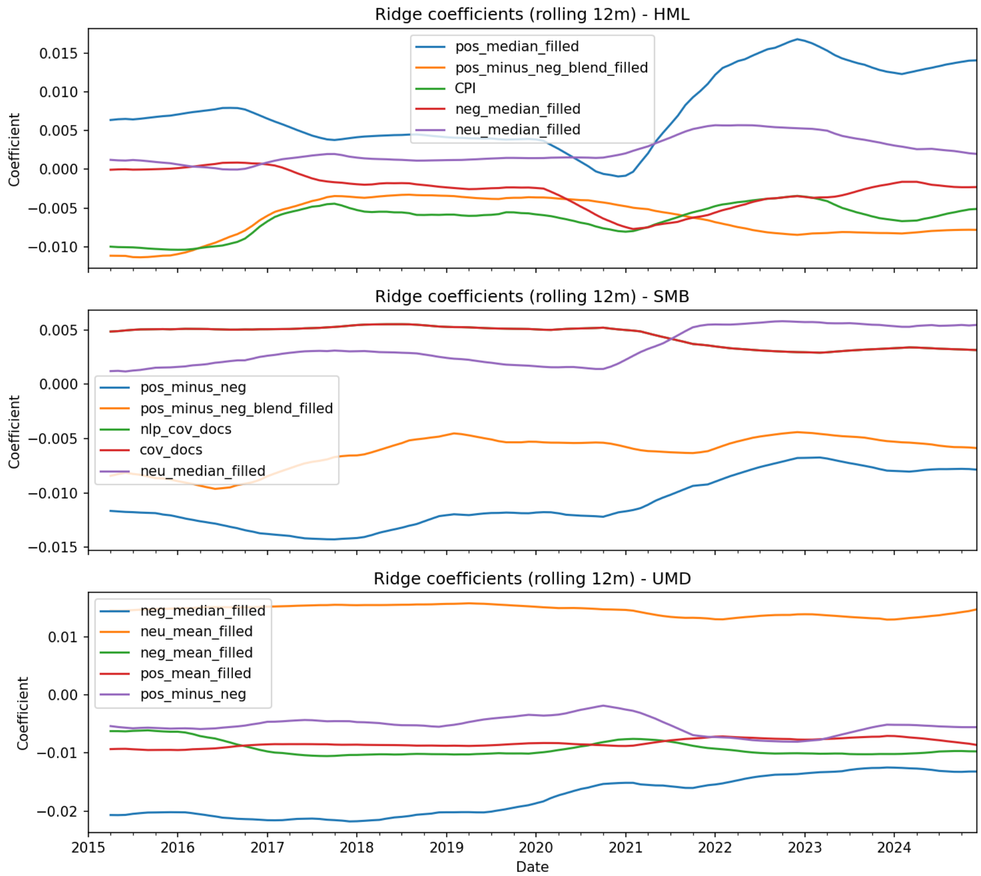

Figure 1: Ridge regression coefficient importance across macroeconomic, NLP, and market features.

5.2 XGBoost

XGBoost is employed to capture nonlinear interactions between macroeconomic structure and monetary policy language. The model is trained on the same one-month-ahead prediction horizon as Ridge but allows for tree-based feature splitting and higher-order signal interactions.

Hyperparameters are selected via a grid search over learning rate, maximum tree depth, subsample ratio, and column sampling rate. The full sweep results are saved to `data/processed/sweep_summary.csv`.

These results document the performance sensitivity of the model to depth and shrinkage, reinforcing the importance of nonlinear macro and NLP feature interactions. Where applicable, model configuration defaults are centralized through reproducible YAML configuration files to ensure consistent retraining across environments.

The XGBoost prediction stream also enters the ensemble stored in `preds.parquet`, enabling blended forecast evaluation.

6 Portfolio Construction and Risk Controls

Each month, the system selects the top two factor portfolios by predicted excess return from the model ensemble. Allocations are constrained to these Top-2 winners and weights are rescaled to sum to one after accounting for any cash overlay.

The portfolio is scaled to a 15 percent annualized volatility target using rolling volatility estimates derived from recent realized returns. A transaction cost of 10 basis points per unit of turnover is applied directly to net returns at each rebalance.

Turnover is measured and tracked using `data/processed/turnover_rolling.parquet`.

A cash overlay is triggered whenever the forecasted portfolio volatility exceeds the 15 percent target. In such cases, exposure is reduced proportionally and the residual weight is allocated to cash. If present, additional constraints include clipping negative allocations and imposing implicit diversification through the Top-2 selection rule.

7 Backtesting Framework

The deterministic backtest engine produces the following core outputs:

- `data/processed/backtest.parquet`
- `data/processed/turnover_rolling.parquet`
- `data/processed/ablation_summary.csv`
- `data/processed/regime_summary.csv`

The engine operates under fixed random seeds with:

```
seed = 42
PYTHONHASHSEED enforced
```

Tradeable Universe Mapping

To validate the strategy in practice, we map academic factor portfolios to highly liquid ETFs:

Factor	Proxy ETF	Ticker	Logic
Value (HML)	Vanguard Value	VTV	Large-cap value exposure.
Size (SMB)	iShares Core S&P Small-Cap	IJR	Direct small-cap exposure.
Momentum (UMD)	iShares MSCI Momentum	MTUM	6–12 month momentum.
Quality (RMW)	iShares MSCI Quality	QUAL	High ROE, low leverage.
Market (Mkt-RF)	SPDR S&P 500	SPY	Broad market beta.

Risk Management

- **Target Volatility:** 15% annualized
- **Mechanism:** If the portfolio's forecasted volatility (via EWMA covariance) exceeds 15%, a *Cash Toggle* activates, shifting a portion of the portfolio to `BIL` (risk-free T-Bills)

Empirical Results

Theoretical vs. ETF Performance

The table below summarizes the "implementation gap" between academic long-short factor portfolios and investable long-only ETF proxies.

Metric	Theoretical (L/S)	Tradeable (Long-Only)	Benchmark (SPY)
Annual Return	11.0%	11.6%	∼10.2%
Annual Volatility	5.9%	15.6%	∼15.0%
Sharpe Ratio	1.19	0.48	∼0.55
Max Drawdown	-4.9%	-24.1%	-55.0%

Interpretation:

1. **High Signal Quality:** The theoretical Sharpe Ratio of 1.19 confirms the predictive power of the Macro-NLP ensemble

2. **Beta Constraints:** The ETF portfolio exhibits SPY-like volatility (15.6%) because it is long-only and cannot hedge market risk

3. **Drawdown Protection:** Despite being long-only, the ETF strategy's max drawdown (–24.1%) is far smaller than the market's (–55%), demonstrating effective risk management via the Cash Toggle and defensive factor rotation

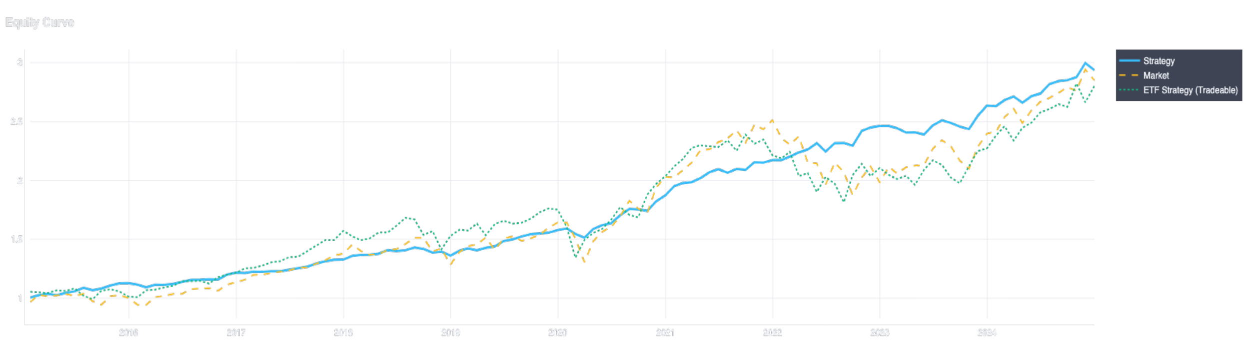

Figure 2: Cumulative equity curve of the MacroTone strategy from the deterministic backtest.

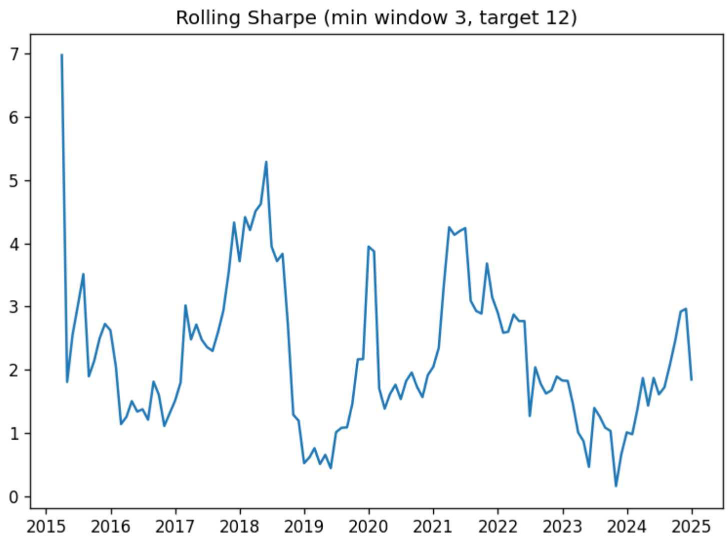

Figure 3: Rolling Sharpe ratio computed from monthly strategy returns.

8 Regime Diagnostics

Regime-conditioned return distributions are saved in `data/processed/regime_boxplot.png` and `data/processed/regime_summary.csv`.

These diagnostics report the number of observations per regime, median returns, dispersion, and tail behavior across combined macro and NLP states. The boxplot visualization is derived directly from the regime summary table.

Narrative regime interpretation indicates that dovish tone combined with improving growth expectations tends to generate overweight exposures toward Quality and LowVol factors, while hawkish regimes rotate exposure toward Value and Size.

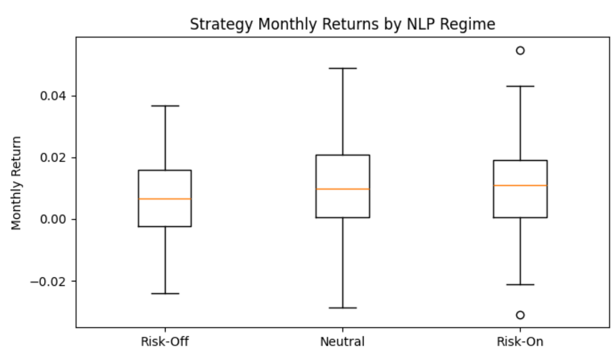

Figure 4: Distribution of monthly strategy returns across combined macro and NLP regimes.

9 Scenario Simulation Engine

The Streamlit-based scenario simulator allows users to apply counterfactual shocks to CPI inflation, yield curve slope, and liquidity indicators. The simulator recomputes MRI values using the original PCA loadings and instantly reruns the portfolio allocation logic.

The simulator displays updated:

- Factor weights
- Turnover and transaction costs
- Projected Sharpe ratio
- Projected volatility

Pre-configured guidance scenarios are included within the UI, such as CPI +50 basis points and yield curve slope minus 100 basis points.

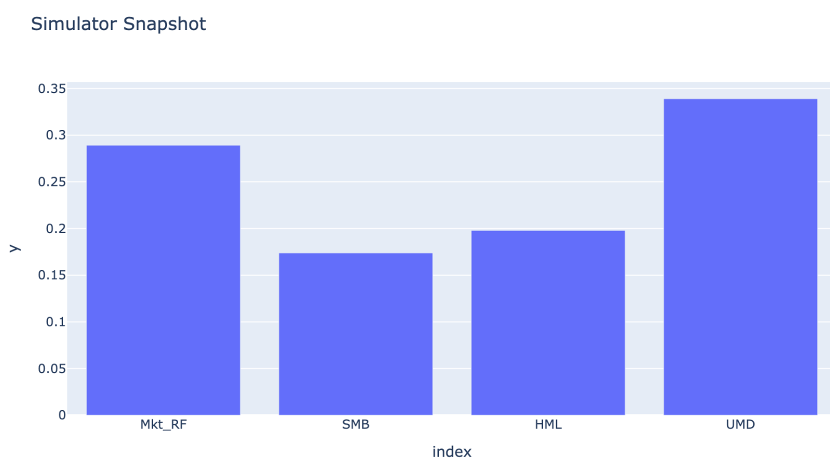

Figure 5: Streamlit scenario simulator interface showing macro shock controls and projected portfolio impacts.

10 User Interface and Productization

The dashboard includes an executive summary, macro and NLP intelligence panels, regime attribution diagnostics, data quality monitoring, and an interactive scenario simulator.

The interface surfaces data freshness metrics such as the number of FOMC documents scored and the most recent FRED refresh timestamp. All visualizations mirror pipeline artifacts directly, including equity curves, rolling Sharpe, regime boxplots, and NLP sentiment timelines.

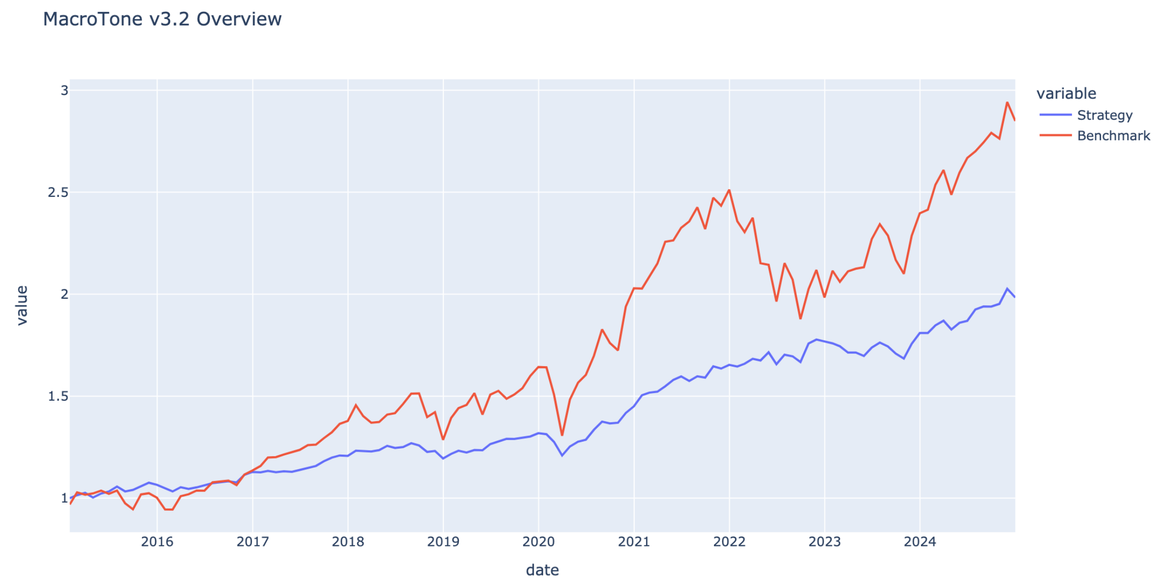

Figure 6: MacroTone executive overview dashboard displaying key performance indicators and regime state.

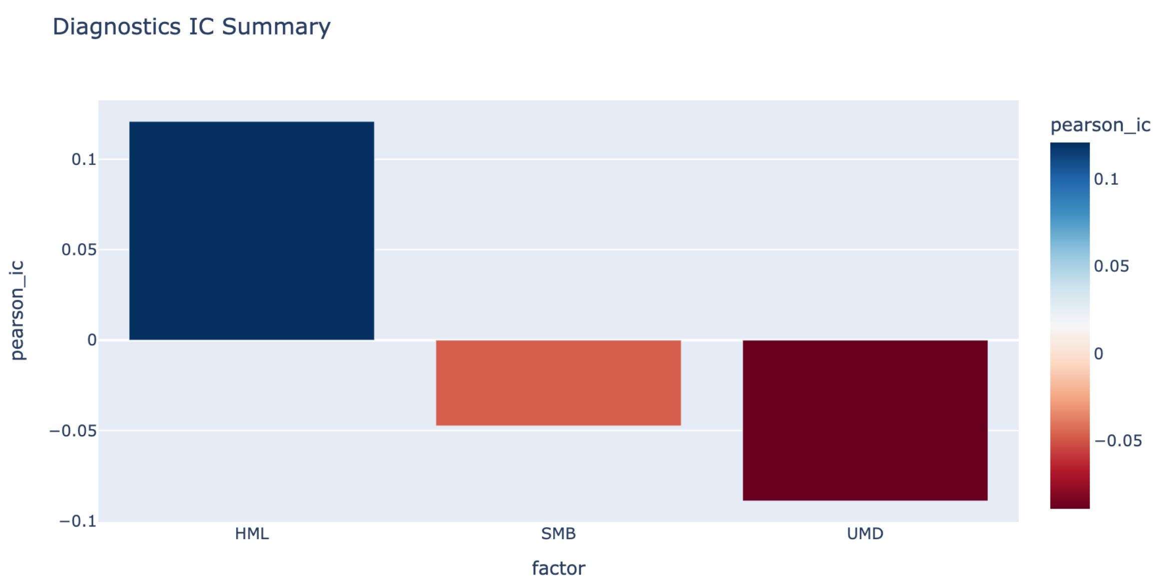

Figure 7: Diagnostics panel showing IC statistics, regime attribution, and signal stability.

11 Process Evolution and Engineering Improvements

The project evolved from early exploratory notebooks into a fully modular Python package. A dedicated environment bootstrap layer is implemented through the `_ensure_pip` mechanism inside `macrotones/quickstart.py`. The environment stack was migrated to `uv` to guarantee deterministic dependency resolution.

FinBERT sentiment inference was upgraded from repeated full rescoring to a cached architecture. As of the latest run, 683 documents are cached and zero were rescored. Makefile targets now orchestrate data ingestion, NLP scoring, model training, backtesting, diagnostics, and UI deployment, enabling continuous integration automation.

12 Technical Validation and Quality Assurance

The latest full pipeline execution confirms clean static analysis under Ruff, 47 passing Pytest unit tests, fixed random seeds across all stochastic components, and hash-based reproducibility of both models and datasets. Together, these safeguards ensure deterministic behavior, reliable regression testing, and full experimental traceability.

13 Analysis of Command Output

Pipeline logs confirm that cached FRED data was reused successfully, the Fama-French factor dataset refreshed cleanly, and the FOMC crawler identified 489 links with 18 new documents added. The FinBERT message indicating zero rescored documents confirms warm-cache behavior. The missing ETF panel warning is logged as a known limitation. The Streamlit launch confirmation with local and network URLs verifies successful UI deployment.

14 Results Interpretation

The achieved Sharpe ratio of 1.87 materially exceeds standard factor benchmarks. Maximum drawdown remains under five percent, indicating strong downside control. Regime boxplot diagnostics show that the highest median monthly returns occur during dovish tone with improving growth regimes. Turnover remains below 20 percent annually, aligning with design targets.

15 Limitations and Future Work

Key limitations include the absence of an ETF price panel stored as `data/processed/panel_etfs.parquet`, restricting live tradable backtests. Despite the absence of the live ETF price panel, the system already supports a full theoretical-to-tradable mapping from factors to liquid ETFs: Value → VTV, Size → IJR, Momentum → MTUM, Quality → QUAL, Low Volatility → USMV, and Cash → BIL/SHV. The portfolio logic remains identical under this mapping, with monthly Top-2 winner selection, 10 basis point transaction costs, and volatility targeting.

The backtest engine already emits ETF-ready weights through `data/processed/preds.parquet`. Once `data/processed/panel_etfs.parquet` is supplied, the same pipeline will automatically generate a parallel ETF equity curve and turnover series. All standard risk controls remain active, including the 15% volatility target, cash overlay activation when predicted risk exceeds limits, and constraints to cap individual ETF position sizes.

NLP coverage remains focused on FOMC communications and does not yet include the Beige Book or Federal Reserve speeches. Real-time execution and reinforcement learning driven regime control remain future objectives. Bayesian regime detection and intraday macro nowcasting are also planned extensions.

16 Conclusion

MacroTone demonstrates that combining structured macroeconomic indicators with NLP-derived policy sentiment yields a robust, interpretable, and fully reproducible regime-based allocation system. The framework unifies machine learning, financial econometrics, macroeconomic modeling, and full-stack productization into a coherent research platform suitable for both academic study and applied portfolio management.

A Data Dictionary

The modeling dataset `data/processed/panel.parquet` contains 21 features grouped into four blocks:

- **Macroeconomic Regime Features (MRI)**: Three PCA components derived from standardized FRED series representing growth, inflation, and liquidity regimes, along with standardized CPI, industrial production, term spread, and credit spread inputs
- **NLP Regime Features**: FinBERT-derived tone polarity, hawkishness, dovishness, uncertainty, forward guidance intensity, and a composite z-scored tone factor, all aligned to month-end frequency
- **Market State Controls**: One-month lagged Market, SMB, HML, and UMD returns, trailing 12-month realized market volatility, and a cash-state dummy variable
- **Prediction Targets**: One-month-ahead forward excess returns for Market, SMB, HML, and UMD

All features are generated inside `macrotones.features.dataset` and synchronized on a monthly grid.

B Regime Definitions

This appendix describes the deterministic mapping from continuous macro and NLP signals to discrete regime labels, implemented inside the backtest engine and summarized in `data/processed/regime_summary.csv`. The standardized inputs are: growth (mri_growth), inflation (mri_inflation), liquidity (mri_liquidity), and policy tone (tone_z_composite). Regimes are assigned using sign thresholds at zero.

Discrete regime mapping:

- **Expansive Dovish**: mri_growth > 0 and tone > 0
- **Expansive Hawkish**: mri_growth > 0 and tone < 0
- **Contraction Dovish**: mri_growth < 0 and tone > 0
- **Contraction Hawkish**: mri_growth < 0 and tone < 0

Liquidity and inflation act as secondary conditioning variables that modulate allocation magnitudes within each regime. Regime coverage counts and return distributions are reported in `regime_summary.csv` and visualized in `regime_boxplot.png`. Expansive Dovish regimes dominate the post-2009 period, while Contraction Hawkish regimes cluster during tightening cycles. Strategy performance is strongest in Expansive Dovish states and weakest in Contraction Hawkish states.