Structure EDA (Exploratory Data Analysis)

1. Data understanding

Variables Available

We are working with the following macroeconomic time-series variables collected from official sources:

- US Inflation (65 observations)
- US Interest Rate (3,000+ observations)
- Europe Inflation (348 observations)
- Europe Interest Rate (46 observations)
- EUR/USD Exchange Rate (target variable)
-

Time Period & Frequency

The dataset clearly combines mixed frequencies:

- US interest rates (3,000+ observations) - daily frequency starting xx/xx/xxxx
- US inflation (65 observations) - monthly data starting xx/xx/xxxx
- European inflation (348 observations) - monthly data starting xx/xx/xxxx
- European interest rate (46 observations) - x data starting xx/xx/xxxx

This mismatch in frequency is a critical structural issue. Before modeling, all variables must be aligned to a common frequency, which will be the daily.

Missing Values

Expected issues:

- Weekends/holidays in EUR/USD
- Macro variables not released daily

Solution:

- Use only trading days
- Forward-fill macro variables
- Drop initial rows before all variables are available

Data Quality Issues

- Structural breaks (2008, 2012, 2020, 2022)
- Zero lower bound in Europe
- Post-COVID inflation shock
- Monetary regime shifts

These create nonlinear regime-dependent behavior.

This will affect ML stability.

2. Univariate Analysis

EUR/USD (Daily)

- Non-stationary in levels
- Random walk behavior
- No clear long-term mean
- Periods of strong directional trends
- Volatility clustering present
- Crisis-driven spikes

Conclusion:

We must use log-returns, not levels.

$rt=\log(EUR/USDt)-\log(EUR/USDt-1)$

US Interest Rate (Daily)

- Step-like pattern
- Long flat periods
- Sudden jumps on FOMC decisions
- Strong regime shifts

Low short-term volatility, high regime shifts.

European Interest Rate

- Long period near zero / negative
- Step-like behavior
- Structural break in 2022 tightening cycle

Lower variability than US rates historically.

Inflation (US & EU)

- Step function once converted to daily
- Long stable periods
- Extreme spike in 2021–2022
- No daily variation (constant within month)

Important:

Inflation daily representation adds information only at release boundaries.

3. Time-Series Properties

3.1 EUR/USD

The EUR/USD series in levels is non-stationary. It does not fluctuate around a fixed mean and does not have constant variance over time. It exhibits long appreciation and depreciation cycles.

For modeling purposes, the appropriate target is daily returns, not the exchange rate level.

Daily returns:

- Fluctuate around a relatively stable mean (close to zero).
- Show periods of calm and periods of turbulence.
- Do not exhibit a deterministic trend.

This transformation is necessary before applying machine learning models.

3.2 Interest Rates

Both US and European interest rates show:

- Long flat periods.
- Sudden discrete adjustments (policy decisions).
- Structural regime shifts (e.g., zero-lower-bound period, tightening cycles).

In levels, they should be treated as non-stationary.

For modeling, either:

- Use daily changes, or
- Use interest rate differentials.
- Using levels directly risks introducing spurious relationships.

3.3 Inflation

After converting to daily frequency (forward-filled within each month), inflation becomes a step function:

- Constant during most days.
- Changes only when a new monthly value is introduced.

In levels, inflation should also be treated as non-stationary.

Because inflation changes infrequently, its predictive power at daily frequency will be limited. The relevant information is more likely embedded in:

- Changes in inflation,
- Or relative inflation between regions.

3.4 Volatility Behavior

Daily EUR/USD returns do not have constant volatility over time.

There are identifiable periods of:

- Low volatility (stable macro conditions),
- High volatility (financial crises, monetary shocks, geopolitical stress).

This means volatility is clustered: large moves tend to be followed by large moves, and calm periods tend to persist.

From a modeling perspective, it is appropriate to include:

- Lagged returns,
- Rolling volatility measures,
- Or other features capturing recent market turbulence.

Ignoring time-varying volatility would reduce model robustness.

4. Multivariate Relationships

4.1 Interest Rate Differential

The most economically grounded explanatory variable is the interest rate differential between the US and the Euro area.

When US rates are higher relative to European rates, capital allocation incentives favor USD-denominated assets. This typically leads to USD appreciation (EUR/USD decline).

Therefore, the interest rate spread is expected to be more informative than using both rates separately.

4.2 Inflation Differential

Inflation affects exchange rates indirectly through monetary policy expectations.

Higher inflation in one region may:

- Trigger tighter monetary policy,
- Influence capital flows,
- Affect currency valuation expectations.

However, at daily frequency, inflation levels change infrequently. As a result:

- The direct daily impact is limited.
- Lagged or differential measures are more relevant than raw levels.

4.3 Interaction Structure

The relationship structure is hierarchical:

Inflation → Monetary policy (interest rates) → Capital flows → Exchange rate

This creates:

- Correlation between predictors,
- Potential redundancy,
- Risk of multicollinearity in linear models.

Feature engineering must account for this, for example by:

- Using spreads instead of individual levels,
- Testing models with reduced predictor sets,
- Comparing linear vs tree-based approaches.

4.4 Regime Dependence

Macroeconomic variables do not influence exchange rates uniformly across time.

During crisis periods:

- Safe-haven flows may dominate macro fundamentals.
- Risk sentiment may overpower rate differentials.

This implies that relationships are not constant over time. Model performance should therefore be evaluated across different subperiods.