**Time Series Structural Break Challenge Methodology**

Sections:

1. **Introduction to Challenge**
2. **Summary of Wavelet Based Detection, Khudhayr et al. 2022 Paper**
3. **Implementation of a feature set for supervised learning.**

**Section 1: Time Series Data Science Structural Break Challenge**

🔗 Competition link: [ADIA Lab Structural Break Challenge](https://hub.crunchdao.com/competitions/structural-break)

**Objective:**  
Detect whether a structural break occurred at a given boundary point in a univariate time series.

**Definition of Structural Break (Summary):**

* A structural break is a **change in the underlying data-generating process (DGP)** at a specific time.
* Can manifest as changes in parameters (e.g., mean, variance, volatility), functional form, or regime transitions.
* Breaks may be abrupt or gradual.
* No break means the process parameters remain stable across the boundary.

**Setup:**

* Each time series is split into two candidate periods:
  + *Period 0*: before the boundary
  + *Period 1*: after the boundary
* Not every boundary represents a true break — some are false splits.
* Task: assign a score ∈ [0,1] indicating likelihood of a break.
* Evaluation: ROC AUC.

**Training Data:**

* X\_train: MultiIndex DataFrame (id, time) with columns:
  + value (float): observed time series value
  + period (0/1): before vs. after the boundary
* y\_train: Series indexed by id, Boolean values:
  + True = genuine structural break
  + False = no break
* Distribution: ~30% True, ~70% False.

**Test Data:**

* Same structure as X\_train, but labels are hidden.
* Submission: Python code with train() and infer() functions, returning one prediction per id.

**Dataset Description (Summary):**

* Tens of thousands of synthetic univariate time series, each ~1,000–5,000 values.
* Each series has a designated boundary point and label (True/False) indicating break or no break.
* Series simulate scenarios from **finance, climate, industry, biomedical signals**, with varying detection difficulty.
* Goal: algorithms must **generalize across diverse scenarios** and detect breaks in unseen data.

**Methodology Suggestions (Summary):**

* Approaches include: change-point detection, distribution comparison tests, anomaly detection, supervised ML, or deep learning.
* Useful strategies: statistical tests before/after the boundary, feature extraction from both parts, time series modeling, automated deep feature learning.
* Careful preprocessing is key for robustness.
* Aim: build reliable, generalizable detection models.

**📊 Dataset Schema Diagram**

**X\_train (DataFrame with MultiIndex)**

Index: [id, time]

Columns:

- value (float)

- period (int: {0,1})

Example:

| **id** | **time** | **value** | **period** |
| --- | --- | --- | --- |
| 0 | 0 | 0.001858 | 0 |
| 0 | 1 | -0.00166 | 0 |
| 0 | 1890 | -0.00590 | 1 |
| 0 | 1891 | 0.007295 | 1 |

**y\_train (Series)**

Index: [id]

Values: bool (True/False)

Example:

| **id** | **y** |
| --- | --- |
| 0 | True |
| 1 | False |
| 2 | False |

**Methodology for Structural Break Detection**

**Core approach:**  
Use supervised learning (e.g., gradient boosting trees, random forests, or neural nets) trained on features, tests, models extracted from periods 0 and 1.

# Section 2. Summary of Boundary Break Test via MODWT-Calibrated Residual Analysis

## Boundary Break Test Via Modwt-Calibrated Residual Analysis

(Adapted from: Rashedi, Ismail, Serroukh, Al Wadi, “Wavelet Based Detection of Outliers in Volatility Time Series Models,” CMC, 2022)

## Goal & Data Setup

* Goal: Test whether the data-generating process (DGP) in Period 1 differs materially from Period 0 at a known boundary.
* Data: Univariate time series with a known split into Period 0 (before boundary) and Period 1 (after boundary).
* Output: A decision (break vs no break), and optionally a p-value or confidence measure.

## Assumptions & Notation

* Let B be the boundary index.
* If prices are strictly positive, work with returns: R\_t = 100 \* (ln X\_t − ln X\_{t−1}); else use first differences.
* Under H0 (no break), one stationary univariate model explains the whole series across the boundary.

## Step-By-Step Procedure

0) PREPROCESSING (promote stationarity)

* If X\_t > 0, convert to returns R\_t. Otherwise difference the series.
* Optionally mean-center R\_t.

1) FIT A UNIVARIATE VOLATILITY MODEL UNDER H0 (NO BREAK)

* Fit on the entire concatenated sequence (Period 0 + Period 1).
* Recommended models: GARCH(1,1), EGARCH(1,1), or GJR-GARCH(1,1).

Example: R\_t = μ\_t + a\_t, a\_t = σ\_t ε\_t, σ\_t^2 = ω + α a\_{t−1}^2 + β σ\_{t−1}^2.

* Choose error law (Normal or Student-t). Student-t is often more robust for fat tails.
* Select model by AIC/BIC if desired.

## 2) Compute Standardized Residuals

* Obtain fitted residuals â\_t and conditional volatilities σ̂\_t.
* Standardize: ε̂\_t = â\_t / σ̂\_t. Under a well-specified H0, ε̂\_t should behave i.i.d.

## 3) Apply Modwt To Residuals

* Use MODWT with a moderate-length wavelet filter (Least Asymmetric LA(8) recommended; Haar is allowed but may cause artifacts).
* Compute wavelet coefficients W\_{j,t} for scales j = 1,…,J (typically J ≤ 3).

4) THRESHOLD CALIBRATION BY MONTE CARLO (PAPER’S CALIBRATION REGIME)

* Calibration target: max statistic per scale, W\_{j,max} = max\_t |W\_{j,t}|.
* Simulate ε\_t i.i.d. under the chosen error law and series length n.
* For each simulation, apply MODWT and record W\_{j,max}.
* Obtain quantiles q\_{j,α}(n) for desired α (e.g., 0.05 or 0.02). These are the critical values used by the paper.
* For localization near the boundary (see Step 5), also simulate window-length quantiles q\_{j,α}(m) for a local window of size m.

## 5) Localize The Test At The Known Boundary

* Define a symmetric window 𝒲 around B: 𝒲 = { t : |t − B| ≤ w }, with window size m.
* Local max statistic per scale: S\_j(B) = max\_{t∈𝒲} |W\_{j,t}|.
* Decision per scale j: Reject H0 at scale j if S\_j(B) > q\_{j,α}(m).
* Combine across scales j = 1,…,J using a multiplicity correction (e.g., Bonferroni). If any adjusted test rejects, declare a break.

Robust alternatives (optional):

* Count test: C\_j(B) = number of t ∈ 𝒲 such that |W\_{j,t}| exceeds a pointwise threshold with nominal exceedance rate α.

Under H0, C\_j(B) ≈ Binomial(m, α). Large C\_j(B) supports a break.

* Energy contrast: Compare E\_{j,0} = ∑\_{t∈Period 0} W\_{j,t}^2 and E\_{j,1} = ∑\_{t∈Period 1} W\_{j,t}^2 (normalized for segment length).

Large deviations imply a scale-specific change (variance/roughness shift). Calibrate via Monte Carlo.

6) OPTIONAL ITERATIVE CLEANING (AS IN THE PAPER)

* If large exceedances are found (global or local), zero-out the largest offending coefficient(s), inverse-MODWT to reconstruct “cleaned” residuals, and refit the volatility model.
* Iterate until no exceedances remain (stabilized residual kurtosis near 3). This step improves model adequacy but is not required for the boundary test itself.

## 7) Final Decision

* Break if any scale-specific boundary-localized test rejects after multiplicity control.
* Otherwise, fail to reject (no break detected).

## Implementation Notes

* Filter choice: LA(8) preferred to avoid blocky artifacts; Haar is acceptable but may be less smooth.
* Number of scales: small J (≤3) is sufficient and consistent with the paper; higher levels produce smaller thresholds and capture coarser changes.
* Error law for calibration: match the fitted residual law (Normal or t). If in doubt, use t with ν≈7–10.
* Window size m: choose so that m ≪ n but large enough to capture boundary-localized energy (e.g., 0.5–2% of n on each side).
* Multiple testing: across scales (and/or across statistics), adjust α (Bonferroni/Simes).
* Power vs specificity: the paper notes a “lower common hard threshold” trick (use a higher-level, smaller quantile across lower levels) to gain sensitivity when tails are heavier than assumed.

PSEUDOCODE (HIGH-LEVEL)

Input: series X\_t with boundary B, significance α, window half-width w, scales J, wavelet filter F.

1. Transform X\_t → R\_t (returns or differences).

2. Fit GARCH-family model on full R\_t; compute standardized residuals ε̂\_t.

3. MODWT\_F(ε̂\_t) → { W\_{j,t} for j=1..J }.

4. Calibrate q\_{j,α}(m) via Monte Carlo (error law = fitted; length = window size m).

5. For each scale j: S\_j(B) = max\_{t∈𝒲} |W\_{j,t}|; reject\_j = [ S\_j(B) > q\_{j,α}(m) ].

6. Apply multiplicity correction across j. If any reject\_j survives correction → declare BREAK; else NO BREAK.

(Optional) Clean-and-refit loop to improve residual diagnostics, following the paper’s procedure.

**Section 3: Implementation towards a rich feature set**

**Synopsis of Khudhayr et al. (2022)**

The paper introduces a procedure to **detect and remove outliers** in financial time series, particularly those modeled with GARCH-family models. The core idea is that if a volatility model (like GARCH) correctly captures the data's structure, its standardized residuals should be random noise (i.i.d.). However, outliers in the original data will manifest as large, sharp spikes in these residuals4.

The authors use the **Maximal-Overlap Discrete Wavelet Transform (MODWT)** to decompose the residual series5. MODWT is chosen for its flexibility over the standard DWT, as it works with any sample size and avoids certain artifacts. Outliers are identified by comparing the maximum absolute value of the wavelet coefficients at different scales against pre-calculated statistical thresholds. These thresholds are determined via Monte Carlo simulations of pure noise. The paper's main application is an iterative "cleaning" process: detect an outlier, remove its effect, and refit the model until the residuals are well-behaved.

**Outline of the Main Methodology**

**Intuitive Outline**

Imagine you're trying to tune an old radio to a clear station.

1. **Tune the Radio (Model the Data):** First, you fit a GARCH model to your financial data. This is like tuning the radio to the main signal (the predictable volatility patterns).
2. **Listen to the Static (Get Residuals):** The leftover noise or "static" that the model can't explain is the series of residuals. If the model is good, this should just be random, quiet static.
3. **Find the "Pops" (Wavelet Analysis):** An outlier in the original data creates a loud, sudden "pop" in the static. The wavelet transform (MODWT) is a special audio analyzer that is extremely good at detecting these sharp, short-lived pops at different frequencies (or "scales").
4. **Set a Volume Threshold (Statistical Test):** To decide if a pop is significant or just random noise, you need a threshold. The authors figure this out by listening to thousands of recordings of pure static to see how loud it gets naturally. Any pop in your data that's louder than this "natural maximum" is flagged as a real outlier.
5. **Clean the Signal (Iterative Removal):** Once a pop is found, the method digitally removes it from the static, tunes the radio again, and listens for more pops, repeating until the static is clean.

**⚙️ Technical Outline**

The paper's procedure for outlier detection is as follows:

1. **Model Fitting:** Fit a GARCH-family model (e.g., GARCH(1,1), EGARCH) to the financial returns series, assuming a specific error distribution (e.g., Normal or Student's t)10101010.
2. **Residual Extraction:** Compute the standardized residuals, ϵ^t​=a^t​/σ^t​. Under the null hypothesis of a correctly specified model with no outliers, these residuals should be i.i.d.
3. **Wavelet Decomposition:** Apply the MODWT with a suitable filter (LA(8) is recommended) to the standardized residuals

ϵ^t​ to obtain wavelet coefficients Wj,t​ for scales j=1,...,J.

1. **Thresholding:** For each scale j, compare the observed maximum absolute wavelet coefficient, Wj,max(obs)​=maxt​∣Wj,t​∣, to a critical value qj,α​.
2. **Calibration:** These critical values qj,α​ are (1-α) quantiles pre-computed via Monte Carlo simulation. The simulation involves generating many series of i.i.d. noise of length

n from the assumed error distribution, applying MODWT, and recording the distribution of Wj,max​.

1. **Iterative Cleaning:** If Wj,max(obs)​>qj,α​ at any scale, an outlier is detected. The procedure then sets the largest offending wavelet coefficient to zero, reconstructs the residuals via an inverse MODWT, and refits the GARCH model on the "cleaned" data. This process is repeated until no more outliers are detected.

**Revised Implementation for the Structural Break Challenge**

Proposed implementation towards a robust, feature-based machine learning solution tailored for the challenge.

**Overall Strategy Shift**

Instead of a binary hypothesis test, we'll **extract multiple statistical features** from the wavelet decomposition of the residuals. These features are designed to quantify different aspects of the change around the boundary. A supervised classifier will then learn to map these features to the probability of a break.

**Proposed Feature Engineering**

For each time series id, perform the following steps:

1. **Preprocessing & Modeling:** As you outlined, convert the full series (Period 0 + Period 1) to returns or differences and fit a single GARCH-family model under the null hypothesis (no break)16. Obtain the full series of standardized residuals

ϵ^t​.

1. **MODWT:** Apply MODWT to the entire ϵ^t​ series to get wavelet coefficients Wj,t​ for scales j=1,2,3.
2. **Feature Extraction:** Now, compute the following features for each scale j.
   * **Feature 1: Localized Maximum Coefficient**
     + This is your statistic

Sj​(B). Define a window

W of size m around the boundary B. Calculate the maximum absolute coefficient within this window. This detects sharp, spike-like changes at the boundary.

* + - Fj,max\_local​=maxt∈W​∣Wj,t​∣
  + **Feature 2: Energy Contrast**
    - This quantifies changes in variance or volatility at different scales. It compares the energy (sum of squared coefficients) in Period 1 to Period 0, normalized by length.
    - Ej,0​=len0​1​∑t∈Period 0​Wj,t2​
    - Ej,1​=len1​1​∑t∈Period 1​Wj,t2​
    - Fj,energy\_ratio​=Ej,0​Ej,1​​ (or log(Ej,1​/Ej,0​) for stability)
  + **Feature 3: Localized Exceedance Count**
    - This captures a "cluster" of unusual activity near the boundary. First, find a global threshold qj,α​(n) based on the full series length n. Then, count how many coefficients inside the window

W exceed this threshold.

* + - Fj,count\_local​=∑t∈W​I(∣Wj,t​∣>qj,α​(n))
    - We can make this feature more sensitive by using the "lower common hard threshold" trick from the paper (e.g., using

q3,α​(n) for all scales j=1,2,3).

* + **Feature 4: Distributional Change of Coefficients**
    - Compare the statistical properties of the wavelet coefficients between the two periods.
    - Fj,mean\_diff​=mean(Wj,t​ in Period 1)−mean(Wj,t​ in Period 0)
    - Fj,std\_ratio​=std(Wj,t​ in Period 1)/std(Wj,t​ in Period 0)

By calculating these 4 types of features for 3 wavelet scales, you generate a 12-dimensional feature vector for each time series id.

**Revised Step-by-Step Algorithm**

**train(X\_train, y\_train) function:**

1. Initialize an empty list for features and corresponding labels.
2. For each id in X\_train:

a. Extract the full time series value.

b. Preprocess, fit GARCH, get standardized residuals.

c. Apply MODWT to the residuals.

d. Calculate the 12-dimensional feature vector as described above.

e. Append the feature vector to the feature list and the corresponding y\_train[id] to the label list.

1. Train a gradient boosting classifier (e.g., LightGBM, XGBoost) on the complete feature matrix and labels.
2. Return the trained model.

**infer(model, X\_test) function:**

1. Initialize an empty list for features.
2. For each id in X\_test:

a. Repeat steps 2a-2d from the training function to generate the same 12-dimensional feature vector for the test series.

b. Append the feature vector to the feature list.

1. Use the trained model.predict\_proba() method on the test feature matrix to get the probability scores for the positive class (break = True).
2. Return these scores as the final predictions.

## Reference

Rashedi, K. A., Ismail, M. T., Serroukh, A., & Al Wadi, S. (2022). Wavelet Based Detection of Outliers in Volatility Time Series Models. Computers, Materials & Continua, 72(2), 3835–3847.