1) Identify relevant regimes in the data

K = 2 regimes (high volatility vs. low volatility)

Alternative approaches: Multivariate Clustering using additional variables

Algorithm	Source package (version)	Category
KMeans	scikit-learn (1.7.2)	Clustering
AgglomerativeClustering	scikit-learn (1.7.2)	Clustering
DBSCAN	scikit-learn (1.7.2)	Clustering
SpectralClustering	scikit-learn (1.7.2)	Clustering
MeanShift	scikit-learn (1.7.2)	Clustering
GaussianMixture	scikit-learn (1.7.2)	Clustering
Birch	scikit-learn (1.7.2)	Clustering
AffinityPropagation	scikit-learn (1.7.2)	Clustering
OPTICS	scikit-learn (1.7.2)	Clustering
MiniBatchKMeans	scikit-learn (1.7.2)	Clustering

Hyperparameter Optimisation

Technique	Source package (version)	Category
GridSearchCV	scikit-learn (1.7.2)	Parameter Optimization

TimeSeriesSplit

CombinatorialPurgedCV

Classical approaches: Univariate econometric modelling (benchmarks)

Markov Regression/Switching

Algorithm	Source package (version)	Category
MarkovRegression	statsmodels (0.14.5)	Regime Switching Regression

$$Pr(I_t = 1 | I_{t-1} = 1) = P$$

$$Pr(I_t = 2 | I_{t-1} = 1) = 1 - P$$

$$Pr(I_t = 2 | I_{t-1} = 2) = Q$$

$$Pr(I_t = 1 | I_{t-1} = 2) = 1 - Q$$

Hamilton suggests that Bayesian inference is an unbiased and efficient way to infer regime probabilities

$$Pr(I_{t-1} = 1 | \Delta s_{t-1}) = \frac{f(\Delta s_{t-1} | I_{t-1} = 1) p_{1t-1}}{f(\Delta s_{t-1} | I_{t-1} = 1) p_{1t-1} + f(\Delta s_{t-1} | I_{t-1} = 2)(1 - p_{1t-1})}$$

and

$$Pr(I_{t-1} = 2 | \Delta s_{t-1}) = \frac{f(\Delta s_{t-1} | I_{t-1} = 2)(1 - p_{1t-1})}{f(\Delta s_{t-1} | I_{t-1} = 1) p_{1t-1} + f(\Delta s_{t-1} | I_{t-1} = 2)(1 - p_{1t-1})}$$

where p_{1t-1} and $p_{2t-1} = 1 - p_{1t-1}$ are called prior probabilities

Using the posteriors we can calculate an expectation of the next periods regime probability as

$$p_{1t} = P\ Pr(I_{t-1} = 1 | \Delta s_{t-1}) + (1 - Q)\ Pr(I_{t-1} = 2 | \Delta s_{t-1})$$

or

$$p_{1t} = P \left[\frac{f_{1t-1} p_{1t-1}}{f_{1t-1} p_{1t-1} + f_{2t-1}(1 - p_{1t-1})} \right] + (1 - Q) \left[\frac{f_{2t-1}(1 - p_{1t-1})}{f_{1t-1} p_{1t-1} + f_{2t-1}(1 - p_{1t-1})} \right]$$

and

$$p_{2t} = 1 - p_{1t}$$

Estimation of the model by maximizing the log-likelihood:

$$L = \sum_{t=1}^{T} log \left[p_{1t} \frac{1}{\sqrt{2\pi h_{1t}}} exp(\Theta_1) + (1 - p_{1t}) \frac{1}{\sqrt{2\pi h_{2t}}} exp(\Theta_2) \right]$$

and

$$\Theta_1 = \frac{-(\Delta s_t - \mu_{1t})^2}{2h_{1t}}, \Theta_2 = \frac{-(\Delta s_t - \mu_{2t})^2}{2h_{2t}}$$

$$Regime1 : \mu_{1t} = \alpha_1 + \beta_1 (i_{t-1} - i^*_{t-1});\ h_{1t} = \sigma_1^2$$

$$Regime2 : \mu_{2t} = \alpha_2 + \beta_2 (i_{t-1} - i^*_{t-1});\ h_{2t} = \sigma_2^2$$