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

Classical approaches: Univariate econometric modelling (benchmarks)

Markov Regression/Switching

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

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

$$Pr(I_{t-1} = 1 | \Delta s_{t-}$$

- 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

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

- Second step: Specification of the regime dependent distribution

$$f(\Delta s_{t-1} | I_{t-1} = 1, 2, \Phi)$$

- Simplest case:

$$\mu_{1t} = c_1; \ \mu_{2t} = c_2$$
$$h_{1t} = \sigma_1^2; \ h_{2t} = \sigma_2^2$$

- Alternatively: e.g.

$$\mu_{1t} = c_1 + \beta_1 \Delta s_{t-1}; \ \ \mu_{2t} = c_2 + \beta_2 \Delta s_{t-1}$$
$$h_{1t} = b_{01} + b_{11} u_{t-1}^2 + b_{21} h_{1t-1}$$
$$h_{2t} = b_{02} + b_{12} u_{t-1}^2 + b_{22} h_{1t-1}$$

- We may suggest that UIP holds in 'normal' times, but is violated in 'non-normal' times
- The sequences of switching between the two states may help us learning what drives the UIP puzzle
- Therefore we need two separate mean equations to be estimated

$$Regime1: \mu_{1t} = \alpha_1 + \beta_1 (i_{t-1} - i_{t-1}^*); \ h_{1t} = \sigma_1^2$$
$$Regime2: \mu_{2t} = \alpha_2 + \beta_2 (i_{t-1} - i_{t-1}^*); \ h_{2t} = \sigma_2^2$$

- Coefficients estimated by maximizing the Log Likelihood as specified above