- There are two broad approaches to model regime shifts in a time series:
  - Assume that the underlying law governing a regime shift can be captured through an threshold variable (TAR – exogenous, SETAR – endogenous)
  - Model the regimes via a variable that is latent (not observable), and that the evolution of the regime variable is governed by a Markov chain (MS-AR)

- Hamilton (1989) proposed a bi-regime, MS-AR model for US GNP

- The bi-regime Markov-switching model with AR order p can be denoted as MS(2)-AR(p), and is defined as:

$$y_t = \mu(s_t) + \sum_{i=1}^{p} \alpha_i (y_{t-i} - \mu(s_{t-i})) + \varepsilon_t, \qquad where \;\; \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

where y_t is an observation in the time series with time index t, α_i are the AR coefficients for $i = 1, 2, \ldots, p$, and s_t is a regime variable of the system such that, $s_t = 1$ corresponds to regime 1 ($\mu(s_t) > 0$, growth) and $s_t = 2$ corresponds to regime 2 ($\mu(s_t) < 0$, recession). The variance and AR model parameters are the same in each regime

- The mean $\mu(s_t)$ switches between the two regimes, such that $\mu(s_t)$ is positive if s_t = 1, and negative if s_t = 2. The transition between the two regimes depends on the variable s_t, governed by a first order Markov chain with transition probabilities P given as:

$$P_{rj} = P(s_{t+1} = j \mid s_t = r)$$

for all $r, j \in \{1, 2\}$ and $P_{r1} + P_{r2} = 1$

- The estimation of this parameter vector has been discussed in detail by Hamilton (1990) and Krolzig (1997), based on maximizing the likelihood
- MS-AR have been commonly used for capturing the periods of growth and recession in a business cycle
- However, the out-of-sample forecast performance has been questionable

Applications of MS-AR

MS-AR applications:

- Hamilton (1989) –modelling US GNP
- Dueker and Neely (2007) – forecasting exchange rates
- Deschamps (2008) – forecasting US unemployment
- Crawford and Fratantoni (2003) – forecasting house prices
- Guidolin (2011) – review of the use of Markov switching models in finance

Analogue prediction methods

- Given a discrete time series, y_t, for $t = 1, 2, \dots, N$
- Construct a m-dimensional delay vector $\boldsymbol{y}_t = (y_{t-(m-1)\tau}, \dots, y_{t-\tau}, y_t)$
- The dimension m and reconstruction delay τ can be estimated using cross-validation
- Select model structure: $y_{t+1} = F(\boldsymbol{y}_t, \boldsymbol{a}) + \varepsilon_{t+1}$
- Employ in-sample data for training the model (estimating model parameters)
- Use distinct out-of-sample data for testing the model
- Fit model:

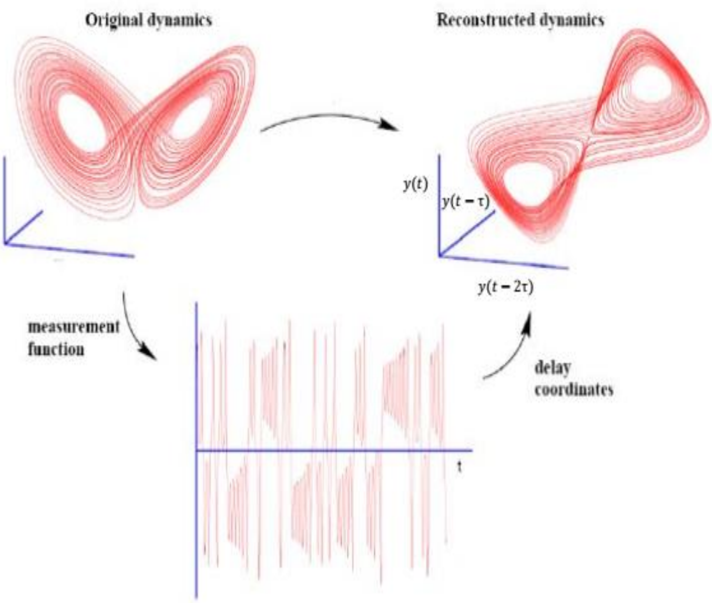
$$\hat{y}_{t+1} = F(\boldsymbol{y}_t, \boldsymbol{a})$$

by minimizing the sum of squared errors:

$$C_{LS} = \sum_i (y_{t+1} - \hat{y}_{t+1})^2$$

- This provides the least square estimates of parameters $\boldsymbol{a}$
- Model approach relies on concepts of state space reconstruction

State space reconstruction



Selecting the reconstruction delay

- For real data, a good choice of delay parameter (τ) is important
- If delay is too small, successive components will be strongly correlated
- If delay is too large, successive components will be unrelated
- Autocorrelation function (ACF) can be employed by selecting the delay as the first zero of the ACF
- Note that the ACF only measures linear relationships
- Nonlinear analogue is obtained using the mutual information

Local analogue prediction method

- To make a prediction from $\boldsymbol{y}_t$:
- First define a local neighbourhood $B(\boldsymbol{y}_t)$ around $\boldsymbol{y}_t$
  - – select k nearest neighbours from the learning data, or
  - – select all nearest neighbours within distance r of $\boldsymbol{y}_t$, or
  - – select a fraction of nearest neighbours from the learning data
- Local analogue: use the future of the nearest neighbour as the prediction,

$$\hat{y}_{t+1} = y_{j+1}, \text{ where } j \text{ minimizes } ||\boldsymbol{y}_t - \boldsymbol{y}_j||$$

- Local average: use the average of the future of all neighbours found in the neighbourhood

$$\hat{y}_{t+1} = \frac{1}{|B(\boldsymbol{y}_t)|} \sum_{j \in B(\boldsymbol{y}_t)} y_{j+1}$$

where $|B(\boldsymbol{y}_t)|$ denotes the number of state vectors in $B(\boldsymbol{y}_t)$

- Any advantage of defining neighbourhood size in terms of k as opposed to r?