An algorithm for inference of a class of Markovian Regime-Switching Models

Angus Lewis with Nigel Bean & Giang Nguyen

AustMS, December, 2018

MRS models

- Theory developed by Hamilton:
  - *Rational expectations econometric analysis of changes in regimes: An investigation of the term structure of interest rates*, 1988
- Motivating idea – Different economic regimes determine behaviour of financial markets
- Widely applied since (10,000+ citations):
  - Weather modelling
  - Speech recognition
- Generalises HMMs
- Extended to hidden semi-Markov models/semi-Markovian regime-switching models
- We consider a new type of MRS specification

MRS models

Typically specified as

$$X_t = \alpha^{(i)} + \phi_1^{(i)} X_{t-1} + \cdots + \phi_p^{(i)} X_{t-p} + \sigma^{(i)} \varepsilon_t \text{ when } R_t = i$$

where

$$\{R_t\}_{t \in \mathbb{N}}$$

is a Markov chain on state space $\mathcal{S} = \{1, 2, ..., M\}$, and

$$\{\varepsilon_t\} \sim \text{ i.i.d. } N(0, 1).$$

Problems & Solutions

Two problems

- Likelihood evaluation
- Likelihood maximisation

Three algorithms

- Forward – likelihood evaluation
- Backward – enables the EM algorithm & quantities of interest
- Expectation Maximisation (EM) – likelihood maximisation

Evaluating likelihoods (naively)

- The regime process, $\{R_t\}$, is hidden
- To evaluate the likelihood.
  Let $\boldsymbol{x}_{0:t} = (x_0, ..., x_t)$,

$$
\begin{aligned}
f_\theta(\boldsymbol{x}_{0:T}) &= \sum_{\boldsymbol{R} \in \mathcal{R}} f_\theta(\boldsymbol{x}_{0:T}, R_0, ..., R_T) \\
&= \sum_{\boldsymbol{R} \in \mathcal{R}} f_\theta(\boldsymbol{x}_{0:T} \mid R_0, ..., R_T) f_\theta(R_0, ..., R_T),
\end{aligned}
$$

  where
  - $\boldsymbol{R} = (R_0, ..., R_t)$ is a hidden regime sequence
  - T is the number of observations
  - Naive complexity is $\mathcal{O}(M^T)$
    - M is the number of regimes

Evaluating likelihoods

HMMs and dependent regime models

The forward algorithm[1]

For $t = 1, ..., T$

$$f_\theta(\boldsymbol{x}_{0:t}, R_t = j) = \overbrace{f_\theta(x_t \mid R_t = j, \boldsymbol{x}_{0:t-1})}^{\text{e.g. AR(p) densities}} \sum_{i \in \mathcal{S}} p_{ij} \underbrace{f_\theta(R_{t-1} = i, \boldsymbol{x}_{0:t-1})}_{\text{value from last iteration}}$$

[1] A computationally stable *normalised* algorithm can be derived from this.

Evaluating likelihoods

HMMs and dependent regime models

The forward algorithm[1]

For $t = 1, \ldots, T$

$$f_\theta(\boldsymbol{x}_{0:t}, R_t = j) = \overbrace{f_\theta(x_t \mid R_t = j, \boldsymbol{x}_{0:t-1})}^{\text{e.g. AR(p) densities}} \sum_{i \in \mathcal{S}} p_{ij} \underbrace{f_\theta(R_{t-1} = i, \boldsymbol{x}_{0:t-1})}_{\text{value from last iteration}}$$

Then

$$L(\theta) = \sum_{j \in \mathcal{S}} f_\theta(\boldsymbol{x}_{0:T}, R_T = j)$$

- $\mathcal{O}(M^2 T)$ complexity

[1] A computationally stable *normalised* algorithm can be derived from this.

Forward algorithm – More than just likelihoods

HMMs and dependent regime models

$$\overbrace{P_\theta(R_t = i \mid \boldsymbol{x}_{0:t})}^{\text{filtered probabilities}} = \frac{f_\theta(R_t = i, \boldsymbol{x}_{0:t})}{\sum_{j \in \mathcal{S}} f_\theta(R_t = j, \boldsymbol{x}_{0:t})}$$

Similarly,

$$\overbrace{P_\theta(R_t = i \mid \boldsymbol{x}_{0:t-1})}^{\text{prediction probabilities}}$$

Smoothed probabilities

HMMs and dependent regime models

Backward algorithm[2]

For $t = T-1, \ldots, 0$

$$\overbrace{P_\theta(R_t = i \mid \boldsymbol{x}_{0:T})}^{\text{smoothed probabilities}} = \overbrace{P_\theta(R_t = i \mid \boldsymbol{x}_{0:t})}^{\text{filtered probabilities}} \sum_{j \in \mathcal{S}} p_{ij} \frac{P_\theta(R_{t+1} = j \mid \boldsymbol{x}_{0:T})}{\underbrace{P_\theta(R_{t+1} = j \mid \boldsymbol{x}_{0:t})}_{\text{prediction probabilities}}}$$

[2] Kim, 1994, *Dynamic linear models with Markov-switching*

Maximising likelihoods

HMMs and dependent regime models

EM algorithm

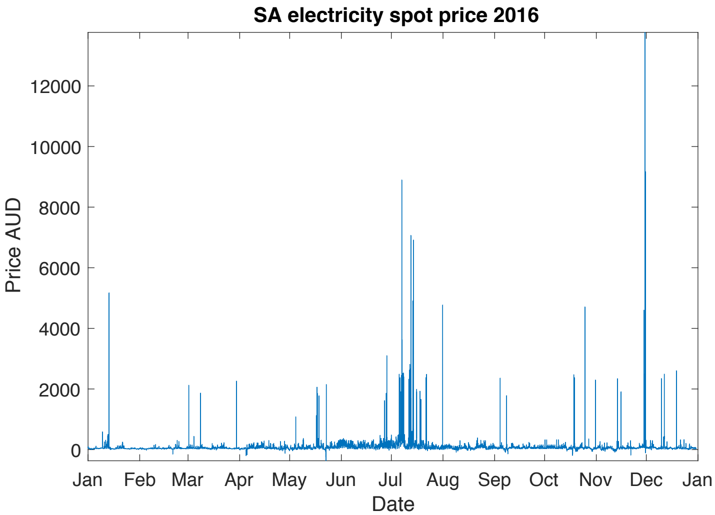
E. Construct

$$
\begin{aligned}
Q(\theta, \theta_n) &= \mathbb{E}[\log f_\theta(\boldsymbol{x}_{0:T}, \boldsymbol{R}) \mid \boldsymbol{x}_{0:T}, \theta_n] \\
&= \sum_{t=1}^{T} \sum_{i \in \mathcal{S}} \overbrace{P_{\theta_n}(R_t = i \mid \boldsymbol{x}_{0:T})}^{\text{smoothed probabilities}} \log f_\theta(x_t \mid R_t = i, \boldsymbol{x}_{0:t-1}) \\
&\quad + \sum_{i,j \in \mathcal{S}} \sum_{t=1}^{T} P_{\theta_n}(R_t = j, R_{t-1} = i | \boldsymbol{x}_{0:T}) \log(p_{ij})
\end{aligned}
$$

M. Set $\theta_{n+1} = \arg\max_\theta Q(\theta, \theta_n)$

Works well when M-step is analytic.

Our motivation – Electricity price modelling



Independent regime MRS models

Consider the MRS model

$$X_t = \begin{cases} B_t^1, & R_t = 1 \\ \vdots & \\ B_t^k, & R_t = k \\ S_t^{k+1} & R_t = k+1 \\ \vdots & \\ S_t^M & R_t = M \end{cases}$$

where

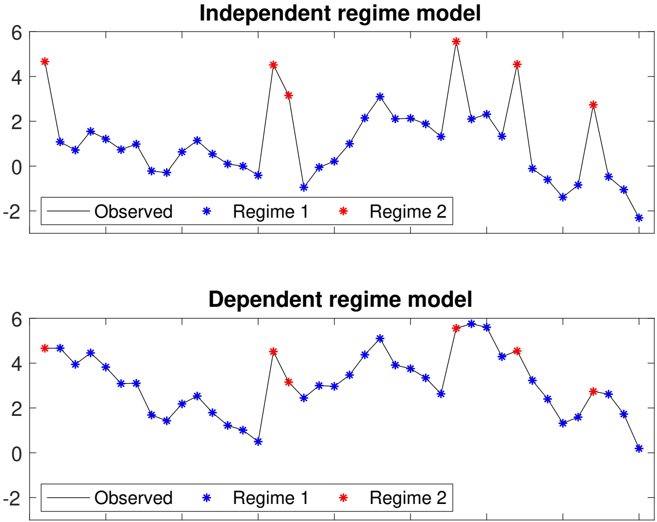
$$B_t^i = \alpha_i + \phi_i B_{t-1}^i + \sigma_i \varepsilon_t^i$$

are AR(1) and

$$S_t^j \sim \text{i.i.d. } N(\mu_j, \sigma_j^2).$$

MRS models of electricity prices

- *Independent regime* MRS models allow 'jumpy' behaviour



Independent vs dependent regimes

Independent regime model

- Multiple time series evolving in parallel
- Which process is observed is determined by the regime

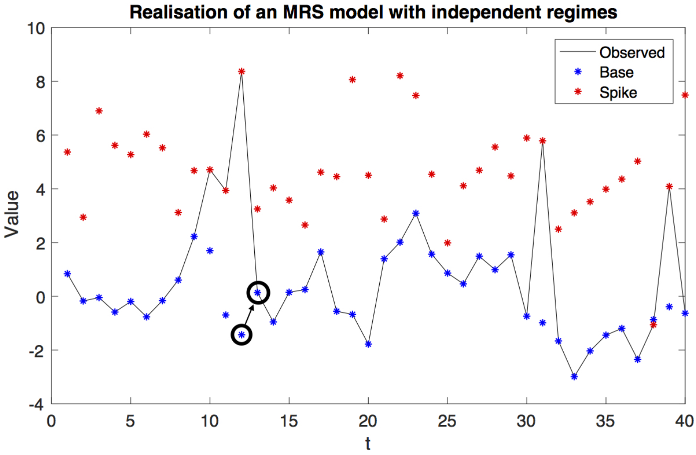
$$X_t = \begin{cases} B_t = \alpha + \phi B_{t-1} + \sigma \varepsilon_t, & R_t = 1 \\ \text{i.i.d. } N(\mu, \sigma_2^2), & R_t = 2 \end{cases}$$

Dependent regime model

- A single time series
- Dynamics (parameters) determined by the regime

$$X_t = \begin{cases} \alpha + \phi X_{t-1} + \sigma \varepsilon_t, & R_t = 1 \\ \text{i.i.d. } N(\mu, \sigma_2^2), & R_t = 2 \end{cases}$$

Independent regimes – Example



- B_t depends on B_{t-1}, but it might be unobserved

Evaluating likelihoods – Independent regime models

This idea motivates an approximate algorithm, the EM-like algorithm

The forward algorithm – Hamilton

For $t = 1, \ldots, T$

$$f_\theta(\boldsymbol{x}_{0:t}, R_t = j) = \overbrace{f_\theta(x_t \mid R_t = j, \boldsymbol{x}_{0:t-1})}^{\text{naive } \mathcal{O}(M^t) \text{ calculation}} \sum_{i \in \mathcal{S}} p_{ij} f_\theta(R_{t-1} = i, \boldsymbol{x}_{0:t-1})$$

Then

$$L(\theta) = \sum_{j \in \mathcal{S}} f_\theta(\boldsymbol{x}_{0:T}, R_T = j)$$

Maximising likelihoods – Independent regime models

An approximate algorithm – EM-like

- At time t suppose R_t and $B_{t-1}^{R_t}$ are known
- Then density of X_t would be known

$$x_t \mid R_t = i, B_{t-1}^i \sim N(\alpha_i + \phi_i B_{t-1}^i, \sigma_i^2).$$

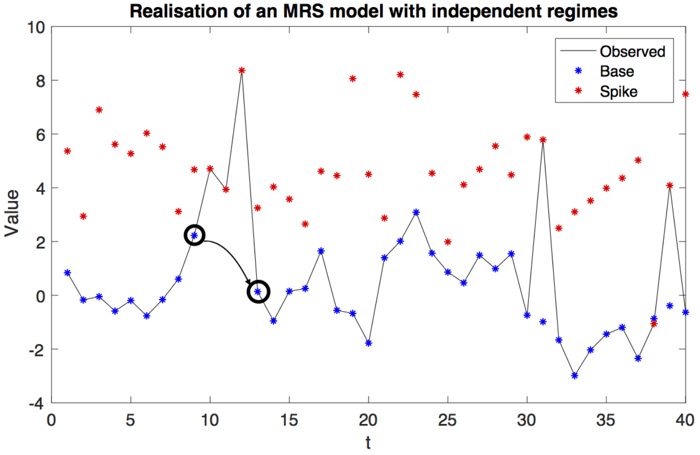
- The EM-like algorithm replaces B_{t-1}^i with approximate values

$$\begin{aligned}\widetilde{B}_{t-1}^i = P_\theta(R_{t-1} = i \mid \boldsymbol{x}_{0:t-1})x_t \\ + P_\theta(R_{t-1} \neq i \mid \boldsymbol{x}_{0:t-1})(\alpha_i + \phi_i \widetilde{B}_{t-2}^i)\end{aligned}$$

calculated iteratively.

Can be shown to perform poorly.

Independent regimes – Example



- Can integrate unobserved prices away
- B_t depends on a random lagged observation

Evaluating likelihoods – Independent regime models

Alternatively, to write down the density of x_t we need to know

- R_t,
- Time since the last visit to R_t.

Evaluating likelihoods – Independent regime models

Alternatively, to write down the density of x_t we need to know

- R_t,
- Time since the last visit to R_t.

Idea: augment $\{R_t\}$ with time-since-last-visit counters

- Let $N_{t,i} = \ell$ denote the event that the last visit to state i before time t was ℓ transitions ago.
- Define an augmented process

$$\{\boldsymbol{H}_t\} = \{(\boldsymbol{N}_t, R_t)\} = \{(N_{t,1}, ..., N_{t,k}, R_t)\}$$

- Similar ideas for hidden semi-Markov models
  - Counters for time since last transition

Facts about $\boldsymbol{H}_t = (\boldsymbol{N}_t, R_t)$

- $\{\boldsymbol{H}_t\}$ is a Markov chain.
- Transition probabilities

$$
\begin{aligned}
&P_\theta(\boldsymbol{H}_{t+1} = (\boldsymbol{N}_{t+1}, j) \mid \boldsymbol{H}_t = (\boldsymbol{N}_t, i)) \\
&= \begin{cases} p_{ij} & \text{for } i \in \{k+1, \ldots, M\}, j \in \mathcal{S}, \boldsymbol{N}_{t+1} = \boldsymbol{N}_t + \mathbf{1}, \\ p_{ij} & \text{for } i \in \{1, \ldots, k\}, j \in \mathcal{S}, \boldsymbol{N}_{t+1} = \boldsymbol{N}_t^{(-i)} + \mathbf{1}, \\ 0 & \text{otherwise}, \end{cases}
\end{aligned}
$$

where $\boldsymbol{N}_t^{(-i)} = (N_{t,1}, \ldots, N_{t,i-1}, 0, N_{t,i+1}, \ldots, N_{t,k})$.

Evaluating likelihoods

Independent regime models

A new forward algorithm[3]

For $t = 1, \ldots, T$

$$f_\theta(\boldsymbol{x}_{0:t}, \boldsymbol{H}_t = (\boldsymbol{N}_t, j)) \\ = f_\theta(x_t \mid \boldsymbol{H}_t = (\boldsymbol{N}_t, j), \boldsymbol{x}_{0:t-1}) f_\theta(\boldsymbol{H}_t = (\boldsymbol{N}_t, j), \boldsymbol{x}_{0:t-1})$$

where

$$f_\theta(\boldsymbol{H}_t = (\boldsymbol{N}_t, j), \boldsymbol{x}_{0:t-1}) \\ = \begin{cases} \sum_{i \in \mathcal{S}} p_{ij} f_\theta(\boldsymbol{H}_{t-1} = (\boldsymbol{N}_t - \mathbf{1}, i), \boldsymbol{x}_{0:t-1}), & N_{t,r} > 1 \forall r \\ \sum_{m=1}^{t} p_{ij} f_\theta(\boldsymbol{H}_{t-1} = (\boldsymbol{N}_t - \mathbf{1} + m\boldsymbol{e}_r, r), \boldsymbol{x}_{0:t-1}), & N_{t,r} = 1 \end{cases}$$

[3] A computationally stable normalised algorithm can also be derived

Evaluating likelihoods

Independent regime models

A new forward algorithm – continued

$$L(\theta) = \sum_{j \in \mathcal{S}} \sum_{\boldsymbol{N}_T} f_\theta(\boldsymbol{x}_{0:T}, \boldsymbol{H}_T = (\boldsymbol{N}_T, j))$$

Forward algorithm – More than just likelihoods

$$\overbrace{P_\theta(\boldsymbol{H}_t = (\boldsymbol{N}_t, i) \mid \boldsymbol{x}_{0:t})}^{\text{filtered probabilities}} = \frac{f_\theta(\boldsymbol{H}_t = (\boldsymbol{N}_t, i), \boldsymbol{x}_{0:t})}{\sum_{j \in \mathcal{S}} \sum_{\boldsymbol{N}_t} f_\theta(\boldsymbol{H}_t = (\boldsymbol{N}_t, j), \boldsymbol{x}_{0:t})}$$

Similarly,

$$\overbrace{P_\theta(\boldsymbol{H}_{t+1} = (\boldsymbol{N}_{t+1}, i) \mid \boldsymbol{x}_{0:t})}^{\text{prediction probabilities}}$$

Smoothed probabilities

Independent regime models

Backward algorithm

For $t = T-1, ..., 0$

$$\overbrace{P_\theta(\boldsymbol{H}_t = (\boldsymbol{N}_t, i) \mid \boldsymbol{x}_{0:T})}^{\text{smoothed probabilities}}$$

$$= \overbrace{P_\theta(\boldsymbol{H}_t = (\boldsymbol{N}_t, i) \mid \boldsymbol{x}_{0:t})}^{\text{filtered probabilities}} \sum_{j \in \mathcal{S}} p_{ij} \frac{P_\theta(\boldsymbol{H}_{t+1} = (\boldsymbol{N}_{t+1}, j) \mid \boldsymbol{x}_{0:T})}{\underbrace{P_\theta(\boldsymbol{H}_{t+1} = (\boldsymbol{N}_{t+1}, j) \mid \boldsymbol{x}_{\boldsymbol{0:t}})}_{\text{prediction probabilities}}}$$

where $\boldsymbol{N}_{t+1} = (N_{t,1}+1, ..., N_{t,i-1}+1, 1, N_{t,i+1}+1, ..., N_{t,k}+1)$ for $i \in \{1, 2, ..., k\}$ and $\boldsymbol{N}_{t+1} = \boldsymbol{N}_t + \boldsymbol{1}$ for $i \in \{k+1, ..., M\}$.

Maximising likelihoods

HMMs and dependent regime models

EM algorithm

Iterate

E. Construct

$$Q(\theta, \theta_n)$$

$$= \sum_{t=1}^{T} \sum_{i \in \mathcal{S}} \sum_{\boldsymbol{N}_t} \overbrace{P_{\theta_n}(\boldsymbol{H}_t = (\boldsymbol{N}_t, i) \mid \boldsymbol{x}_{0:T})}^{\text{smoothed probabilities}} \log f_\theta(x_t \mid R_t = i, \boldsymbol{N}_t, \boldsymbol{x}_{0:t-1})$$

$$+ \sum_{i,j \in \mathcal{S}} \log(p_{ij}) \sum_{t=1}^{T} P_{\theta_n}(R_t = j, R_{t-1} = i | \boldsymbol{x}_{0:T})$$

M. Set $\theta_{n+1} = \arg\max_\theta Q(\theta, \theta_n)$

Complexity

- Complexity is $\mathcal{O}(M^2 T^{k+1} k^k)$
- Compare to naive complexity $\mathcal{O}(M^T)$
  - M – the number of regimes
  - T – the number of observations
  - k – the number of AR(1) processes/counters

- Number of accessible states of $\{\boldsymbol{N}_t\}$ by time t is

$$\sum_{m=0}^{\min(t,k)} \binom{t}{m}\binom{k}{m} m! \approx \mathcal{O}(t^k k^k)$$

given

$$\boldsymbol{H}_0 = \begin{cases} (\mathbf{1}, i) & \text{with probability } \pi_i \\ (\boldsymbol{N}_0, j) & \text{with probability 0 otherwise,} \end{cases}$$

for some distribution $\boldsymbol{\pi} = (\pi_1, ..., \pi_M)$.

Truncation

- $\mathcal{O}(M^2 T^{k+1} k^k)$ may be large
- Truncate

$$N_{t,i} \in \{1, 2, ..., D-1, D\}, \text{ where } D << T$$

  and allow self-transitions $D \to D$
- Complexity is now $\mathcal{O}(M^2 D^k T k^k)$
- Compare to the complexity of the naive approach with the same assumptions: $\mathcal{O}(M^D T)$
- We set

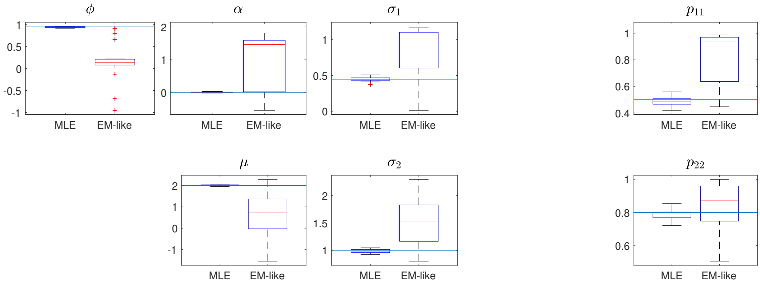
$$f(x_t \mid R_t = i, N_{t,i} = D, \boldsymbol{x}_{0:t-1}) = f(x_t \mid R_t = i, N_{t,i} = \infty)$$

  i.e. the stationary density for Regime i

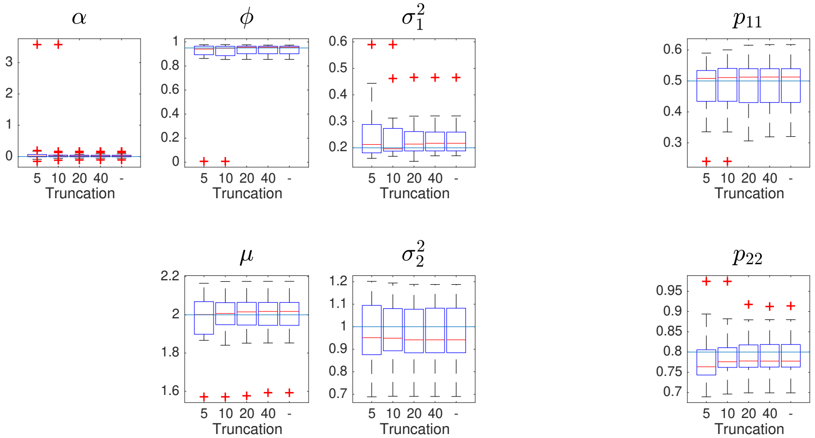
Comparison with existing method

$$X_t = \begin{cases} B_t, & \text{if } R_t = 1, \\ S_t, & \text{if } R_t = 2, \end{cases}$$

- $B_t = 0.95 B_{t-1} + 0.2\varepsilon_t$,
- $S_t \sim N(2, 1)$
- $p_{11} = 0.5$, $p_{22} = 0.8$



An example – Truncation



Summary

- Developed novel algorithms for independent regime MRS models
  - Forward, backward, EM
  - Efficient approximations to all of the above
  - Important concept was the augmented hidden chain $\{\boldsymbol{H}_t\}$
- We believe these ideas can be extended to more general processes
  - more general AR(p) processes
  - other Markovian processes?
- *Estimation of Markovian-Regime-Switching models with independent regimes*, Lewis, Nguyen, Bean, To be submitted

An example – Consistency