Modeling Time Series with Changes in Regime

Paweł Jamer

Faculty of Mathematics and Information Science
Warsaw University of Technology

January 17, 2012

Contents:

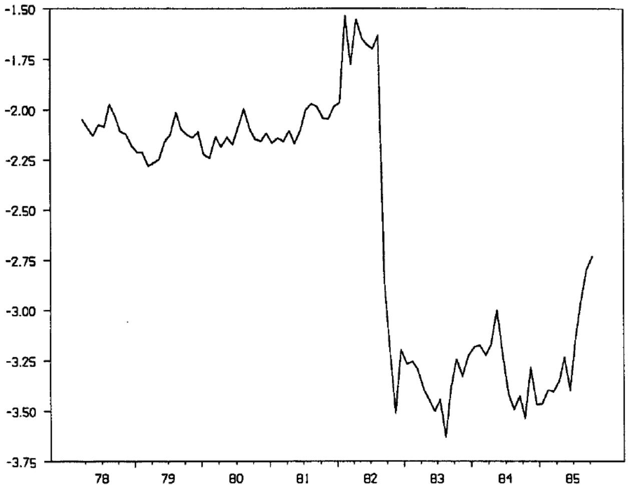

Figure: Rogers, 1992

Model 1

We might use a model such as

$$\begin{cases} y_t - \mu_1 = \phi\left(y_{t-1} - \mu_1\right) + \varepsilon_t, & \text{before 1982} \\ y_t - \mu_2 = \phi\left(y_{t-1} - \mu_2\right) + \varepsilon_t, & \text{after 1982} \end{cases},$$

where $\mu_1 > \mu_2$.

However...

- If the process has changed in the past, clearly it could also change again in the future, and this prospect should be taken into account in forming a forecast.
- Change in regime should not be regarded as the outcome of a perfectly foreseeable, deterministic event.

Model 2

By modification of model 1 we obtain

$$y_t - \mu_{s_t^*} = \phi \left(y_{t-1} - \mu_{s_t^*} \right) + \varepsilon_t,$$

where

- s_t^* — unobserved random variable called regime.
  - $\{s_t^*\}$ take on only discrete values.
  - We can consider $\{s_t^*\}$ as a Markov chain.

Definition

Markov chain is a stochastic process $\{s_t\}$ which satisfy Markov property

$$P\left(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots\right) = P\left(s_t = j | s_{t-1} = i\right).$$

Define:

- $p_{ij} = P\left(s_t = j | s_{t-1} = i\right)$ — transition probability,
  - $(\forall i) \sum_{j=1}^{N} p_{ij} = 1.$
- $\mathbf{P} = \left[p_{ij}\right]_{i,j=1,2,\ldots,N}$ — transition matrix.
- π such that $\boldsymbol{P}\pi = \pi$ — ergodic probabilities.
  - $\lim_{n\to\infty} \boldsymbol{P}^n = \pi\mathbf{1}'$.

Let's define new random variable

$$\xi_t = \begin{cases} \boldsymbol{e}_1 = (1,0,0,\dots,0)' & \text{when} \quad s_t = 1 \\ \boldsymbol{e}_2 = (0,1,0,\dots,0)' & \text{when} \quad s_t = 2 \\ \vdots & \vdots \qquad \vdots \\ \boldsymbol{e}_N = (0,0,0,\dots,1)' & \text{when} \quad s_t = N \end{cases}.$$

We have

$$P\left(\xi_{t+1} = \boldsymbol{e}_j \mid \xi_t = \boldsymbol{e}_i\right) = P\left(s_{t+1} = j \mid s_t = i\right) = p_{ij}.$$

Thus, the conditional expectation of ξ_{t+1} given $\xi_t = \boldsymbol{e}_i$ is given by

$$\mathbb{E}\left(\xi_{t+1} \mid \xi_t = \boldsymbol{e}_i\right) = \boldsymbol{P}\boldsymbol{e}_i$$

and finally

$$\mathbb{E}\left(\xi_{t+1} \mid \xi_t, \xi_{t-1}, \dots\right) = \mathbb{E}\left(\xi_{t+1} \mid \xi_t\right) = \boldsymbol{P}\xi_t.$$

So it is possible to express a Markov chain in the form

First-Order Vector Autoregression

$$\xi_{t+1} = \boldsymbol{P}\xi_t + \boldsymbol{v}_{t+1},$$

where

- $\boldsymbol{v}_{t+1} = \xi_{t+1} - \mathbb{E}\left(\xi_{t+1} \mid \xi_t, \xi_{t-1}, \ldots\right)$. — martingale difference sequence (MDS).

Above expression implies that

$$\xi_{t+m} = \boldsymbol{v}_{t+m} + \boldsymbol{P}\boldsymbol{v}_{t+m-1} + \boldsymbol{P}^2\boldsymbol{v}_{t+m-2} + \ldots + \boldsymbol{P}^{m-1}\boldsymbol{v}_{t+1} + \boldsymbol{P}^m\xi_t,$$

If follows that **m-period-ahead forecasts** for a Markov chain can be calculated from

$$\mathbb{E}\left(\xi_{t+m} \mid \xi_t, \xi_{t-1}, \ldots\right) = \boldsymbol{P}^m\xi_t.$$

Definition

Density of y_t conditional on the random variable s_t taking on the value j is

$$f\left(y_t | s_t = j; \theta\right) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{\left(y_t - \mu_j\right)^2}{2\sigma_j^2}}, j = 1, 2, \ldots, N,$$

where

- θ is a vector of population parameters that includes $\mu_1, \mu_2, \ldots, \mu_N$ and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$.

- $\pi_j = P\left(s_t = j; \theta\right), j = 1, 2, \ldots, N.$

Let $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta} = \left(\mu_1, \ldots, \mu_N, \sigma_1^2, \ldots, \sigma_N^2, \pi_1, \ldots, \pi_N\right)'.$$

From the formula of conditional probability

$$p\left(y_t, s_t = j; \boldsymbol{\theta}\right) = f\left(y_t \mid s_t = j; \boldsymbol{\theta}\right) P\left(s_t = j; \boldsymbol{\theta}\right) = \frac{\pi_j}{\sqrt{2\pi}\sigma_j} e^{-\frac{\left(y_t - \mu_j\right)^2}{2\sigma_j^2}}.$$

The unconditional density of y_t can be found by summing above expression over all possible values for j

$$f\left(y_t; \boldsymbol{\theta}\right) = \sum_{j=1}^{N} p\left(y_t, s_t = j; \boldsymbol{\theta}\right) = \sum_{j=1}^{N} \frac{\pi_j}{\sqrt{2\pi}\sigma_j} e^{-\frac{\left(y_t - \mu_j\right)^2}{2\sigma_j^2}}.$$

Regime Inference

$$P\left(s_t = j \mid y_t; \theta\right) = \frac{p\left(y_t, s_t = j; \theta\right)}{f\left(y_t; \theta\right)} = \frac{\pi_j f\left(y_t \mid s_t = j; \theta\right)}{f\left(y_t; \theta\right)}$$

EM Algorithm

1. Choose initial value of $\theta^{(0)}$.
2. For $j = 1, 2, \ldots, N$ calculate
   1. $\hat{\mu}_j^{(k+1)} = \sum_{t=1}^{T} y_t P\left(s_t = j | y_t; \hat{\theta}^{(k)}\right) / \sum_{t=1}^{T} P\left(s_t = j | y_t; \hat{\theta}^{(k)}\right)$,
   2. $\left(\hat{\sigma}_j^2\right)^{(k+1)} = \sum_{t=1}^{T} \left(y_t - \hat{\mu}_j\right)^2 P\left(s_t = j | y_t; \hat{\theta}^{(k)}\right) / \sum_{t=1}^{T} P\left(s_t = j | y_t; \hat{\theta}^{(k)}\right)$,
   3. $\hat{\pi}_j^{(k+1)} = \frac{1}{T} \sum_{t=1}^{T} P\left(s_t = j \mid y_t; \hat{\theta}^{(k)}\right)$.
3. If $\left|\theta^{(k+1)} - \theta^{(k)}\right| > \varepsilon$ repeat step 2.

Let:

- $\boldsymbol{y}_t$ — $n \times 1$ vector of observed endogenous variables,
- $\boldsymbol{x}_t$ — $k \times 1$ vector of observed exogenous variables,
- $\mathcal{Y}_t = (\boldsymbol{y}'_t, \boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{-m}, \boldsymbol{x}'_t, \boldsymbol{x}'_{t-1}, \ldots, \boldsymbol{x}'_{-m})'$ — vector containing all observation obtained through date t,
- s_t — regime.

Conditional density of $\boldsymbol{y}_t$

$$f\left(\boldsymbol{y}_t \mid s_t = j, \boldsymbol{x}_t, \mathcal{Y}_{t-1}, \alpha\right)$$

- α — vector of parameters characterizing the conditional density,
- η_t — $N \times 1$ vector contains all conditional densities.

Example

First-Order Autoregression

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t,$$

where

- ε_t iid $\mathcal{N}\left(0, \sigma^2\right)$,
- $\{s_t\}$ — N-state Markov chain,
- s_{t_1}, ε_{t_2} — independent for all t_1 and t_2.

- $\boldsymbol{\alpha} = \left(c_1, c_2, \ldots, c_N, \phi_1, \phi_2, \ldots, \phi_N, \sigma^2\right)$.
- $f\left(\boldsymbol{y}_t \mid s_t = j, \boldsymbol{x}_t, \boldsymbol{\mathcal{Y}}_{t-1}, \boldsymbol{\alpha}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(y_t - c_j - \phi_j y_{t-1}\right)^2}{2\sigma^2}}$

Example

Define variable s_t as follows:

$$s_t = \begin{cases} 1 & \text{if } s_t^* = 1 \text{ and } s_{t-1}^* = 1 \\ 2 & \text{if } s_t^* = 2 \text{ and } s_{t-1}^* = 1 \\ 3 & \text{if } s_t^* = 1 \text{ and } s_{t-1}^* = 2 \\ 4 & \text{if } s_t^* = 2 \text{ and } s_{t-1}^* = 2 \end{cases},$$

where s_t^* — two-state Markov chain.

Then s_t follows a four-state Markov chain with transition matrix

$$\boldsymbol{P} = \begin{bmatrix} p_{11}^* & 0 & p_{11}^* & 0 \\ p_{12}^* & 0 & p_{12}^* & 0 \\ 0 & p_{21}^* & 0 & p_{21}^* \\ 0 & p_{22}^* & 0 & p_{22}^* \end{bmatrix},$$

where $p_{ij}^* = P\left(s_t^* = j \mid s_{t-1}^* = i\right)$

Let

$$\hat{\xi}_{t|T} = \mathbb{E}\left(\xi_t \mid \mathcal{Y}_T; \theta\right) = \begin{bmatrix} P\left(s_t = 1 \mid \mathcal{Y}_T; \theta\right) \\ P\left(s_t = 2 \mid \mathcal{Y}_T; \theta\right) \\ \vdots \\ P\left(s_t = N \mid \mathcal{Y}_T; \theta\right) \end{bmatrix}$$

and

$$\eta_t = \begin{bmatrix} f\left(\boldsymbol{y}_t \mid s_t = 1, \boldsymbol{x}_t, \mathcal{Y}_{t-1}; \theta\right) \\ f\left(\boldsymbol{y}_t \mid s_t = 2, \boldsymbol{x}_t, \mathcal{Y}_{t-1}; \theta\right) \\ \vdots \\ f\left(\boldsymbol{y}_t \mid s_t = N, \boldsymbol{x}_t, \mathcal{Y}_{t-1}; \theta\right) \end{bmatrix}.$$

By multiplying $\hat{\xi}_{t|T}$ and η_t we obtain

$$\hat{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t = \left[p\left(\boldsymbol{y}_t, s_t = j \mid \boldsymbol{x}_t, \boldsymbol{\mathcal{Y}}_{t-1}; \boldsymbol{\theta}\right)\right]_{j=1,2,\ldots,N}.$$

So we can write

$$f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \boldsymbol{\mathcal{Y}}_{t-1}; \boldsymbol{\theta}\right) = \mathbf{1}'\left(\hat{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t\right).$$

Above equations gives us

$$P\left(s_t = j \mid \boldsymbol{\mathcal{Y}}_t; \boldsymbol{\theta}\right) = P\left(s_t = j \mid \boldsymbol{y}_t, \boldsymbol{x}_t, \boldsymbol{\mathcal{Y}}_{t-1}; \boldsymbol{\theta}\right)$$

$$= \frac{p\left(\boldsymbol{y}_t, s_t = 1 \mid \boldsymbol{x}_t, \boldsymbol{\mathcal{Y}}_{t-1}; \boldsymbol{\theta}\right)}{f\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \boldsymbol{\mathcal{Y}}_{t-1}; \boldsymbol{\theta}\right)} = \frac{p\left(\boldsymbol{y}_t, s_t = 1 \mid \boldsymbol{x}_t, \boldsymbol{\mathcal{Y}}_{t-1}; \boldsymbol{\theta}\right)}{\mathbf{1}'\left(\hat{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t\right)}.$$

From where directly follows

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t}{\mathbf{1}'\left(\hat{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t\right)}.$$

Let's take expectations of

$$\xi_{t+1} = \boldsymbol{P}\xi_t + \boldsymbol{v}_{t+1},$$

conditional on $\mathcal{Y}_t$

$$\mathbb{E}\left(\xi_{t+1} \mid \mathcal{Y}_t\right) = \mathbb{E}\left(\boldsymbol{P}\xi_t + \boldsymbol{v}_{t+1} \mid \mathcal{Y}_t\right) = \boldsymbol{P}\mathbb{E}\left(\xi_t \mid \mathcal{Y}_t\right) + \mathbb{E}\left(\boldsymbol{v}_{t+1} \mid \mathcal{Y}_t\right).$$

Note that $\boldsymbol{v}_{t+1}$ is a martingale difference sequence with respect to $\mathcal{Y}_t$, so that above equality becomes

$$\xi_{t+1|t} = \boldsymbol{P}\xi_{t|t}.$$

Regime Inference

1. $\hat{\xi}_{t|t} = \frac{\left(\hat{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t\right)}{\mathbf{1}'\left(\hat{\xi}_{t|t-1} \odot \boldsymbol{\eta}_t\right)}$,
2. $\hat{\xi}_{t+1|t} = \boldsymbol{P}\hat{\xi}_{t|t}$.

- Starting value: $\hat{\xi}_{1|0} = \boldsymbol{\rho}$
  - $\boldsymbol{\rho}$ — $N \times 1$ vector of nonnegative constants summing to unity.
  - For example $\boldsymbol{\rho} = \frac{1}{N} \cdot \mathbf{1}$, $\boldsymbol{\rho} = \boldsymbol{\pi}$.

Regime Inference

1. **Forecast**: $\hat{\xi}_{t+m|t} = \boldsymbol{P}^m \hat{\xi}_{t|t}$.
2. **Smoothed**: $\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \left(\boldsymbol{P}' \cdot \left(\hat{\xi}_{t+1|T} (\div) \hat{\xi}_{t+1|t}\right)\right), t < T$.

Forecast depends on regime

$$\mathbb{E}\left(\boldsymbol{y}_{t+1} \mid s_{t+1} = j, \mathcal{Y}_t; \theta\right) = \int \boldsymbol{y}_{t+1} f\left(\boldsymbol{y}_{t+1} \mid s_{t+1} = j, \mathcal{Y}_t; \theta\right) d\boldsymbol{y}_{t+1}$$

Example

For example, for the AR(1) specification

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t,$$

a forecast is given by

$$\mathbb{E}\left(y_{t+1} \mid s_{t+1} = j, \mathcal{Y}_t; \theta\right) = c_j + \phi_j y_t.$$

$$\mathbb{E}\left(\boldsymbol{y}_{t+1} \mid \boldsymbol{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right) = \int \boldsymbol{y}_{t+1} f\left(\boldsymbol{y}_{t+1} \mid \boldsymbol{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right) d\boldsymbol{y}_{t+1}$$

$$= \int \boldsymbol{y}_{t+1} \left(\sum_{j=1}^{N} p\left(\boldsymbol{y}_{t+1}, s_{t+1} = j \mid \boldsymbol{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right) \right) d\boldsymbol{y}_{t+1}$$

$$= \int \boldsymbol{y}_{t+1} \left(\sum_{j=1}^{N} f\left(\boldsymbol{y}_{t+1} \mid s_{t+1} = j, \boldsymbol{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right) P\left(s_{t+1} = j \mid \boldsymbol{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right) \right) d\boldsymbol{y}_{t+1}$$

$$= \sum_{j=1}^{N} P\left(s_{t+1} = j \mid \boldsymbol{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right) \int \boldsymbol{y}_{t+1} f\left(\boldsymbol{y}_{t+1} \mid s_{t+1} = j, \boldsymbol{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right) d\boldsymbol{y}_{t+1}$$

Forecast independent of regime

$$\mathbb{E}\left(\boldsymbol{y}_{t+1} \mid \mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right) = \sum_{j=1}^{N} P\left(s_{t+1} = j \mid \mathcal{Y}_t; \boldsymbol{\theta}\right) \mathbb{E}\left(\boldsymbol{y}_{t+1} \mid s_{t+1} = j, \boldsymbol{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\right).$$

Log-Likelihood Function

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \log f(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \mathcal{Y}_{t-1}; \theta),$$

where

- $f(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \mathcal{Y}_{t-1}; \theta) = \mathbf{1}' \left(\hat{\xi}_{t|t-1} \odot \eta_t\right)$

Numerical Methods:

- Grid Search,
- Steepest Ascent,
- Newton-Raphson,
- Davidon-Fletcher-Powell.

Assume that

- $p_{ij} \geq 0$,
- $(\forall i) \sum_{j=1}^{N} p_{ij} = 1$,
- $\hat{\xi}_{i|0} = \rho$ — fixed value unrelated to the other parameters.

Then the maximum likelihood estimates for the transition probabilities satisfy

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} P\left(s_t = j, s_{t-1} = i \mid \mathcal{Y}_T; \hat{\theta}\right)}{\sum_{t=2}^{T} P\left(s_{t-1} = i \mid \mathcal{Y}_T; \hat{\theta}\right)}$$

If the vector of initial probabilities ρ is

- regarded as a separate vector of parameters
- constrain only by $\mathbf{1}'\rho = 1$ and $\rho \geq 0$.

The maximum likelihood estimate of ρ turns out to be the smoothed inference about the initial state:

$$\hat{\rho} = \hat{\xi}_{1|T}$$

The maximum likelihood estimate of the vector α is characterized by

$$\sum_{t=1}^{T}\left(\frac{\partial \log \eta_t}{\partial \alpha'}\right)' \hat{\xi}_{t|T} = \mathbf{0}.$$

Bibliography

James D. Hamilton,
Time Series Analysis,
Princeton University Press.

Thank you!