1.3 Regime switching models

A potentially useful approach to model nonlinearities in time series is to assume different behavior (structural break) in one subsample (or regime) to another. If the dates of the regimes switches are known, modeling can be worked out with dummy variables. For example, consider the following regression model

$$y_t = \mathbf{x}_t{}'\beta_{S_t} + u_t, \quad t = 1, \ldots, T,$$

where

$$u_t \sim \text{NID}(0, \sigma^2_{S_t}),$$

$$\beta_{S_t} = \beta_0(1 - S_t) + \beta_1 S_t,$$

$$\sigma^2_{S_t} = \sigma^2_0(1 - S_t) + \sigma^2_1 S_t,$$

and

$$S_t = 0 \text{ or } 1, \quad \text{(Regime 0 or 1)}.$$

Thus under regime 1(0), the coefficient parameter vector is $\beta_{1(0)}$ and error variance $\sigma^2_{1(0)}$.

For the sake of simplicity consider an AR(1) model. Usually it is assumed that the possible difference between the regimes is a mean and volatility shift, but no autoregressive change. That is

$$y_t = \mu_{S_t} + \phi(y_{t-1} - \mu_{s_{t-1}}) + u_t,\ u_t \sim \text{NID}(0, \sigma^2_{S_t}),$$

where $\mu_{S_t} = \mu_0(1 - S_t) + \mu_1 S_t$, and $\sigma^2_{S_t}$ as defined above. If S_t, $t = 1, \ldots, T$ is known a priori, then the problem is just a usual dummy variable autoregression problem.

In practice, however, the prevailing regime is not usually directly observable. Denote then

$$P(S_t = j | S_{t-1} = i) = p_{ij}, \quad (i, j = 0, 1),$$

called transition probabilities, with $p_{i0} + p_{i1} = 1$, $i = 0, 1$. This kind of process, where the current state depends only on the state before, is called a Markov process, and the model a Markov switching model in the mean and the variance.

The probabilities in a Markov process can be conveniently presented in matrix form:

$$\begin{pmatrix} P(S_t = 0) \\ P(S_t = 1) \end{pmatrix} = \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} \begin{pmatrix} P(S_{t-1} = 0) \\ P(S_{t-1} = 1) \end{pmatrix}$$

Estimation of the transition probabilities p_{ij} is usually done (numerically) by maximum likelihood as follows.

The conditional probability density function for the observations y_t given the state variables S_t, S_{t-1} and the previous observations $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ is

$$f(y_t|S_t, S_{t-1}, \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2_{S_t}}} \exp\left\{-\frac{[y_t - \mu_{S_t} - \phi(y_{t-1} - \mu_{S_{t-1}})]^2}{2\sigma^2_{S_t}}\right\},$$

because

$$u_t = y_t - \mu_{S_t} - \phi(y_{t-1} - \mu_{S_{t-1}})) \sim \text{NID}(0, \sigma^2_{S_t}).$$

The chain rule for conditional probabilities** yields then for the joint probability density function for the variables y_t, S_t, S_{t-1} given past information $\mathcal{F}_{t-1}$

$$f(y_t, S_t, S_{t-1}|\mathcal{F}_{t-1}) = f(y_t|S_t, S_{t-1}, \mathcal{F}_{t-1})P(S_t, S_{t-1}|\mathcal{F}_{t-1}),$$

such that the log-likelihood function to be maximized with respect to the unknown parameters becomes (exercise)

$$\ell(\theta) = \sum_{t=1}^{T} \ell_t(\theta),$$

where

$$\ell_t(\theta) = \log\left[\sum_{S_t=0}^{1}\sum_{S_{t-1}=0}^{1} f(y_t|S_t, S_{t-1}, \mathcal{F}_{t-1})P(S_t, S_{t-1}|\mathcal{F}_{t-1})\right],$$

$\theta = (p, q, \phi, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2)$ and the transition probabilities $p := P(S_t = 0|S_{t-1} = 0)$, and $q := P(S_t = 1|S_{t-1} = 1)$.

**<u>Chain Rule for conditional probabilities:</u>

$$P(A \cap B|C) = P(A|B \cap C) \cdot P(B|C)$$

In order to find the conditional joint probabilities $P(S_t, S_{t-1}|\mathcal{F}_{t-1})$ we use again the chain rule for conditional probabilities:

$$\begin{aligned} P(S_t, S_{t-1}|\mathcal{F}_{t-1}) &= P(S_t|S_{t-1}, \mathcal{F}_{t-1})P(S_{t-1}|\mathcal{F}_{t-1}) \\ &= P(S_t|S_{t-1}) \cdot P(S_{t-1}|\mathcal{F}_{t-1}), \end{aligned}$$

where we have used the Markov property $P(S_t|S_{t-1}, \mathcal{F}_{t-1}) = P(S_t|S_{t-1})$.

We note that the problem reduces to estimating the time dependent state probabilities $P(S_{t-1}|\mathcal{F}_{t-1})$, and weighting them with the transition probabilities $P(S_t|S_{t-1})$ to obtain the joint probabilities $P(S_t, S_{t-1}|\mathcal{F}_{t-1})$.

This can be achieved as follows:

First, let $P(S_0 = 1|\mathcal{F}_0) = P(S_0 = 1) = \pi$ be given (such that $P(S_0 = 0) = 1 - \pi$). Then the probabilities $P(S_{t-1}|\mathcal{F}_{t-1})$ and the joint probabilities $P(S_t, S_{t-1}|\mathcal{F}_{t-1})$ are obtained using the following two steps algorithm

1. Given $P(S_{t-1} = i|\mathcal{F}_{t-1})$, $i = 0, 1$, at the beginning of time t (the t'th iteration),

$$P(S_t = j, S_{t-1} = i|\mathcal{F}_{t-1}) = P(S_t = j|S_{t-1} = i)P(S_{t-1} = i|\mathcal{F}_{t-1}).$$

2. Once y_t is observed, we update the information set $\mathcal{F}_t = \{\mathcal{F}_{t-1}, y_t\}$ and the probabilities by backwards application of the chain rule and using the law of total probability:

$$P(S_t = j, S_{t-1} = i|\mathcal{F}_t) = P(S_t = j, S_{t-1} = i|\mathcal{F}_{t-1}, y_t)$$

$$= \frac{f(S_t=i, S_{t-1}=j, y_t|\mathcal{F}_{t-1})}{f(y_t|\mathcal{F}_{t-1})}$$

$$= \frac{f(y_t|S_t=j, S_{t-1}=i, \mathcal{F}_{t-1})P[S_t=j, S_{t-1}=i|\mathcal{F}_{t-1}]}{\sum_{s_t, s_{t-1}=0}^{1} f(y_t|s_t, s_{t-1}, \mathcal{F}_{t-1})P[S_t=s_t, S_{t-1}=s_{t-1}|\mathcal{F}_{t-1}]}$$

We may then return to step 1 by applying again the law of total probability:

$$P(S_t = s_t|\mathcal{F}_t) = \sum_{s_{t-1}=0}^{1} P(S_t = s_t, S_{t-1} = s_{t-1}|\mathcal{F}_t).$$

Once we have the joint probability for the time point t, we can calculate the likelihood $\ell_t(\theta)$. The maximum likelihood estimates for θ is then obtained iteratively maximizing the likelihood function by updating the likelihood function at each iteration with the above algorithm.

Steady state probabilities

$P(S_0 = 1|\mathcal{F}_0)$ and $P(S_0 = 0|\mathcal{F}_0)$ are called the steady state probabilities, and, given the transition probabilities p and q, are obtained as (exercise):

$$P(S_0 = 1|\mathcal{F}_0) = \frac{1-p}{2-p-q},$$
$$P(S_0 = 0|\mathcal{F}_0) = \frac{1-q}{2-p-q}.$$

Smoothed probabilities

Recall that the state S_t is unobserved. However, once we have estimated the model, we can make inferences on S_t using all the information from the sample. This gives us

$$P(S_t = j|\mathcal{F}_T), \quad j = 0, 1,$$

which are called the smoothed probabilities.

Note. In the estimation procedure we derived $P(S_t = j|\mathcal{F}_t)$ which are usually called the filtered probabilities.

Expected duration

The expected length the system is going to stay in state j can be calculated from the transition probabilities. Let D_j denote the number of periods the system is in state j. Application of the chain rule and the Markov property yield for the probability to stay k periods in state j (exercise)

$$P(D_j = k) = p_{jj}^{k-1}(1 - p_{jj}),$$

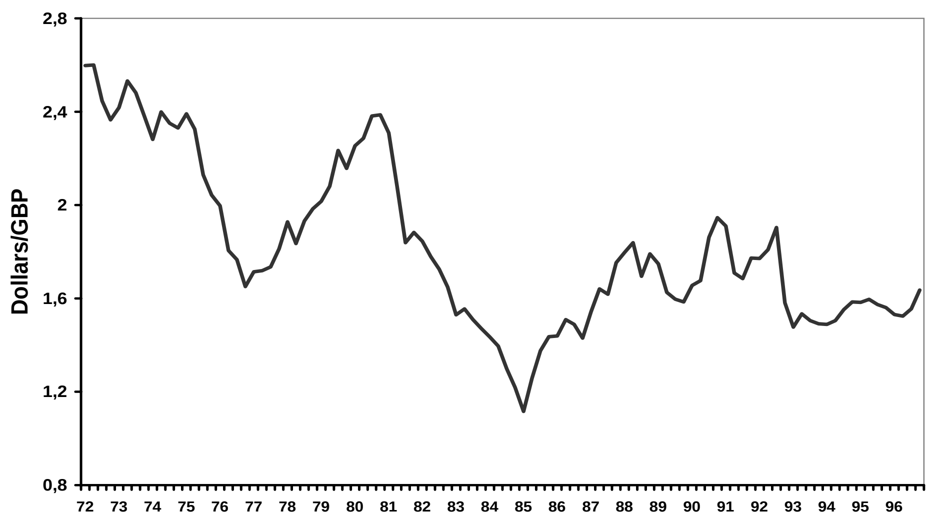
which implies for the expected duration of that state (exercise)

$$\mathsf{E}(D_j) = \sum_{k=0}^{\infty} kP(D_j = k) = \frac{1}{1 - p_{jj}}.$$

Note that in our case $p_{00} = p$ and $p_{11} = q$.

Example. Are there long swings in the dollar/sterling exchange rate?

Consider the following time series of the USD/GBP exchange rate from 1972I to 1996IV:



It appears that rather than being a simple random walk, the time series consists of distinct time periods of both upwards and downwards trends. In that case it may be put in a Markov switching framework as follows.

Model changes Δx_t in the exchange rate as

$$\Delta x_t = \alpha_0 + \alpha_1 S_t + \epsilon_t,$$

so that $\Delta x_1 \sim N(\mu_0, \sigma_0^2)$ when $S_t = 0$ and $\Delta x_t \sim N(\mu_1, \sigma_1^2)$, when $S_t = 1$, where $\mu_0 = \alpha_0$ and $\mu_1 = \alpha_0 + \alpha_1$. Parameters μ_0 and μ_1 constitute two different drifts (if $\alpha_1 \neq 0$) in the random walk model.

Estimating the model from quarterly with sample period 1972I to 1996IV gives

Parameter	Estimate	Std err
μ_0	2.605	0.964
μ_1	-3.277	1.582
σ_0^2	13.56	3.34
σ_1^2	20.82	4.79
p (regime 1)	0.857	0.084
q (regime 0)	0.866	0.097

The expected length of stay in regime 0 is given by $1/(1-p) = 7.0$ quarters, and in regime 1 $1/(1-q) = 7.5$ quarters.

Example. Suppose we are interested whether the market risk of a share is dependent on the level of volatility on the market. In the CAPM world the market risk of a stock is measured by β.

Consider for the sake of simplicity only the cases of high and low volatility.

The market model is

$$y_t = \alpha_{S_t} + \beta_{S_t} x_t + \epsilon_t,$$

where $\alpha_{S_t} = \alpha_0(1 - S_t) + \alpha_1 S_t$, $\beta_{S_t} = \beta_0(1 - S_t) + \beta_1 S_t$ and $\epsilon_t \sim N(0, \sigma^2_{S_t})$ with $\sigma^2_t = \sigma^2_0(1 - S_t) + \sigma^2_1 S_t$.
Estimating the model yields

Parameter	Estimate	Std Err	t-value	p-value
α_0	-0.0075	0.0186	-0.40	0.685
α_0	0.0849	0.0499	1.70	0.089
β_0	0.9724	0.0224	43.47	0.000
β_1	1.8112	0.0666	27.19	0.000
σ^2_0	0.7183	0.0150	48.01	0.000
σ^2_1	1.3072	0.0267	48.89	0.000
State Prob				
P(High|High)	0.96340			
P(Low|High)	0.03660			
P(High|Low)	0.01692			
P(Low|Low)	0.98308			
P(High)	0.68393			
P(Low)	0.31607			
Log-likelihood -3186.064				

The empirical results give evidence that the stock's market risk depends on the level of stock volatility. The expected duration of high volatility is $1/(1 - .9634) \approx 27$ days, and for low volatility 59 days.

Market returns with high-low volatility probabilities