SWITCHING REGIME ESTIMATION.

Introduction.

Usually in econometrics we assume that observed data has been drawn from some data generating mechanism which may be related to some specific economic theory or simply represent the true relationship between a set of variables. An underlying assumption in all econometric models is that all the observations have been drawn from the same distribution conditional on some constant parameter set. It is very unlikely that economic time series can be characterized in such a way since we expect changes in the properties of the data when there is a change in macroeconomic policy, say from free floating to target exchange rates, or even a more dramatic change from war time to peace time.

The standard econometric approach consist of trying to detect the existence of these changes in regime using different types of parameter constancy tests, and then impose dummy variables to account for these changes. By doing this, the econometrician could ensure parameter constancy within regime. Nevertheless this procedure might be very rigid and may lead to the use of models with too many dummy variables. How should we model a change in the process followed by a particular time series? Suppose that the series under scrutiny has a break in the unconditional mean at time t_1 . For data prior to t_1 we might use a model such as

$$y_t - \mu_1 = \phi(y_{t-1} - \mu_1) + \varepsilon_t \quad for\ t_1 < t$$

and for data after t_1

$$y_t - \mu_2 = \phi(y_{t-1} - \mu_2) + \varepsilon_t \quad for\ t \geq t_1$$

Even if this specification may capture the break at t_1 , is not a satisfactory time series model. For example, how are we to forecast a series described as above. Also, if the process has change in the past it could also change again. The change in regime does not need to be the outcome of a perfectly foreseeable, deterministic event. *Rather the change itself may be regarded as a random variable.* A complete time series model would therefore include a description of the probability law governing the change from μ_1 to μ_2

We might consider the process to be influenced by an unobserved random variable x_t , which is called the *state* or *regime.* This variable may take different values at date t; if $x_t = 1$, then the process is in regime 1, while $x_t = 2$ means that the process is in regime 2. Therefore we can write this model as

$$y_t - \mu_{x_t} = \phi_1(y_{t-1} - \mu_{x_{t-1}}) + \varepsilon_t$$

where the unconditional mean takes the values μ_1 when $x_t = 1$ and the value μ_2 when $x_t = 2$.

We then need a description of the time series process for the unobserved variable. Since x_t only takes discrete values we need to model this process using a discrete-valued random variable. The easiest is a Markov chain.

Properties of Markov Processes.

Let x_t be a random variable (which denotes the unobserved state of the system) that can take values 0 and 1. If the probability that x_t takes a particular value at time t, only depends on its value at $t-1$, this variable is governed by a Markov process.

$$P((x_t = i|x_{t-1} = j, x_{t-2} = k...) = P((x_t = i|x_{t-1} = j)$$

The process is summarized by the probabilities: $P(x_t = 0|x_{t-1} = 0) = q$ and $P(x_t = 1|x_{t-1} = 1) = p$.

These information is usually summarized in what is called the transition Matrix or transition probability matrix.

	0	1	(time t-1)
0	q	$(1-p)$	
1	$(1-q)$	p	
(time t)			

Autoregressive Representation of Markov Process.

$$\begin{bmatrix} 1 - x_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} q & (1-p) \\ (1-q) & p \end{bmatrix} \begin{bmatrix} 1 - x_t \\ x_t \end{bmatrix} + \begin{bmatrix} \zeta_{1,t+1} \\ \zeta_{2,t+1} \end{bmatrix}$$

or

$$W_{t+1} = \mathbf{P} W_t + U_{t+1}$$

where

$W_{t+1} = \begin{bmatrix} 1 - x_{t+1} \\ x_{t+1} \end{bmatrix}$ and $U_{t+1} = \begin{bmatrix} \zeta_{1,t+1} \\ \zeta_{2,t+1} \end{bmatrix}$ and $E_t\zeta_{1,t+1} = 0$, $E_t\zeta_{2,t+1} = 0$, so $E_t U_{t+1} = 0$

Then it follows that W_t is a random vector that takes the value

$W_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ when $x_t = 0$

and

$W_t = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ = when $x_t = 1$

$E(W_{t+1}|x_t = i) = \begin{bmatrix} q \\ 1-q \end{bmatrix}$ when $i = 0$

$E(W_{t+1}|x_t = i) = \begin{bmatrix} 1-p \\ p \end{bmatrix}$ when $i = 1$.

The above vectors are simply column $i + 1$ of the transition matrix.

Then

$$E(W_{t+1}|W_t) = \mathbf{P}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$U_{t+1} = W_{t+1} - E(W_{t+1}|W_t) = W_{t+1} - \mathbf{P}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Now notice that the second row gives

$$x_{t+1} = (1-q) + (-1+p+q)x_t + \zeta_{2,t+1}$$

This expression can be recognized as an AR(1) process with constant term 1−q and autoregressive coefficient $(-1+p+q)$

The process is stationary whenever the autoregressive coefficient is smaller than 1, i.e., $(-1+p+q) < 1$,or ,$p+q < 2$.

Then the expected value of x_t is given by

$$E(x_{t+1}) = (1-q) + (-1+p+q)E(x_t)$$

or

$$E(x_t) = (1-q)/(2-p-q)$$

since $E(x_{t+1}) = E(x_t)$ for a stationary process. Also notice that the expected value may be written as

$$E(x_t) = 0.\ \ P(x_t = 0) + 1P(x_t = 1) = P(x_t = 1)$$

therefore the **unconditional probabilities** of being in state 1 and zero are

$$P(x_t = 1) = (1-q)/(2-p-q)$$

and

$$P(x_t = 0) = 1 - P(x_t = 1) = 1 - (1-q)/(2-p-q) = (1-p)/(2-p-q)$$

Conditional and Unconditional Probabilities of States 0 and 1

(An alternative derivation).

Notice that to start the Markov chain we need information of the probabilities of state 0 and 1 at time zero. This is given by the unconditional probabilities $P(x_0 = 0)$ and $P(x_0 = 1)$. To get the unconditional probabilities of the states at time 1, you simply need to multiply the unconditional probabilities at time zero by the matrix of transition probabilities.

$$\begin{bmatrix} P(x_1 = 0) \\ P(x_1 = 1) \end{bmatrix} = \begin{bmatrix} q & (1-p) \\ (1-q) & p \end{bmatrix} \begin{bmatrix} P(x_0 = 0) \\ P(x_0 = 1) \end{bmatrix}$$

It is shown in Cox and Miller (1965) that if there exists a statistical equilibrium in which the state equilibrium is independent of the initial conditions, then the state probability π satisfies the condition that $\Pi = \mathbf{P}\Pi$, where $\mathbf{P}$ is the matrix of transition probabilities.

$$\begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix} = \begin{bmatrix} q & (1-p) \\ (1-q) & p \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}$$

where $\pi_0 = P(x_{t-j} = 0)$ and $\pi_1 = P(x_{t-j} = 1)$ for all values of j.

Then using $\pi_0 + \pi_1 = 1$ we get the following values:

$$\pi_0 = \frac{(1-p)}{(2-p-q)} \text{ and } \pi_1 = \frac{(1-q)}{(2-p-q)}$$

where π_0 and π_1 are the equilibrium unconditional probabilities. If there exists a pair of initial values that introduce stationarity in the stochastic process, then choice of the initial value is of great importance. Note that the initial values can be the equilibrium unconditional probability.

Given the following initial values,

$$\begin{bmatrix} p^0(0) \\ p^0(1) \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}$$

it can be shown (by multiplying n times by the transition probability matrix) that the unconditional probability vector at time n is:

$$\begin{bmatrix} p^n(0) \\ p^n(1) \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}$$

Therefore, the distribution does not change with time and the stochastic process is always in equilibrium.

Forecasts for a Markov chain

Sometimes it is also useful to know the probability of being in state 1 (0)at time $t+n$ given that state 1 (0) prevailed at time t. A clear example is Hamilton's (1988) application of filter to test the (rational) expectational hypothesis of the term structure of interest rates which requires the forecasting of the state n periods ahead given the information of the state at time t.

. A n-periods ahead forecast for a Markov chain can be obtained simply by multiplying n times by the transition probability.

$$\begin{bmatrix} P(x_{t+n} = 0) \\ P(x_{t+n} = 1) \end{bmatrix} = \begin{bmatrix} q & (1-p) \\ (1-q) & p \end{bmatrix}^n \begin{bmatrix} P(x_t = 0) \\ P(x_t = 1) \end{bmatrix}$$

Following Cox and Miller (1965), $\mathbf{P}^n$ is derived in the following way:

1) Find the eigen-values of the transition probability Matrix.

$$\lambda_1 = 1, \quad \lambda_2 = -1 + p + q,$$

2) Find the associated eigen-vectors.

$$\left[\begin{array}{c} \frac{(1-p)}{(2-p-q)} \\ \frac{(1-q)}{(2-p-q)} \end{array} \right], \left[\begin{array}{c} -1 \\ 1 \end{array} \right]$$

3) Express $\mathbf{P}$ as $T \Lambda T^{-1}$, where

$$T = \left[\begin{array}{cc} \frac{(1-p)}{(2-p-q)} & -1 \\ \frac{(1-q)}{(2-p-q)} & 1 \end{array} \right],$$

is the matrix of eigen-vectors and

$$\Lambda = \left[\begin{array}{cc} 1 & 0 \\ 0 & -1 + p + q \end{array} \right],$$

a diagonal matrix of eigen-values.

4) Use the result that

$$\mathbf{P}^n = T \Lambda^n T^{-1}$$

or

$$\mathbf{P}^n = \left[\begin{array}{cc} \frac{(1-p)}{(2-p-q)} & -1 \\ \frac{(1-q)}{(2-p-q)} & 1 \end{array} \right] \left[\begin{array}{cc} 1 & 0 \\ 0 & -1 + p + q \end{array} \right] \left[\begin{array}{cc} 1 & 1 \\ -\frac{(1-q)}{(2-p-q)} & \frac{(1-p)}{(2-p-q)} \end{array} \right]$$

	0	1	(time t)
0	$\frac{(1-p)}{(2-p-q)} + \frac{(1-q)(p+q-1)^n}{(2-p-q)}$	$\frac{(1-p)}{(2-p-q)} - \frac{(1-p)(p+q-1)^n}{(2-p-q)}$	
1	$\frac{(1-q)}{(2-p-q)} - \frac{(1-q)(p+q-1)^n}{(2-p-q)}$	$\frac{(1-q)}{(2-p-q)} + \frac{(1-p)(p+q-1)^n}{(2-p-q)}$	
(time t+n)			

Note that by making $n = 1$ in the above matrix we end with the matrix of transition probabilities, $\mathbf{P}$.

Note also that when n tends to infinity, the conditional tends to the unconditional probability. The fact that x_t has been in state 0 or 1 "infinite" number of periods ago does not provide any useful information.

In addition, we can derive the conditional expectations:

$$
\begin{aligned}
E(x_{t+n}|x_t &= 0) = \frac{(1-q)}{(2-p-q)} - \frac{(1-q)(p+q-1)^n}{(2-p-q)} \\
E(x_{t+n}|x_t &= 1) = \frac{(1-q)}{(2-p-q)} + \frac{(1-p)(p+q-1)^n}{(2-p-q)}
\end{aligned}
$$

The expected value of x_t at time n conditional on s at time zero is:

$$E(x_{t+n}|x_t = x_t) = \frac{(1-q)}{(2-p-q)} + (x_t - \frac{(1-q)}{(2-p-q)})(p+q-1)^n$$

This result is derived assuming that x_t is observed (conditional on x_t), in most of the empirical applications we will assume that the states are unobserved and must be predicted from the realizations of y_t (the observed variable). Hamilton has done this using a non-linear filter consisting of five steps that are described below. His procedure for drawing inferences about x_t is an iterative one. Given an initial inference about x_{t-1} based on data observed through date $t-1$, iteration t produces an inference about x_t based on data observed through data t.

A Brief Description of Hamilton's Non-Linear Filter.

The filter developed by Hamilton assumes that discrete states (say high or low inflation) of the economy are not known and therefore have to be inferred from the data. He assumes that the states follow a discrete Markov process. Hamilton constructed an optimal non-linear forecast, which can be thought of as arising from a two step procedure which involves obtaining an optimal inference about the current state given the past values of the variable that is to be forecast, and then using the outcome of the filter to generate future forecasts of this variable.

Hamilton's starting point is the assumption that the economy has two possible states, let us say, state zero and state one. For the sake of simplicity, he also assumes that the unconditional mean and the variance are the only parameters that are allowed to vary between regimes. It should be noted that there is no reason to assume a priori that the other parameters will remain constant from one regime to the other, and this is something that can easily be tested using traditional procedures.

The observed variable, y_t , is assumed to follow an autoregressive process of order m, allowing the mean and the variance to vary from state 0 to state 1. This can be represented in the following way.

$$y_t - \mu_{x_t} = \phi_1(y_{t-1} - \mu_{x_{t-1}}) + \phi_2(y_{t-2} - \mu_{x_{t-2}}) + .. + \phi_m(y_{t-m} - \mu_{x_{t-m}}) + \sigma_{x_t} v_t$$

v_t is distributed N(0,1) and μ_{x_t} is parameterized as $\alpha_0 + \alpha_1 x_t$ and σ_{x_t} as $w_0 + w_1 x_t$

That is the mean is equal to α_0 in state 0 (when $x_t = 0$), and equal to $\alpha_0 + \alpha_1$ in state 1 (when $x_t = 1$) and the standard deviation is equal to w_0 in state zero and equal to $w_0 + w_1$ in state one.

The error v_t is assumed to be independent of all $x_{t-j} \geq 0$

4.2.2 Hamilton's Filter.

Step_ 1. Calculate the joint density of the m past states and the current state conditional on the information included in y_{t-1} and all past values of y, where y is the variable that is observed. This is done by using the Markov property which says that only the information of the last state is relevant.[1]

$$p(x_t, x_{t-1}, ., x_{t-m}|y_{t-1}, y_{t-2}, ., y_0) = p(x_t|x_{t-1})p(x_{t-1}, x_{t-2}, ., x_{t-m}|y_{t-1}, y_{t-2}, ., y_0)$$

$p(x_t|x_{t-1})$ is given by (4.1). As in all the subsequent steps the second term on the right-hand-side is obtained from the preceding step of the filter. In this case, $p(x_{t-1}, x_{t-2}, ..., x_{t-m}|y_{t-1}, y_{t-2}, ..., y_0)$ is known from the input to the filter, which in turn represents the result of the iteration at date $t-1$ (from step 5).

To begin with the iteration, it is necessary to assign some initial values to the parameters, and to impose some initial conditions on the Markov process. For the sake of simplicity, the unconditional distribution $p(x_{m-1}, x_{m-2}, ...x_0)$ has been chosen for the first observation.

$$p(x_{m-1}, x_{m-2}, ...x_0) = p(x_{m-1}|x_{m-2})...p(x_1|x_0)p(x_0)$$

where $p(x_0)$ are the equilibrium unconditional probabilities as defined above.

Step_ 2. Calculate the joint conditional distribution of y_t and $(x_t, x_{t-1}, ..., x_{t-m})$.

$$\begin{aligned} p(y_t, x_t, x_{t-1}, ..., x_{t-m}|y_{t-1}, y_{t-2}, ..., y_0) &= p(y_t|x_t, x_{t-1}, ..., x_{t-m}, y_{t-1}, y_{t-2}, ..., y_0). \\ &\quad p(x_t, x_{t-1}, x_{t-2}, ..., x_{t-m}|y_{t-1}, y_{t-2}, ..., y_0) \end{aligned}$$

where we assume that

$$\begin{aligned} & p(y_t|x_t, x_{t-1}, ., x_{t-m}, y_{t-1}, y_{t-2}, ., y_0) \\ = \; & \frac{1}{\sqrt{2\pi}(\omega_0 + \omega_1 x_t)} \exp[-\frac{1}{2[\omega_0 + \omega_1 x_t]^2}((y_t - \alpha_1 x_t - \alpha_0) \\ & -\phi_1(y_{t-1} - \alpha_1 x_{t-1} - \alpha_0) - .. - \phi_m(y_{t-m} - \alpha_1 x_{t-m} - \alpha_0))^2] \end{aligned}$$

[1] This means that the probability of the current state conditional on the previous state is equal to the probability of the current state conditional on the m past states, that is;
$P(X_t|X_{t-1}) = (X_t|X_{t-1}, ..., X_{t-m})$

Note that $p(y_t|x_t, x_{t-1}, ., x_{t-m}, y_{t-1}, y_{t-2}, ., y_0)$ involves $(x_t, x_{t-1}, ., x_{t-m})$ which is a vector which can take 2^{m+1} values.

Step_ 3. Marginalise the previous joint density with respect to the states giving the conditional density, from which the (conditional) likelihood function is calculated.

$$p(y_t|y_{t-1}, y_{t-2}, ., y_0) = \sum_{x_t=0}^{1} \sum_{x_{t-1}=0}^{1} .. \sum_{x_{t-m}=0}^{1} p(y_t, x_t, x_{t-1}, ., x_{t-m}|y_{t-1}, y_{t-2}, ., y_0)$$

Step_ 4. Combine the results from steps 2 and 3 to calculate the joint density of the state conditional on the observed current and past realizations of y

$$p(x_t, x_{t-1}, ., x_{t-m}|y_t, y_{t-1}, y_{t-2}, ., y_0) = \frac{p(y_t, x_t, x_{t-1}, ., x_{t-m}|y_{t-1}, y_{t-2}, ., y_0)}{p(y_t|y_{t-1}, y_{t-2}, ., y_0)}$$

.

Step_ 5. The desired output is then obtained from

$$p(x_t, x_{t-1}, ., x_{t-m+1}|y_t, y_{t-1}, y_{t-2}.., y_0) = \sum_{x_{t-m}=0}^{1} p(x_t, x_{t-1}, ., x_{t-m}|y_t, y_{t-1}, y_{t-2}, ., y_0)$$

The output of step 5 is used as an input to the filter in the next iteration. Note that to iterate, estimates of the parameters are required. Maximum likelihood estimates can be obtained numerically from Step 3 as a by-product of the filter, and this is the approach followed in Hamilton (1988). The sample conditional likelihood is:

$$\ln p(y_t, y_{t-1}, y_{t-2}, ., y_m)|y_{m-1}, ., y_0) = \sum_{t=m}^{T} \ln p(y_t|y_{t-1}, ., y_0).$$

which can be maximized numerically with respect to the unknown parameters $(\alpha_1, \alpha_0, p, q, \omega_0, \omega_1, \phi_1, \phi_2 ... \phi_m)$.

Notice that p and q, the parameters of the transition matrix, are also estimated by maximum likelihood. As we said in Note 1, p and q are the parameters of the transition matrix associated with remaining in the previous state.

Hamilton's filter requires the numerical optimization of a very complicated non-linear function. Cramer (1986) has pointed out that maximum likelihood estimation suffers from several specific types of failures. First, the parameter vector may change direction at ever increasing speed toward absurd values, while still increasing log likelihood at each step. Second, the iterative process may also enter a loop and keep repeating the same movements of the parameters. Third, collinearity of the data or under identification of the model can produce a close to singular information matrix. We must bear in mind that most of these problems can occur in the examples to be analyzed as a result of the non-linear nature of Hamilton's filter.