Notes Section 8

Markov–Switching Vector Autoregressions

How might we model timeseries processes that undergo changes in regime? We have considered processes with known breakpoints, and evaluating the possible location of one or more breakpoints through Perron's models of unit roots with unknown structural break. But what if there are two (or more) regimes, and some likelihood in every period of regime change? If we have observe that a process has changed in the past, we should consider that it could change in the future when forming a forecast of the process. The change in regime should not be regarded as the outcome of a perfectly foreseeable, deterministic event, but rather as a random variable. A complete time series model would include a description of the probability law governing regime change.

The leading approach to this problem involves an unobserved, or latent, random variable s_t^*, which will be termed the state or regime of the process at date t. A simple regime–specific autoregression might be written as:

$$\begin{aligned} y_t - \mu_1 &= \phi(y_{t-1} - \mu_1) + \epsilon_t \\ y_t - \mu_2 &= \phi(y_{t-1} - \mu_2) + \epsilon_t \end{aligned}$$

where each state has its own mean, but is presumed to follow the same dynamics in both regimes (an assumption that may readily be relaxed). Using the state variable, we may rewrite both regimes as

$$y_t - \mu_{s_t^*} = \phi(y_{t-1} - \mu_{s_{t-1}^*}) + \epsilon_t$$

where the regime–dependent means are now expressed in terms of the latent variable. Note that the latent variable is discrete, taking on values 1 and 2. The simplest time series model for a discrete–valued random variable is a **Markov chain.**

Let s_t be a random variable that can assume only positive integer values $(1, 2, ..., N)$. Suppose that the probability that s_t equals a particular value j depends only upon the most recent value of s_t :

$$P\{s_t = j | s_{t-1} = i, s_{t-2} = k, ...\} = P\{s_t = j | s_{t-1} = i\} = p_{ij}$$

A process that is described as a n–state Markov chain with transition probabilities p_{ij}, giving the probability that state i will be followed by state j. It must therefore be so that $p_{i1} + p_{i2} + ... + p_{iN} = 1$. The transition probabilities may be assembled in a **transition matrix** $P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}$

If a 2x2 transition matrix $\begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix}$ is upper triangular, then state 1 is an absorbing state: once the process enters state 1, it will never leave it, since $p_{11} = 1$. The Markov chain in this case is said to be reducible. If both p_{11} and p_{22} are strictly less than unity, the chain is said to be irreducible. This generalizes to an n-state chain: if the transition matrix is block-upper triangular, with a $k \times k$ block in the upper left corner, then once the process enters one of those states, it will not leave them.

Since every column of P must sum to unity, $P'1 = 1$ where 1 is a $n \times 1$ vector of ones. This implies that P must have one unit eigenvalue, with associated eigenvector 1. If all other eigenvalues of P are inside the unit circle, the Markov chain is said to be ergodic, with a vector of ergodic probabilities π which satisfy $P\pi = \pi$. These are the limiting values of powers of the transition matrix:

$$\lim_{m \to \infty} P^m = \pi 1'$$

This is of relevance since a $m-$period-ahead forecast of the process will involve P^m. This also implies that a long-run forecast of the process will be independent of the current state. The ergodic probabilities can be viewed as indicating the unconditional probability of being in each of the N states. An ergodic Markov chain is a covariance-stationary process, yet it may be expressed as a VAR with a unit root, since one of the eigenvalues is unity. The VAR is stationary, despite the unit root, since the variance-covariance matrix of its error process is singular. How may we calculate the unconditional probabilities in π? The vector π has the properties that $P\pi = \pi$ and $1'\pi = 1$. Thus we seek a vector π satisfying $A\pi = e_{N+1}$ where e_{N+1} denotes the $(N+1)^{st}$ column of I_{N+1}, and

$$A_{(N+1)\times N} = \begin{bmatrix} I_N - P \\ 1' \end{bmatrix},$$

a solution for which may be found by premultiplying by

$(A'A)^{-1}A'$:

$$\pi = (A'A)^{-1}A'e_{N+1}$$

That is, π is the $(N+1)^{st}$ column of the matrix $(A'A)^{-1}A'$.

i.i.d. Mixture Distributions

We will consider autoregressive models in which the parameters of the autoregression can change as the result of a regime–shift variable, where the regime is the outcome of an unobserved Markov chain. Let the regime at date t be indexed by the latent variable s_t which takes on one of N possible values. When the process is in regime 1, the observed variable y_t is presumed to have been drawn from a $N(\mu_1, \sigma_1^2)$ distribution; when in regime 2, from a $N(\mu_2, \sigma_2^2)$ distribution, and so on. The vector of population parameters θ includes the N values of μ and the N values of σ^2, .as well as a set of unconditional probabilities. The regime s_t is presumed to have been generated from a probability distribution, with unconditional probabilities given by

$$P\left\{s_t = j; \theta\right\} = \pi_j$$

for each of the N values of j. We may then form the joint density–distribution function of y_t and s_t, which describes the density of y_t, given a particular value for s_t. The unconditional density for y_t may then be calculated as:

$$f\left(y_t; \theta\right) = \sum_{j=1}^{N} p\left(y_t, s_t = j; \theta\right)$$

where $p()$ is the density–distribution function. If the regime variable is distributed $i.i.d.$ across time, the log likelihood for the observed data y_t may be calculated as:

$$\pounds\left(\theta\right)=\sum_{t=1}^{T}\log f\left(y_t;\theta\right)$$

which must be maximized subject to the constraints that $\sum \pi = 1$ and $\pi_j \geq 0$. The maximum likelihood estimates of θ represent the solution to a system of nonlinear equations, which are often solved by application of the EM (expectations maximization) algorithm.

Once one has obtained estimates of θ , it is possible to make an inference about the regime likely to have been in force in each period of the observed sample: more specifically, the probability that the data for time t were generated by regime j.

Time series models of changes in regime

We consider a model that allows a variable to follow a different time series process over different subsamples. Consider an $AR(1)$ where both the constant term and the autoregressive coefficient are indexed by the regime:

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

with $\epsilon\tilde{}N(0,\sigma^2)$. This problem generalizes to processes where the probability that $s_t = j$ depends upon not only its prior period's value (in a strict Markov chain) but also upon a vector of other observed variables (the time–varying transition probability (TVTP) model). A regime–switching model can include a set of observed exogenous regressors; in general, we consider y_t as an $n-$vector of observed endogenous variables, and x_t a $k-$vector of observed exogenous variables. In the example given above, n

is 1, and x includes only a units vector. The unknown parameters in this problem then include $(c_1, ...c_N)$, $(\phi_1, ..., \phi_N)$, and σ^2, for the N regimes. If N equals 2, there are then 5 parameters to be estimated. A generalization of the model would allow the error variance to vary across regimes as well. An important issue with these models is that any inference must be made conditional on the assumption of a specific number of regimes; one cannot readily test for the number of regimes with a likelihood ratio test, since under the null hypothesis of $N-1$ states, the parameters defining the N^{th} state are unidentified. Alternatively, one may fit the $N-1$ state model and examine its adequacy to determine whether a $N-$state model is needed.

This strategy may be generalized to the modelling of the joint evolution of a set of variables, in terms of a Markov-switching VAR: that is, a VAR model in which each of the equatioons is subject to regime shifts.

A number of numerical examples of this modelling strategy are available in Krolzig's $MSVAR$ package for Ox 3.0. He also describes how the MS-VAR model can be related to several other models proposed for the analysis of multiple regimes: for example, the self–exciting threshold autoregressive ($SETAR$) model and the smooth transition autoregressive ($STAR$) model. All three are special cases of what Krolzig calls an endogenous selection Markov–switching VAR model, in which the transition probabilities are not constant, but rather functions of the observed time series vector y_{t-d}.

References

Hamilton, J.D., 1990. Analysis of time series subject to changes in regime. Journal of Econometrics, 45, 39-70.

Hamilton, J.D., 1994. Time Series Analysis. Princeton: Princeton University Press.

Krolzig, H.-M., 1998. Econometric modelling of Markov-Switching vector autoregressions using MSVAR for Ox. From the Ox website.