Nonlinear Time Series Models 2: Markov Switching Models, TVP models and Neural Nets

Advanced Time Series Econometrics

Scottish Graduate Programme in Economics

Markov Switching Models

- Popular class of regime-switching model
- Similar idea to state space model
- Latent variable (similar to state) denotes which regime you are in
- Automatically classifies observations into regimes
- Illustrate for AR(1) model with two regimes
- Ideas go through for AR(p) (or time series regression or VARs) with many regimes

Markov Switching Models

- $$y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_{1t} \text{ if } s_t = 1 \\ \rho_2 y_{t-1} + \varepsilon_{2t} \text{ if } s_t = 2 \end{cases}$$
- Different AR models depending on whether $s_t = 1$ or 2
- Similar structure to TAR
- Would be a TAR if we were to define "regime indicator" or "state" s_t as
- $s_t = 1$ if $z_t \leq \tau$
- $s_t = 2$ if $z_t > \tau$
- But Markov switching model defines s_t differently

Markov Switching Models

- s_t is hidden two-state Markov chain
- What does this mean?
- "Hidden" means not directly observed (latent) similar to states
- Wikipedia's definition of Markov Chain:
- 'random process that undergoes transitions from one state to another on a state space. It must possess a property that is usually characterized as memorylessness: the probability distribution of the next state depends only on the current state and not on the sequence of events that preceded it.'

Markov Switching Models

- E.g. Suppose $s_t = 1$,what is s_{t+1}?
- Can either stay in regime 1, or switch (transition) to regime 2

$$\begin{aligned}\Pr\left(s_{t+1} = 1 | s_t = 1\right) &= p_{11} \\ \Pr\left(s_{t+1} = 2 | s_t = 1\right) &= p_{12}\end{aligned}$$

- Probability of switching depends only on s_t (not on s_{t-1} or any past data, etc. = memoryless)
- E.g. suppose $s_t = 1$ is recession, 2 is expansion
- p_{11} is probability of staying in recession next period
- p_{12} is probability of switching from regime 1 to 2
- I.e. out of recession into expansion (turning point of business cycle)
- $p_{12} = 1 - p_{11}$

Markov Switching Models

- Similarly can define a p_{22} as probability of staying in regime 2 given currently in regime 2
- $$\begin{aligned} \Pr\left(s_{t+1} = 1 | s_t = 2\right) &= p_{21} \\ \Pr\left(s_{t+1} = 2 | s_t = 2\right) &= p_{22} \end{aligned}$$
- $p_{21} = 1 - p_{22}$ is probability of switching from regime 2 to 1
- e.g. switching from expansion to recession
- Used for dating turning points in business cycles, calculating probability economy will go into recession, etc.

Markov Switching Models

- Definition:

$$y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_{1t} \text{ if } s_t = 1 \\ \rho_2 y_{t-1} + \varepsilon_{2t} \text{ if } s_t = 2 \end{cases}$$

- $$\Pr\left(s_{t+1} = j | s_t = i\right) = p_{ij}$$

- $var\left(\varepsilon_{jt}\right) = \sigma_j^2$
- Can have $\sigma_1^2 \neq \sigma_2^2$ (regimes have different error variances)
- Or can have $\sigma_1^2 = \sigma_2^2$ (homoskedastic: regimes have same error variances)

Markov Switching Models

- Can estimate Markov switching model, e.g., in Stata using maximum likelihood
- Estimates of p_{ij} and s_t for $t = 1, .., T$ provided
- Can use information criteria to choose between Markov switching and other models
- s_t tells you which observation is in which regime

Markov Switching Models

- Important difference with TAR
- TAR you (the researcher) chooses the regimes
- E.g. GDP growth and choice of $z_t = y_{t-1}$ and threshold $\tau = 0$ implies:
- One regime where last period's growth was negative, another positive
- The researcher has imposed: regime 1 is recessionary, regime 2 is expansionary
- Markov switching *estimates* which observation lies in which regime
- Maybe the classification could accord with expansion/recession, but maybe not
- Could get any division into regimes

Example: Markov Switching in US GDP growth

- Markov switching AR(1) model (homoskedastic)
- Include an intercept in each model (called α_j in tables below)
- Stata will produce:
- Estimates of intercept and AR coeff. for each regimes
- Error variance
- transition probabilities
- See table on next slide

	Estimate	St. Error	95% Confidence Interval	
α_1	1.805	0.280	1.257	2.353
ρ_1	0.379	0.056	0.269	0.488
α_2	14.816	2.365	10.180	19.451
ρ_2	-0.091	0.223	-0.528	0.346
p_{11}	0.987	0.009	0.948	0.997
p_{12}	0.013	0.009	0.003	0.052
p_{21}	0.641	0.250	0.175	0.938
p_{22}	0.359	0.250	0.062	0.825

Example: Markov Switching in US GDP growth

- How are results interpreted?
- p_{11} is close to one: if in regime 1, stay there with high probability
- p_{22} is much smaller (0.359): if in regime 2, tend to switch to regime 1
- Stata (command: estat durations) will estimate durations of each regime
- Estimated duration of regime 1 is 75.4 quarters
- Estimated duration of regime 2 is 1.6 quarters
- Regime 1 long, regime 2 very short
- Regime 2 has few observations in it so imprecise estimation (wide confidence intervals)
- Next slide has estimates of probability each period is in regime 1 (filtered)
- Probabilities for regime 2 are one minus this

State1, filter probabilities
0
.2
.4
.6
.8
1
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
2020q1
quarters

Example: Markov Switching in US GDP growth

- Figure shows regime 1 holds almost all the time
- Regime 2 only holds for a few periods
- If you look at data (remember we are using % change since previous quarter at annual rate):
- 7 quarters have very fast ($>$11%) growth rates: 1950Q1, Q2, Q3, 1955Q1, 1971Q1, 1978Q2
- These are exactly the ones classified as regime 2
- Regime 2 $=$ outliers of unusually fast growth

Example: Markov Switching in US GDP growth

- Note: fitted AR(1) model in regime 2 is:

$$y_t = 14.82 - 0.091 y_{t-1}$$

- AR coefficient is insignificant, hence regime 2 is (approx.) saying $y_t = 14.82$ (i.e. predicted GDP growth in regime 2 is very high)
- This (homoskedastic) Markov switching model has been asked to divide data in two regimes
- Answer: regime 1 is "normal growth", regime 2 is "small number of outliers"

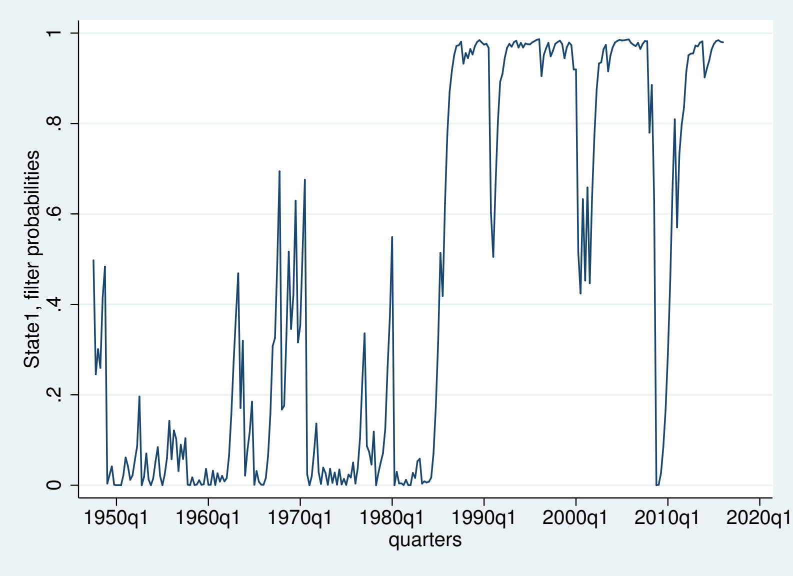
Example: Markov Switching in US GDP growth

- Repeat the analysis with heteroskedastic model
- Now $\sigma_1^2 \neq \sigma_2^2$
- All other specification choices the same
- Next table gives parameter estimates

	Estimate	St. Error	95% Confidence Interval	
α_1	2.205	0.340	1.538	2.857
ρ_1	0.247	0.100	0.051	0.442
α_2	2.152	0.453	1.264	3.040
ρ_2	0.387	0.078	0.239	0.534
p_{11}	0.983	0.014	0.920	0.997
p_{12}	0.017	0.014	0.003	0.080
p_{21}	0.014	0.012	0.003	0.068
p_{22}	0.986	0.012	0.932	0.997
σ_1^2	1.949	0.137	1.698	2.238
σ_2^2	4.545	0.271	4.043	5.109

Example: Markov Switching in US GDP growth

- How are results interpreted?
- p_{11} and p_{22} now both much closer to one
- Once in a regime, tend to stay there for long time
- Estimated duration of regime 1 is 59.9 quarters
- Estimated duration of regime 2 is 70.9 quarters
- Fitted regression lines in two regimes similar to one another
- Regimes differ in error variances: $\widehat{\sigma}_1^2 = 1.949$ and $\widehat{\sigma}_2^2 = 4.043$
- Next slide has estimates of probability each period is in regime 1 (filtered)
- Probabilities for regime 2 are one minus this



State1, filter probabilities
0
.2
.4
.6
.8
1
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
2020q1
quarters

Example: Markov Switching in US GDP growth

- Figure shows regime 1 holds for most of time since 1983 (with some exceptions)
- Post-1983 often called Great Moderation of the Business Cycle
- Exceptions around 2008-2009 (Financial Crisis) and 2001 (bursting of dotcom bubble)
- Regime 2 holds for these exceptions plus most of earlier part of sample
- This (heteroskedastic) Markov switching model has been asked to divide data in two regimes
- Its answer: regime 1 is “low volatility”, regime 2 is “high volatility”

Example: Markov Switching in US GDP growth

- Model comparison between hetero and homo versions:
- AIC for homo is 1484.6 and for hetero is 1446.8
- BIC for homo is 1509.9 and for hetero is 1475.8
- Heteroskedastic version of model clearly preferred
- Comparing to TAR and AR, heteroskedastic Markov switching is best
- Of all the models considered in this lecture for GDP growth best one is:
- Markov switching model involving high and low volatility regimes

Time-varying Parameter AR Model

- Markov switching: abrupt switches between regimes
- TVP-AR allows for constant gradual evolution of parameters
- Illustrate with TVP-AR(1)
- Same ideas work with TVP-AR(p), TVP regression or TVP-VARs
- They are state space models
- All our state space tools (Kalman filter, etc.) can be used

Time-varying Parameter AR Model

- Measurement equation

$$y_t = \rho_t y_{t-1} + \varepsilon_t$$

- Note t subscript on AR coefficient
- State equation

$$\rho_{t+1} = \rho_t + u_t$$

The TVP-AR

- Remember our general Normal Linear State Space model
- Measurement equation:

$$y_t = W_t \delta + Z_t \beta_t + \varepsilon_t$$

- State equation:

$$\beta_{t+1} = D_t \beta_t + u_t$$

- TVP-AR is special case of this with:
- $W_t = 0$ (although can add explanatory variables with constant coeffs through W_t)
- $Z_t = y_{t-1}$
- $\beta_t = \rho_t$
- $D_t = 1$

Example: Estimating the TVP-AR using GDP Growth Data

- Will not discuss econometric estimation and model comparison for TVP-AR
- Already covered in state space model lecture
- Stata cannot estimate the TVP-AR (without further programming)
- It only allows for Normal linear state space model where $Z_t = Z$ (constant over time)
- With TVP-AR $Z_t = y_{t-1}$ (varying over time)
- Allowing for time-varying intercept in Stata possible
- That is, estimating

$$y_t = \alpha_t + \beta x_t + \varepsilon_t$$

  can be done in Stata (can have regression terms playing role like $W_t \delta$ in Normal linear state space model)
- But not

$$y_t = \alpha_t + \beta_t x_t + \varepsilon_t$$

  using Stata's state space commands

Example: Time Varying CAPM

- Tsay textbook has example with CAPM with time-varying α and β
- CAPM = capital asset pricing model
- Key model in finance (we will consider similar models in next lecture on Factor Models)
- Here outline Tsay's Time varying version of CAPM
- r_t = excess return on asset of interest (e.g. General Motors stock)
- $r_{M,t}$ = excess return on the market as a whole (e.g. S&P500 index)

Example: Time Varying CAPM

- Conventional (constant parameter) CAPM

$$r_t = \alpha + br_{M,t} + \varepsilon_t$$

- b is CAPM beta
- Idea: b is measure of risk, measure of volatillity of GM stock relative to market as a whole
- $b < 1$ GM is less volatile than stock market as a whole
- $b > 1$ GM is more volatile than stock market as a whole
- α is CAPM alpha
- Idea: α is expect to be zero, if positive it measures abnormal returns (on risk adjusted basis) investor gets from hold GM stock
- Excess return on portfolio/mutual fund above what an equilibrium model like CAPM might suggest

Example: Time Varying CAPM

- Tsay suggests CAPM alpha and/or beta might be changing over time
- E.g. There are some times mutual fund manager can enjoy abnormal returns ($\alpha > 0$) other times not ($\alpha \approx 0$)
- E.g. Financial crisis caused correlation between stock market as a whole and individual stocks to change (β changes)
- Time-varying CAPM:

$$
\begin{aligned}
r_t &= \alpha_t + b_t r_{M,t} + \varepsilon_t \\
\alpha_{t+1} &= \alpha_t + u_t^a \\
b_{t+1} &= b_t + u_t^\beta
\end{aligned}
$$

Example: Time Varying CAPM

- But this is a Normal linear State Space Model with:
- $W_t = 0$
- $Z_t = [1, r_{M,t}]$
- $\beta_t = \begin{pmatrix} \alpha_t \\ b_t \end{pmatrix}$
- $D_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Artificial Neural Networks

- Artificial Neural Network (ANN) very popular in machine learning
- Less popular in time series econometrics, but growing in popularity
- Hence, I provide brief introduction to them
- Cannot easily estimate ANN's in Stata
- ANNs provide valid approximation a huge class of nonlinear functions
- So if you think nonlinearity is likely, but do not know functional form, ANN's can be useful
- Black box method (can be hard to interpret results)
- For forecasting this may not be a problem, but for structural economic analysis can be a problem

Artificial Neural Networks

- An example of a neural network involving dependent variable y_t
- $$y_t = \alpha\xi_{t-h} + \sum_{i=1}^{n} \gamma_i G(\xi_{t-h}\beta_i) + \varepsilon_t$$
- ξ_t is some data.
- E.g. $\xi_t = (1, y_t, y_{t-1}, .., y_{t-p+1})$ is you want p lags of the depedendent variable to be used as predictors
- h is forecast horizon.
- E.g. typical to set $h = 1$ but can set $h > 1$ for longer forecast horizons
- $G(.)$ is the logistic function (other forms possible):

$$G(x) = \frac{1}{1 + exp(x)}$$

Artificial Neural Networks

- General idea: y_t is nonlinear function of p lags of $\xi_t = (1, y_t, y_{t-1}, .., y_{t-p+1})$
- Why this particular function?
- ANN theory says highly flexible, capable of approximating virtually anything
- So if you do not know function form, can try ANN
- ANN terminology:
- There are n "hidden units" in the ANN
- G is "activation function"
- ξ_t are "inputs" that enter the activation function
- β_i are "connection strengths"
- γ_i are "weights" that determine the "output layer" (y_t)

Artificial Neural Networks

- This is a "univariate single layer feed-forward neural network"
- "Univariate" since one dependent variable
- "Single layer" since can have more layers (see next slide)
- "feed forward" since past information (ξ_{t-h}) on right hand side "feeds forward" in time to predict the left hand side variable
- Neural networks (and associated terminology) inspired by how learning happens in the brain

Artificial Neural Networks

- Univariate single layer feed-forward neural network very flexible
- But can be made even more flexible by having multiple hidden layers
- Double layer feed-forward neural betwork:

$$y_t = \alpha\xi_{t-h} + \sum_{i=1}^{n_1} \gamma_i G(\sum_{j=1}^{n_2} \lambda_{j,i} G(\xi_{t-h}\beta_i)) + \varepsilon_t$$

Econometric of ANNs

- Use nonlinear least squares methods to estimate unknown parameters in α, β_i and γ_i for $i = 1, .., n$
- Use information criteria to choose n
- Evaluate information for $n = 1$, then $n = 2$, etc. and choose best one
- If you have different possible choices for $G(.)$ can also use information criteria
- To decide whether single versus double layer can use information criteria, etc.

Interpreting Results from ANNs

- When forecasting interpration of parameters is unimportant (all you care is whether forecasts are good or not)
- Econometric techniques will provide estimates of β_i and γ_i for $i = 1, .., n$
- Hard for the economist to directly interpret their meaning
- If ξ_t contains only a few elements can plot fitted regression line
- E.g. $h = 1$ and $\xi_{t-1} = (1, y_{t-1})$ then can plot

$$y_t = \alpha\xi_{t-h} + \sum_{i=1}^{n} \gamma_i G(\xi_{t-h}\beta_i) + \varepsilon_t$$

for various values of y_{t-1}

Interpreting Results from ANNs

- Put y_{t-1} on X-axis and fitted value of y_t on Y-axis
- If an AR(1) model is appropriate then such a plot should be straight line
- With ANN such a plot could be nonlinear
- Manner in which ANN deviates from linearity could be informative to the economist
- E.g. Does ANN plot reveal different value above/below a threshold? This suggests TAR behaviour
- etc.

Summary

- This lecture covers range of nonlinear time series models, including:
- Markov switching models: parameters switching abruptly between different regimes
- TVP-AR: Gradual change over time
- Neural nets: Very flexible approach suitable when nonlinear form is unknown