Nonlinear Time Series Models

Introduction

- Regressions, ARMA models and VARs are all linear
- Illustrate with a regression:

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

- Straight line relationship between y and x
- β is marginal effect of x on y
- Marginal effect is same regardless of value of x
- Nonlinear models = anything that is not linear
- Marginal effect may be different depending on what value of x is
- E.g. quadratic relationship

$$y_t = \alpha + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$$

- Marginal effect is $\beta_1 + 2\beta_2 x_t$

Types of Nonlinearity

- In macroeconomics and finance certain types of nonlinearities of particular interest
- Abrupt change: at a point in time β abruptly changes
- Structural break models
- E.g. financial crisis fundamentally altered the financial markets and relationships between variables changed
- Gradual change: β is gradually changing over time
- Time varying parameter (TVP) models
- E.g. financial liberalizations throughout the 1980s and 1990s gradually changed the relationships between financial variables

Types of Nonlinearity

- Regime switching models: β is different in different regimes
- E.g. Regime $=$ state of the business cycle
- Expansion and Recession might be two regimes
- E.g. β measures the effectiveness of monetary policy
- Measures how much inflation increases when money supply increased
- Printing money in recessionary times might not cause inflation ($\beta = 0$)
- Printing money when economy is booming will (β large)

Types of Nonlinearity

- So far talked about nonlinearities in β
- Nonlinearities in error variance (σ^2)
- E.g. high volatility regime versus low volatility regime in stock markets
- E.g. Risk on/Risk off behaviour of financial market participants
- E.g. Great Moderation of Business Cycle
- In early 1980s volatility of many macro variables dropped

Why Worry About Nonlinearities?

- Strong empirical evidence it exists in many (most?) macroeconomic and financial time series
- Macro variables: Stock and Watson (1996) "Evidence on Structural Instability in Macroeconomic Time Series Relations" *Journal of Business and Economic Statistics*
- Financial Variables: Ang and Bekaert (2002) "Regime Switches in Interest Rates" *Journal of Business and Economic Statistics*
- Many other papers present similar findings
- If nonlinearities exist linear models are mis-specified
- Linear researcher gives wrong policy advice/bad forecasts, etc.
- E.g. How much does inflation increase when money supply increase?
- Linear methods give average estimate over expansions and recessions
- Average over time where $\beta = 0$ and β large might give β fairly large

The Econometrics of Nonlinear Time Series Models

- This lecture goes through various models of breaks/regime switching/TVP
- Focus is on the models, their properties and how to use them in practice
- Only a little about theory of estimation and hypothesis testing
- Maximum likelihood or Bayesian methods or least squares typically done

The Econometrics of Nonlinear Time Series Models

- Stata will produce usual likelihood based model choice methods including:
- Information criteria, tests of significance of individual parameters and likelihood ratio tests
- Reading 1: Chapters 9 and 10 of Ghysels and Marcellino
- Reading 2: Chapter 4 of Tsay

Models with Structural Breaks

- Use AR(1) model to illustrate ideas
- Extensions to AR(p) easy (just add more lags)
- And usually want to include an intercept (results below do)
- Case of structural breaks in regression coefficients is easy (just replace y_{t-1} by x_t)
- A simple structural break model:

$$y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_t \text{ if } t \leq \tau \\ \rho_2 y_{t-1} + \varepsilon_t \text{ if } t > \tau \end{cases}$$

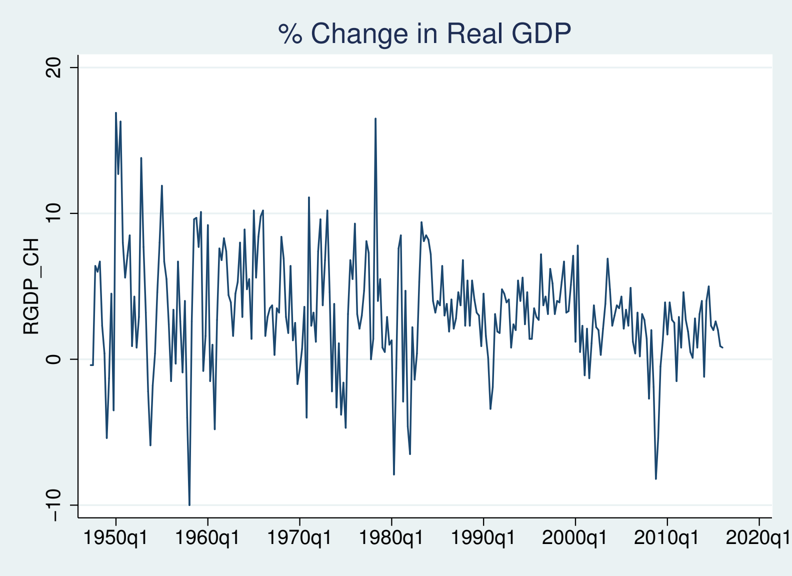
- Different AR(1) models before or after *break date* τ
- For now assume break date known
- Note: extension to multiple breaks is straightforward

Example: Breaks in US GDP growth

- Data: Real Gross Domestic Product, Percent Change from Preceding Period, Quarterly, Seasonally Adjusted Annual Rate
- Plotted on next graph
- Can write structural break model as

$$y_t = \rho_1 y_{t-1} + \gamma x_{t-1} + \varepsilon_t$$

  where $x_t = D_t y_t$
- Dummy variable: $D_t = 1$ if $t > \tau$, else $D_t = 0$
- If $t \leq \tau$ then coefficient on y_{t-1} is ρ_1
- If $t > \tau$ coefficient on y_{t-1} is $\rho_1 + \gamma$ (same as ρ_2 on previous slide)


% Change in Real GDP
RGDP_CH
20
10
0
−10
1950q1
1960q1
1970q1
1980q1
1990q1
2000q1
2010q1
2020q1

Example: Breaks in US GDP growth

- OLS estimates of the AR(1) model give fitted regression line

$$y_t = 2.04 + 0.37y_{t-1}$$

- Adding x_{t-1} using a break date $\tau = 1983Q1$ (beginning of Great Moderation of the Business Cycle?) gives

$$y_t = 2.06 + 0.38y_{t-1} - .02x_{t-1}$$

- Two fitted regressions look pretty similar, so maybe no break in $1983Q1$?
- Can test this: if $\gamma = 0$, same AR model before/after break
- T-stat for $H_0 : \gamma = 0$ is -0.20 with p-value 0.84
- Accept H_0 there is no break in $1983Q1$
- Another way to check is information criteria
- AR(1) gives AIC and BIC of 1495 and 1502
- AR(1) with break gives AIC and BIC of 1496 and 1507
- Choose AR(1) over model with break

Example: Breaks in US GDP growth

- Preceding assumed same error variance before and after break
- What if model is:

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

- $var\left(\varepsilon_{1t}\right) = \sigma_1^2$ and $var\left(\varepsilon_{2t}\right) = \sigma_2^2$
- Can estimate this model by simply dividing the data into two parts
- Part 1: all observations with $t \leq \tau$
- Part 2: all observations with $t > \tau$

Example: Breaks in US GDP growth

- Divide the data into pre- and post-1983 samples
- Two OLS regressions yield:
- Fitted regression line with pre-1983 sample:

$$y_t = 2.41 + 0.34y_{t-1}$$

- $s^2 = 20.85$
- Post-1983:

$$y_t = 1.46 + 0.49y_{t-1}$$

- $s^2 = 5.14$
- Looks like break in error variance:
- Great Moderation of the Business Cycle: s^2 much lower after 1983
- But how confident statistically?

Example: Breaks in US GDP growth

- Three models: AR(1), AR(1) with break in mean (i.e. ρ changes) and AR(1) with break in both mean and error variance
- Use information criteria (IC) to choose between them
- Note: IC for model with break in mean and variance just add together ICs from pre- and post-break regressions
-

	AIC	BIC
AR(1)	1494.74	1501.98
AR(1) with break in mean	1496.70	1507.55
AR(1) using pre-break data	842.18	848.10
AR(1) using post-break data	592.56	598.32
AR(1) with break in mean and variance	1434.74	1446.43

- AR(1) with breaks in mean and variance has lowest IC's – it is best model

Chow Test

- General test used with regression or AR (here use for break)
- Do two parts of your sample have same regression line?
- $H_0 : \rho_1 = \rho_2, \sigma_1^2 = \sigma_2^2$
- Steps in Chow Test using break date τ
- Estimate AR(1) model using entire sample and get Sum of Squared Residuals (SSR_0)
- Estimate AR(1) model using $t \leq \tau$ sample and get SSR_1
- Estimate AR(1) model using $t > \tau$ sample and get SSR_2

Chow Test

- Chow test statistic is:

$$Chow = \frac{\frac{SSR_0 - (SSR_1 + SSR_2)}{k}}{\frac{SSR_1 + SSR_2}{T_1 + T_2 - 2k}}$$

- T_1 = number of observations in $t \leq \tau$ sample
- T_2 = number of observations in $t > \tau$ sample
- k = number of explanatory variables (plus intercept) in model
- In AR(1) $k = 2$
- If H_0 is true $Chow \sim F\left(k, T_1 + T_2 - 2k\right)$
- Get critical value from F statistical tables

Example: Breaks in US GDP growth

- In our example $k = 2,\ T_1 = 132,\ T_2 = 142$
- $SSR_0 = 3640.1, SSR_1 = 2940.6, SSR_2 = 667.6$
- Plugging these in we get $Chow = 1.19$
- 5% critical value from $F\left(2, 270\right)$ is 3.09
- Test statistic less than critical value so accept H_0
- In contrast to IC results, this indicates no break
- Classical hypothesis testing: only reject H_0 if overwhelming evidence against it
- BIC is (approximately) proportional to log of probability model generated data
- ratio of exp of BICs for two models = relatively probabilities of each model
- Hypothesis tests/ICs have different interpretations so sometimes they give different results

The Threshold Autoregressive (TAR) Model

- TAR models have same form as structural break models
- But different variable triggers regime change

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

- $var\left(\varepsilon_{1t}\right) = \sigma_1^2$ and $var\left(\varepsilon_{2t}\right) = \sigma_2^2$
- τ is called the threshold

The TAR Model

- z_t (the trigger) can be any exogenous or lagged dependent variable
- Common choice: $z_t = y_{t-d}$
- d =delay parameter
- Must have $d > 0$
- Having y_t being dependent variable and also trigger runs into endogeneity problems
- z_t could be some other variable (e.g. change in oil price)
- Structural break model is special case of TAR with $z_t = t$

The TAR Model

- Assume τ (and d) known (talk about estimating them later)
- Econometrics exactly the same as structural break model (with known break data) covered previously
- E.g. GDP growth and $z_t = y_{t-1}$ and $\tau = 0$
- Two regimes: recession ($y_{t-1} \leq 0$) and expansion ($y_{t-1} > 0$)
- Approach 1: TAR with change in mean
- Create $D_t =$ a dummy variable $= 1$ for recessions, 0 for expansions
- Create $x_t = D_t y_t$
- Add x_{t-1} as explanatory variable to AR(1) model
- Approach 2: TAR with change in mean and variance
- Run separate AR(1) models for recession and expansion

Example: Regime Switches in US GDP growth

- GDP growth data
- $z_t = y_{t-1}$ and $\tau = 0$
- Table of information criteria in same format as for structural breaks
-

	AIC	BIC
AR(1)	1494.74	1501.98
TAR(1) with change in mean	1496.54	1507.39
AR(1) using $y_{t-1} \geq 0$	1241.28	1248.19
AR(1) using $y_{t-1} < 0$	247.96	251.38
TAR(1) with change in mean and variance	1489.24	1499.57

- Similar to structural break results

Example: Regime Switches in US GDP growth

- Little evidence that recessionary and expansionary regimes have different AR coefficients
- But there is evidence that error variances differ between recessions and expansions
- Larger error variances in recessions
- In substantive application, would consider longer lag length, different thresholds, triggers and delays
- Or even multiple regimes
- E.g. 4 regimes: 1) recessions, 2) recovery from recessions 3) normal times and 4) over-heating
- For financial applications, might want structural break or regime switching in volatility
- E.g. Tsay, page 182-183 has threshold GARCH model
- Asymmetric responses to positive and negative return to an asset.

What if Break Dates or Thresholds Unknown?

- Have now gone through basic ideas of structural break and TAR class of regime switching models
- But always assumed τ (and z_t and d) known
- The idea underlying estimation is simple:
- Estimate model for every possible choice for τ (and z_t and d if relevant)
- Choose value of τ with highest value for likelihood (MLE)
- Choose value for τ with lowest sum of squared residuals (least squares estimator)
- May want to restrict possible values for τ to make sure regimes contain a minimum number of observations
- E.g. break does not occur in first or last 5% of the observations

What if Break Dates or Thresholds Unknown?

- You have to program in Stata to do estimation in this way using loops
- Loops are commands of the form:
- "for each [value of τ] in [a grid of possible values] [estimate the TAR]"
- And you have to specify things in [...]
- I won't ask you to estimate τ in computer labs

What if Break Dates or Thresholds Unknown?

- With multiple breaks and other parameters can become computationally demanding
- E.g. Monthly data for 40 years ($T = 480$)
- 3 regimes (2 thresholds) TAR model (τ_1, τ_2) with 12 values for delay ($d = 1, .., 12$)
- On the order of $12 \times 480^2 = 2764800$ TARs to estimate
- Can reduce this somewhat by imposing restrictions (e.g. $\tau_2 > \tau_1$ and each regime must contain at least 40 observations)
- But still a lot of TARs to estimate

Testing for Nonlinearity

- Can always use information criteria to choose between different nonlinear time series models
- Or choose between nonlinear and linear
- What about formal hypothesis testing methods?
- If τ is known can use Chow test
- But what is τ is unknown?
- Tests break into two groups:
- 1. Test in a particular model (e.g. testing TAR/structural break versus AR)
- 2. general test for departures from linearity

Testing for Thresholds/Structural Breaks

- Same issues hold for TAR and structural break model, illustrate with former
- Bottom line: econometric theory hard and hard to do these tests in Stata (need to program)
- Consider hypothesis test for whether TAR is preferred to AR
- $$y_t = \begin{cases} \rho_1 y_{t-1} + \varepsilon_{1t} \text{ if } z_t \leq \tau \\ \rho_2 y_{t-1} + \varepsilon_{2t} \text{ if } z_t > \tau \end{cases}$$
- $H_0 : \rho_1 = \rho_2, \sigma_1^2 = \sigma_2^2$
- $H_1 : \rho_1 \neq \rho_2, \sigma_1^2 \neq \sigma_2^2$

Testing for Thresholds/Structural Breaks

- But what about τ? H_0 says nothing about it as it does not appear in AR model
- Jargon: Nuisance parameter unidentified under the null or Davies' Problem
- Complicates the econometric theory of deriving asymptotic distribution of test statistic
- Also have to program in Stata
- Beyond scope of this course, but see page 211 of Tsay textbook if interested
- Davies' problem make several nonlinear tests complicated
- Tsay, page 212 describes another TAR test which does not run into Davies' problem (but complicated in other ways)

Testing for Thresholds/Structural Breaks

- A simple test statistic for TAR can be obtained using the sums of squared residuals for the linear (SSR_0) and TAR (SSR_1) relative to estimated error variance from TAR (s^2)
- But SSR_1 and s^2 depend on choice for threshold so write them as $SSR_1(\tau)$ and $s^2(\tau)$
- Hansen test statistic is based on

$$H(\tau) = \frac{SSR_0 - SSR_1(\tau)}{s^2(\tau)}$$

- Test statistic is $H(\tau^*)$ where τ^* is chosen to make the test statistic as large as possible

Testing for Thresholds/Structural Breaks

- To do hypothesis testing you need test statistic and critical value
- We have a test statistic $H(\tau^*)$, what is critical value?
- To get critical value you need to know distribution of test statistic
- Distribution of $H(\tau^*)$ is not know analytically
- Need to use numerical methods
- This is not done in standard software like Stata

Tests for Departures from Linearity

- Several tests exist to check whether linear model is adequate without specifying alternative model
- I will discuss one of these which can be easily done in Stata
- Intuition: if linear model is okay, then its errors should be white noise
- Estimate linear model, calculate residuals and test if they are white noise
- If not, then better consider a nonlinear model such as structural break or TAR
- Note: sometimes looking at patterns in residuals can suggest what form of nonlinearity.
- E.g. if large residuals bunch at one point in time might indicate break

RESET Test

- Can be used with regression or AR(p) models, illustrate with AR(1):

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

- Obtain OLS residuals $\widehat{\varepsilon}_t$ and their sum of squares, SSR_0
- If linear model is correct, these residuals should be white noise (unpredictable)
- Run a second OLS regression of:

$$\widehat{\varepsilon}_t = \phi_1 \widehat{\varepsilon}_{t-1} + \phi_2 \widehat{\varepsilon}^2_{t-1} + v_t$$

  and get sum of squared residuals, SSR_1
- Note: can have higher powers (e.g. $\widehat{\varepsilon}^3_{t-1}$) on right hand side of second regression

RESET Test

- $H_0 : \phi_1 = \phi_2 = 0$
- If H_0 true then $SSR_1 = SSR_0$
- If H_1 true then $SSR_1 < SSR_0$
- This intuition forms basis of RESET test
- Test statistic

$$RS = \frac{(SSR_0 - SSR_1)/g}{SSR_1/(T - p - g)}$$

- Note: p is AR lag length and and g is highest power on $\widehat{\varepsilon}_{t-1}$ in second regression minus one
- If H_0 is true, $RS \sim F(g, t - p - g)$
- Get critical values from F-statistical tables

Practical Recommendations for Model Choice with Nonlinear Time Series Models

- Often researcher faces questions
- Is a time series model linear or nonlinear?
- If nonlinear what type of nonlinearity?
- Sometimes economic theory and knowledge about the economy can help answer
- If not, then experiment:
- Try some generic test (like RESET or others in textbook)
- If any indication of nonlinearity, look at IC's for a few different nonlinear models
- E.g. TAR (different choices for z_t, d and τ), structural break models (different choices for τ)
- Choose specification with best IC

Other Threshold Autoregressive Models

- Several variants on the TAR model I will not go through
- STAR = smooth transition autoregressive model (see chapter 9 of textbook)
- e.g. TAR switches abruptly from one regime to another, STAR does this more gradually
- My models and examples involve switching between two regression models
- Can do very similar things with error variances
- E.g. Tsay (Chapter 3) has threshold GARCH model (TGARCH)
- GARCH involves equation for volatilities of asset returns, r_t

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 r_{t-1}^2$$

- TGARCH allows this to differ across regimes
- e.g. one GARCH equation if $r_{t-1} < \tau$, another if $r_{t-1} \geq \tau$

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

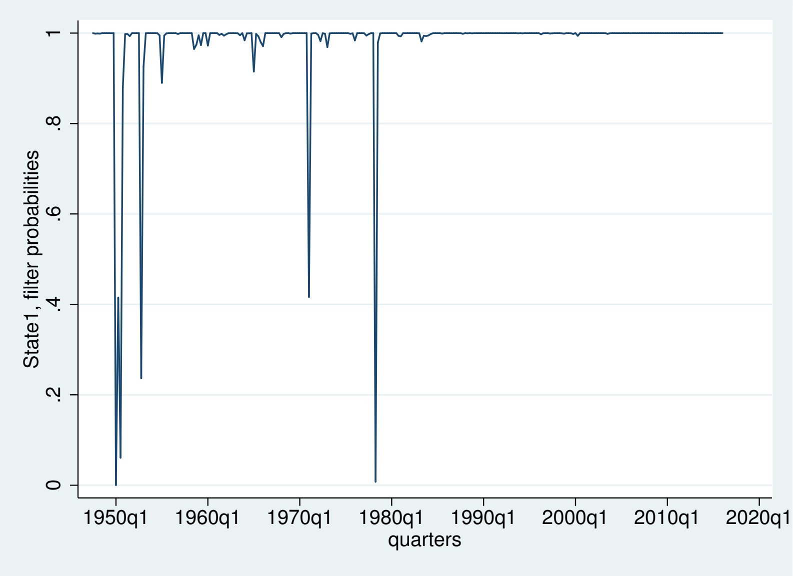
Example: Markov Switching in US GDP growth

- 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"

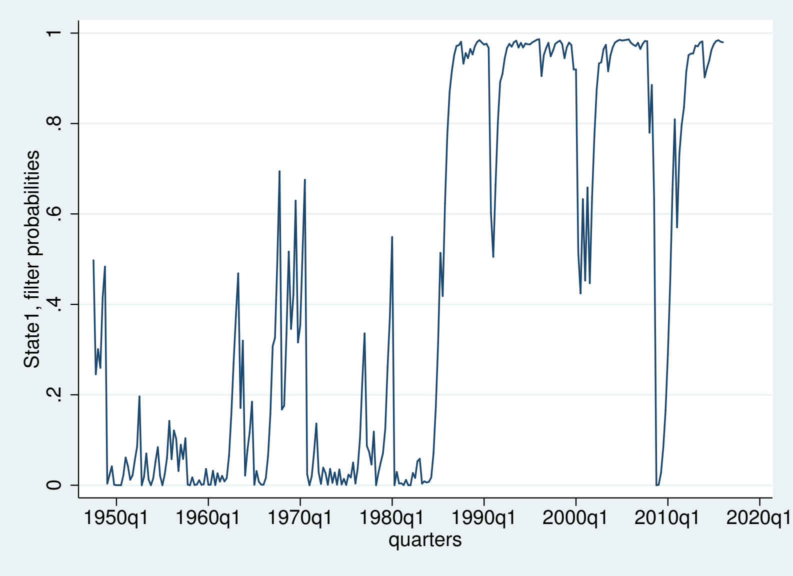
Example: Markov Switching in US GDP growth

- 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

- In many (most?) macro and finance time series evidence of parameter change
- Often need nonlinear time series model to avoid mis-specification
- But what kind of nonlinearity?
- I have gone through variety of models in this lecture
- Abrupt break at specific time: Structural break
- Gradual change over time: TVP-AR (or STAR)
- Switching between regimes: TAR or Markov switching
- ANNs are very flexible approach suitable when nonlinear form is unknown
- Lecture focusses on AR(1) but these models also work with AR(p), regressions, VARs or even volatilities