ADVANCED TIME SERIES ECONOMETRICS LAB 2

Ping Wu

University of Strathclyde

ping.wu@strath.ac.uk

TAR Model: checklist

- Create dummy variable to represent threshold.
- Run the regression.
- Interpret the output of each model:
  Does the AR coefficient change across regimes?
  Can you comment on the magnitude of the AR coefficient in each regime?
  Is this change statistically significant?
- Compare models:
  Does the AR(1), homoscedastic or heteroskedastic model have the lowest IC?

Ex1: Starting

- Create variables

```
generate logSP500 = log(SP500)
generate CH_SP500 = D.logSP500
```

log first difference of stock prices (interpretation?)

- Run AR (1) regression

```
///AR(1) MODEL-----------------------------------------

regress CH_SP500 L.CH_SP500
```

can you write the equation for this model?

- Estimate IC

```
///INFORMATION CRITERIA

estat ic
```

Ex1: Starting

- Write out the TAR model:
  Using two equations where we have coefficients ρ_1 and ρ_2 (see lecture slide 19).
  Using one equation and a dummy variable where we have two coefficients ρ_1 and γ see lecture slide 10).
- Create dummy variable

```
*following material for TAR model----------------------------
generate SP_thresh1 = 0
replace SP_thresh1 = CH_SP500 if CH_SP500>0
```

  generate a variable x_t which has a value of 0
  $x_t = y_t$ if $y_t > 0$
- Run Homoskedastic TAR using dummy variable

```
///BREAK IN MEAN (HOMOSKEDASTIC TAR)

regress CH_SP500 L.CH_SP500 L.SP_thresh1
estat ic
```

  regress y_t on y_{t-1} and x_{t-1}

Ex1: Starting

- Interpret coefficients
  What are the values of ρ_1, ρ_2, and γ? (see slide 10)
  Are coefficients different across regimes?
  Is this statistically significant?

Ex1: Tips

- Gary has given you code to estimate the heteroskedastic TAR. This model is estimated using two separate regressions and no dummy variables.
- Hint 1: we obtain ρ_1, ρ_2 straight away.
- Hint 2: we get two different information criterion from our two regressions.
- Hint 3: we need one overall information criterion for the heteroskedastic TAR.
- Which is best: AR (1), homoscedastic or heteroskedastic?
- Hint 4: information criteria and p-values might not always give us the same answer.

Ex1F: Model

- AR(1) Model

$$y_t = \alpha y_{t-1} + u_t$$

- TAR Model

$$y_t = \begin{cases} \rho_1 y_{t-1} + \epsilon_{1t}, & (y_{t-1} - \bar{y})^2 \leq \overline{(y_{t-1} - \bar{y})^2} \\ \rho_2 y_{t-1} + \epsilon_{2t}, & (y_{t-1} - \bar{y})^2 > \overline{(y_{t-1} - \bar{y})^2} \end{cases}$$

- Interpretation of regimes:
  Stock market volatility last month $\leq$ average stock market volatility;
  Stock market volatility last month $>$ average stock market volatility.

Ex1F: CODE Homoskedastic TAR

```
///PART F

sum CH_SP500

generate z1 = (L.CH_SP500 - 0.005303)^2
generate z2 = (CH_SP500 - 0.005303)^2

sum z1
sum z2

generate SP_thresh2 = 0
replace SP_thresh2 = CH_SP500 if z2> .0012676

///BREAK IN MEAN (HOMOSKEDASTIC TAR)

regress CH_SP500 L.CH_SP500 L.SP_thresh2
estat ic
```

- Sum used to obtain mean of stock price index growth, $\bar{y}$
- Generate threshold trigger, $\mathrm{z}_{2,\mathrm{t}}$
- Sum used to obtain mean of threshold trigger, $\overline{Z}_2$
- Generate dummy $= 0$ if $\mathrm{z}_{2,\mathrm{t}} \leq \overline{z_2}$;
  Generate dummy $= y_t$ if $\mathrm{z}_{2,\mathrm{t}} > \overline{z_2}$
- Use the lagged dummy in the regression

Ex1F: CODE Heteroskedastic TAR

```
///BREAK IN MEAN AND VARIANCE (HETEROSKEDASTIC TAR)

///Post-break data

regress CH_SP500 L.CH_SP500 if z1>=.0012694
estat ic

///Pre-break data

regress CH_SP500 L.CH_SP500 if z1<.0012694
estat ic

///Using Stata Threshold function - tau is estimated
threshold CH_SP500, regionvars(l.CH_SP500) threshvar(z1)
```

- Sum used to obtain mean of stock price index growth, $\bar{y}$
- Generate threshold trigger, $\mathrm{z}_{1,\mathrm{t}}$
- Sum used to obtain mean of threshold trigger, $\overline{z_1}$
- Run regression using $\mathrm{z}_{1,\mathrm{t}} \geq \bar{z}$
- Run regression using $\mathrm{Z}_{1,\mathrm{t}} < \bar{z}$

- Model 1: AR (1)

CH_SP500	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
CH_SP500						
L1.	.2435219	.0369678	6.59	0.000	.1709386	.3161051
_cons	.0040433	.0013316	3.04	0.002	.0014289	.0066578

```
.
. ///INFORMATION CRITERIA
>
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	690	1321.97	1343.071	2	-2682.142	-2673.069

Note: BIC uses N = number of observations. See [R] BIC note.

- $y_t = 0.004 + 0.244 y_{t-1}$

Ex1F: Some output

- Model 2: AR(1) with Break in Mean:

CH_SP500	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
CH_SP500						
L1.	.0997625	.079729	1.25	0.211	-.0567793	.2563043
SP_thresh2						
L1.	.1839458	.0904436	2.03	0.042	.0063667	.3615249
_cons	.0049715	.0014048	3.54	0.000	.0022134	.0077296

```
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	690	1321.97	1345.142	3	-2684.284	-2670.674

- $y_t = 0.005 + 0.100 y_{t-1} + 0.184 x_{t-1}$ where $x_t = D_t y_t$
- $H_0 : \gamma = 0$
- P-value $= 0.042$
- Reject H_0, evidence of a break in mean at the 5% significance level.
- Model 2 IC $<$ Model 1 IC, evidence of a break in mean.

Ex1F: Some output

- Model 3: AR(1) with Break in Mean and Variance ($z < \bar{z}$)

Source	SS	df	MS
Model	.001667275	1	.001667275
Residual	.52691225	526	.001001734
Total	.528579525	527	.001002997

Number of obs	=	528
F(1, 526)	=	1.66
Prob > F	=	0.1976
R-squared	=	0.0032
Adj R-squared	=	0.0013
Root MSE	=	.03165

CH_SP500	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
CH_SP500						
L1.	.0964228	.0747399	1.29	0.198	-.0504025	.2432481
_cons	.0051385	.0015047	3.41	0.001	.0021824	.0080945

- $y_t = 0.005 + 0.096 y_{t-1}$, $var\left(\widehat{\epsilon_{1t}}\right) = 0.527$
- Note: coefficient ρ_1 is insignificant

Ex1F: Some output

- Model 3: AR(1) with Break in Mean and Variance ($z \geq \bar{z}$)

```
. regress CH_SP500 L.CH_SP500 if z1>=.0012694

      Source |       SS           df       MS      Number of obs   =       162
-------------+----------------------------------   F(1, 160)       =     29.89
       Model |  .05447524         1  .05447524   Prob > F        =    0.0000
    Residual |  .291651038       160  .001822819   R-squared       =    0.1574
-------------+----------------------------------   Adj R-squared   =    0.1521
       Total |  .346126278       161  .002149853   Root MSE        =    .04269

------------------------------------------------------------------------------
    CH_SP500 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    CH_SP500 |
         L1. |   .2832879   .0518203     5.47   0.000     .1809478    .3856279
             |
       _cons |   .0045138   .0033604     1.34   0.181    -.0021226    .0111502
------------------------------------------------------------------------------
```

- $y_t = 0.005 + 0.283 y_{t-1}$, $var\left(\widehat{\epsilon_{2t}}\right) = 0.292$
- Note: coefficient ρ_2 is significant, lag of stock market growth has greater explanatory power in high volatility regime
- The improved model fit results in lower estimated error variance, $var\left(\widehat{\epsilon_{2t}}\right) < var\left(\widehat{\epsilon_{1t}}\right)$

Ex1F: Some output

- Model 2: AR(1) with Break in Mean:

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	690	1321.97	1345.142	3	-2684.284	-2670.674

- Model 3: AR(1) with Break in Mean and Variance:

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	162	268.1643	282.0352	2	-560.0705	-553.8953

- Model 3 IC: Add together IC for two regimes
  AIC: -2706, BIC: -2691
  Model 3 IC $<$ Model 2 IC, evidence of a break in mean and variance.
- Conclusions:
  When modelling stock market index growth, it matters whether stock market volatility is high or low.
  Evidence that high and low volatility regimes have different AR coefficients and error variances.

MARKOV SWITCHING model: checklist

- Estimate model using “mswitch” command
- Interpret the output of each model:
  Do coefficients/error variances change across regimes?
  How long does each regime last for?
  What are the probabilities of staying in the regime or changing to a different regime?
  What is the probability of being in each regime at each point in time?
- Compare models:
  Look at which model specification has the lowest IC.

Ex2: Tips

```
mswitch dr RGDP_CH, switch(L.RGDP_CH L2.RGDP_CH L3.RGDP_CH L4.RGDP_CH) varswitch
```

- *dr* = dynamic regression, allow a quick adjustment after the process changes states
- $switch(L.RGDP_C H)$, regime switching in AR coefficients
- *varswitch*, regime switching in error variance
- HINT: use help "mswitch" to figure our how to add more lags and states
- Main output
  Do coefficients vary across regimes?
  Do error variances (labelled "sigma 1"/"sigma 2") vary across regimes?

Ex2: Tips

- estat transition
  p11 gives probability of staying in regime 1
  p12 gives probability of switching from regime 1 to regime 2
  p22 gives probability of staying in regime 2
  p21 gives probability of switching from regime 2 to regime 1
- estat duration estimated duration of each regime
- Note: confidence intervals may be wide if we have few observations and estimation is imprecise

Ex2: Tips

- predict fprob, pr smethod (filter)
  Estimates the probability of being in regime 1 at each point in time
  This graph in particular may provide a clue about how we can interpret our regimes

A note on persistence

- If we have: $y_t = \rho_1 y_{t-1} + u_t$
- We want to check for persistence e.g. is the interest rate this period related to last period's value?
- If $\rho_1 > 0$ we have persistence
- The larger ρ_1 is, the more persistent our variable
- By looking at the size of ρ_1 and ρ_2 we can determine whether the level of persistence is different across regimes.
- An example of a persistent variable (in developed countries) is the interest rate. We want interest rate changes to be "smooth".
- An example of a less persistent variable is GDP growth.