Lecture 9: Structural Breaks and Threshold Model

Structural Breaks

- Consider the modified AR(1) model

$$y_t = \begin{cases} b_0 + b_1 y_{t-1} + e_t, & (t = 1, 2, \ldots, T_b) \\ a_0 + a_1 y_{t-1} + e_t, & (t = T_b + 1, T_b + 2, \ldots, T) \end{cases} \quad (1)$$

- T_b denotes the break date. Before the break date, the intercept is b_0 and slope is b_1; after the break date they become a_0 and a_1.

Dummy Variable

- Define a dummy variable as

$$d_t = \begin{cases} 0, & (t = 1, 2, \ldots, T_b) \\ 1, & (t = T_b + 1, T_b + 2, \ldots, T) \end{cases} \tag{2}$$

- Then the original model can be rewritten as

$$y_t = \phi_0 + \phi_1 y_{t-1} + \gamma_0 d_t + \gamma_1 (d_t \times y_{t-1}) + e_t \tag{3}$$

It follows that

$$\begin{aligned} \phi_0 &= b_0 \\ \phi_1 &= b_1 \\ \gamma_0 &= a_0 - b_0 \\ \gamma_1 &= a_1 - b_1 \end{aligned}$$

Chow Test

- The null hypothesis is that there is NO structural break, i.e.,

$$H_0 : b_0 = a_0; a_1 = b_1 \tag{4}$$

- In the dummy variable model, this null hypothesis is equivalent to

$$H_0 : \gamma_0 = 0; \gamma_1 = 0 \tag{5}$$

- Gregory Chow proposes the F test for (5), which follows the F distribution under the assumption that T_b is known.

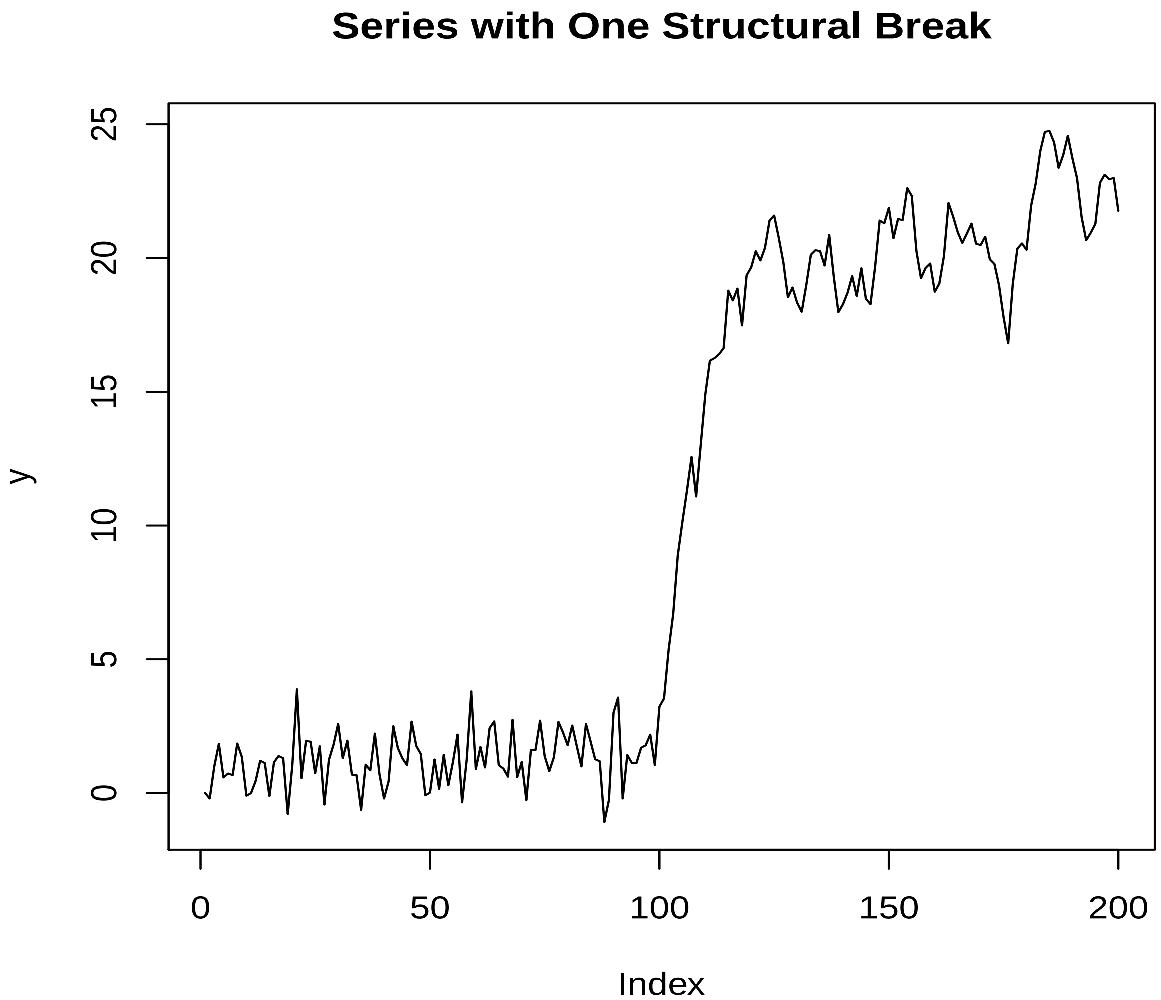
Unknown Break Date

- If T_b is unknown, we can estimate it by grid search.
  1. That means we run regression (3) using the first potential break date. Save the residual sum of squares (RSS)
  2. Then we move to next period, and treat it as a new potential break date. Run the regression again, and save RSS
  3. We repeat until we try all the potential break dates. The regression with the smallest RSS produces the estimated break date
- In R, we need to implement the grid search using a loop.

Example

```
b0 = 1
b1 = 0.2
a0 = 2
a1 = 0.9
br = 100
for (i in 2:br) {
y[i] = b0+b1*y[i-1] + e[i]
}
for (i in (br+1):n) {
y[i] = a0+a1*y[i-1] + e[i]
}
```

Plot



Series with One Structural Break
y
0
5
10
15
20
25
0
50
100
150
200
Index

Dummy Variable and Interaction Term

```
tr = 1:n
d = rep(0,n)
Tb = 100
d[tr>Tb] = 1             # dummy variable
cbind(y,d)

y.1 = c(NA, y[1:n-1])  # first lag
i.t = d*y.1              # interaction term
```

Here we assume the break date is known, and it is $T_b = 100$

Unrestricted and Restricted Regressions

```
# unrestricted model
m.u = lm(y~y.1+d+i.t)
summary(m.u)

# restricted model, which imposes the null hypothesis
m.r = lm(y~y.1)
summary(m.r)
```

Unrestricted Regression

```
lm(formula = y ~ y.1 + d + i.t)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.16229    0.15418   7.538 1.75e-12 ***
y.1          0.08569    0.09735   0.880   0.3798
d            0.94418    0.47700   1.979   0.0492 *
i.t          0.81400    0.10004   8.137 4.68e-14 ***

Residual standard error: 0.9598 on 195 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared: 0.9899,     Adjusted R-squared: 0.9898
F-statistic:  6386 on 3 and 195 DF,  p-value: < 2.2e-16
```

Remarks

1. $\hat{\gamma}_0 = 0.94418$, which is significant, and is close to the true value of $a_0 - b_0 = 2 - 1 = 1$

2. $\hat{\gamma}_1 = 0.81400$, which is significant, and is close to the true value of $a_1 - b_1 = 0.9 - 0.2 = 0.7$

3. Multiple R-squared of 0.9899 can be used to construct the F test.

Chow Test and P value

```
chowtest = ((0.9899-0.9844)/2)/((1-0.9899)/195)
pvalue = pf(chowtest, 2, 195, lower.tail = FALSE)
c(chowtest, pvalue)
[1] 5.309406e+01 3.905581e-19
```

Because the unrestricted and restricted models have the same dependent variable, we can construct the F test using R^2 :

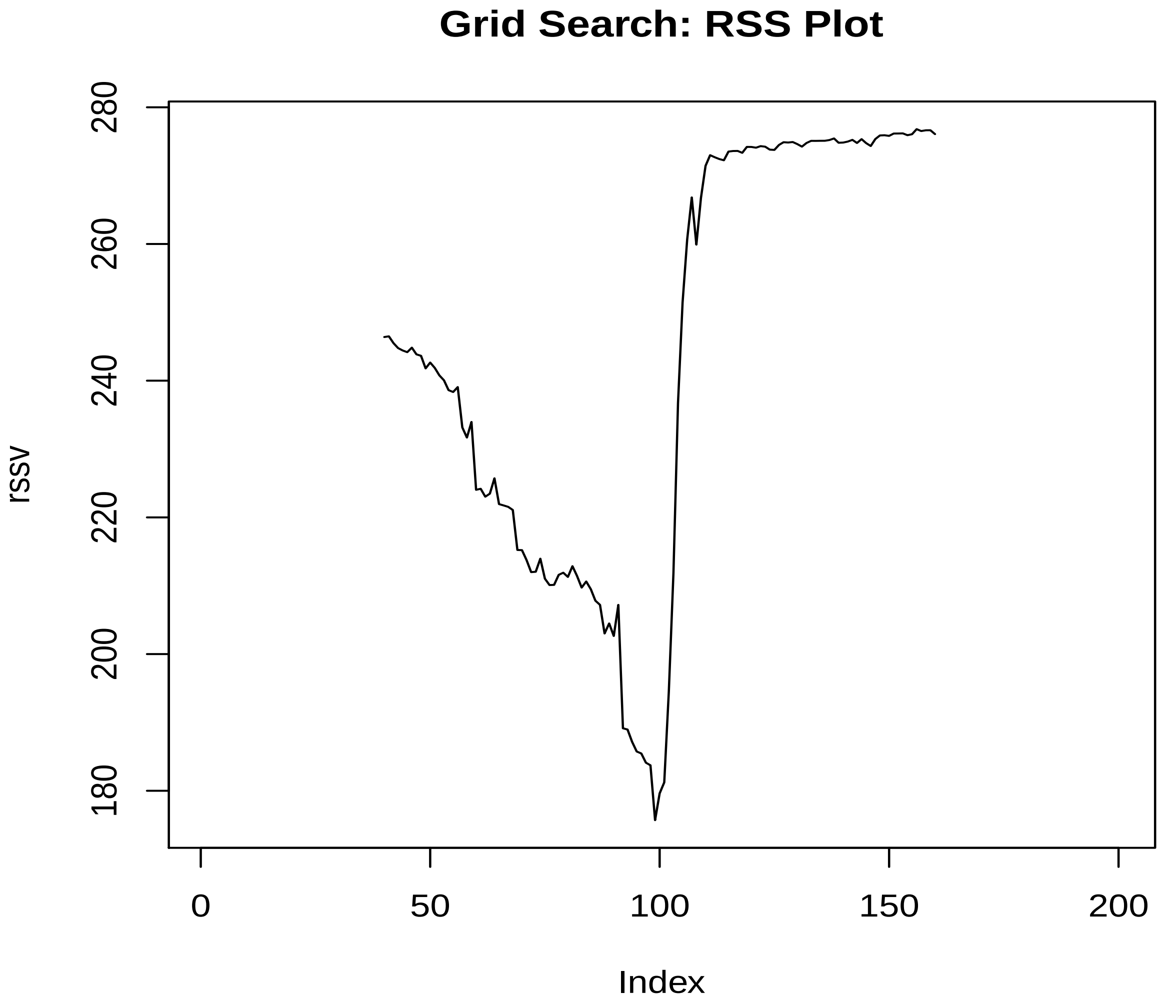
$$\texttt{f test} = \frac{(R_u^2 - R_r^2)/q}{(1-R_u^2)/\texttt{df}_u} = \frac{(0.9899-0.9844)/2}{(1-0.9899)/195} = 53.09$$

In this case, the p value is much smaller than 0.05. So the null hypothesis of no break is rejected.

Grid Search T_b

```
min.br = (0.2*n)     # lower limit of searching range
max.br = (0.8*n)     # upper limit of searching range
rssv = rep(NA, n)    # the vector that saves RSS
for (tb in min.br:max.br) {
d = rep(0,n)
d[tr>tb] = 1
y.1 = c(NA, y[1:n-1])
i.t = d*y.1
m.u = lm(y~y.1+d+i.t)
rssv[tb] = sum(resid(m.u)^2)
}
plot(rssv, type = "l", main = "Grid Search: RSS Plot")
```

Plot of RSS against the break date



Estimated T_b

```
min.rss = min(rssv,na.rm = TRUE)
est.breakdate = which(rssv==min.rss)
est.breakdate
[1] 99
```

The break date is estimated as 99, close to the true value of 100.

Chow Test with Estimated Break Date

- Next we can do the Chow test using the estimated T_b.
- However, the Chow test no longer follows the F distribution
- Instead it follows a nonstandard distribution that involves Brownian Bridge.
- To learn more, google "sup Wald statistic" or "Quandt Likelihood Ratio (QLR) statistic". By definition

$$\texttt{QLR} = \texttt{max}\left[F(T_{b1}), F(T_{b2}), \ldots, F(T_{bk})\right]$$

Davies Problem

- Note that under the null hypothesis

$$H_0 : \gamma_0 = 0; \gamma_1 = 0 \tag{6}$$

  the dummy variable disappears, so the break date T_b cannot be identified.

- In that case, T_b becomes a nuisance parameter which can be identified only under the alternative hypothesis.
- This technical issue is called Davies problem, first raised by Davies (1977, Biometrika)
- The QLR test follows nonstandard distribution due to the Davies problem

Threshold Model

Threshold Autoregression

- Consider another modified AR(1) model, called threshold autoregression (TAR)

$$y_t = \begin{cases} b_0 + b_1 y_{t-1} + e_t, & \text{if } (y_{t-1} < \tau) \\ a_0 + a_1 y_{t-1} + e_t, & \text{if } (y_{t-1} \geq \tau) \end{cases} \quad (7)$$

- τ denotes the threshold parameter. There are two regimes: in regime one $y_{t-1} < \tau$; in regime two $y_{t-1} \geq \tau$.
- For example, regime 1 may be the recession, while regime 2 is the economic boom

Dummy Variable

- Define a new dummy variable as

$$d_t = \begin{cases} 0, & \text{if } (y_{t-1} < \tau) \\ 1, & \text{if } (y_{t-1} \geq \tau) \end{cases} \tag{8}$$

- Then the TAR model can be rewritten as

$$y_t = \phi_0 + \phi_1 y_{t-1} + \gamma_0 d_t + \gamma_1 (d_t \times y_{t-1}) + e_t \tag{9}$$

- Similar to the Chow test, now we can test the null hypothesis of no regime-switching

$$H_0 : \gamma_0 = 0; \gamma_1 = 0$$

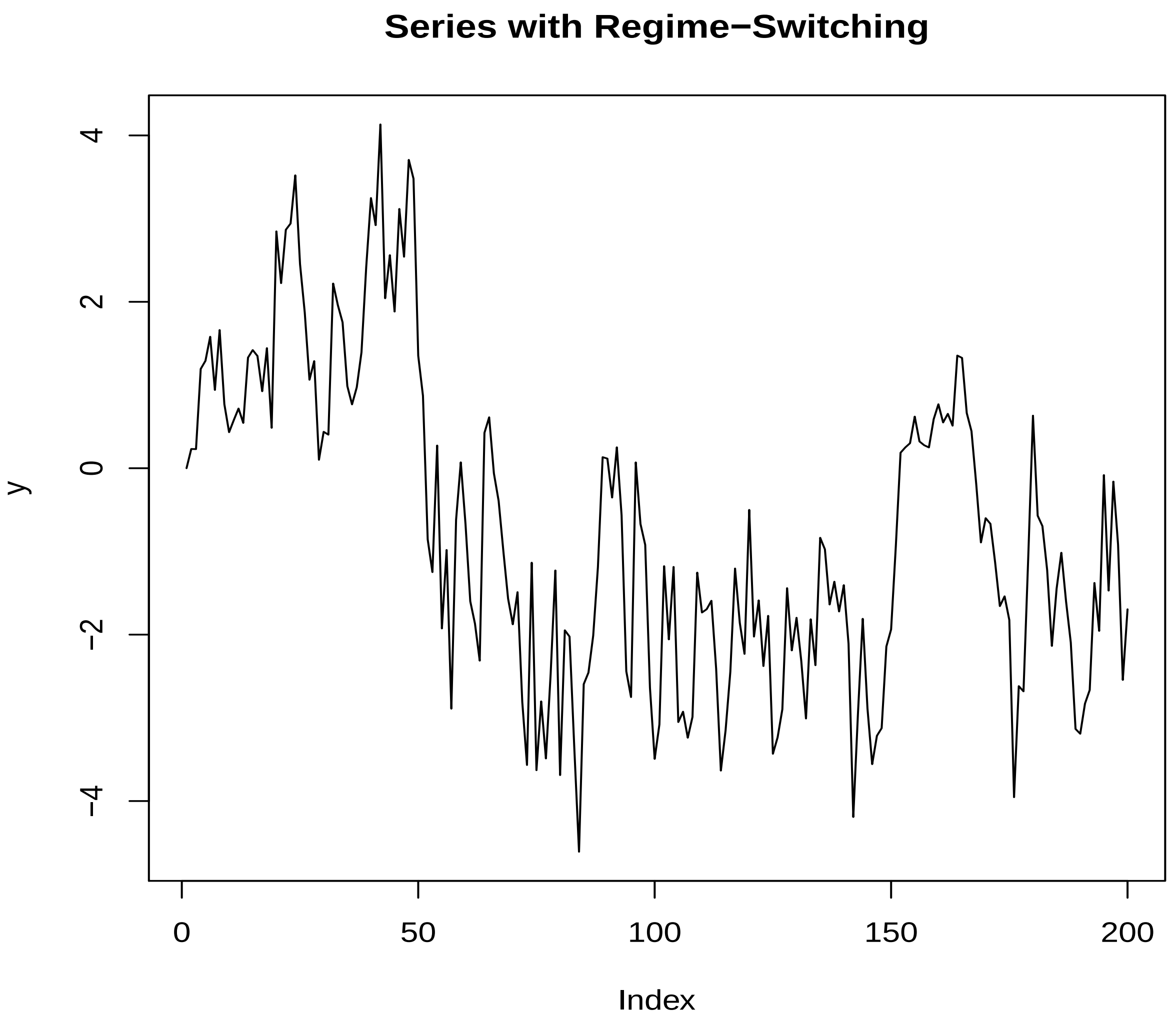
and the test follows nonstandard distribution if τ is estimated by grid search

Example

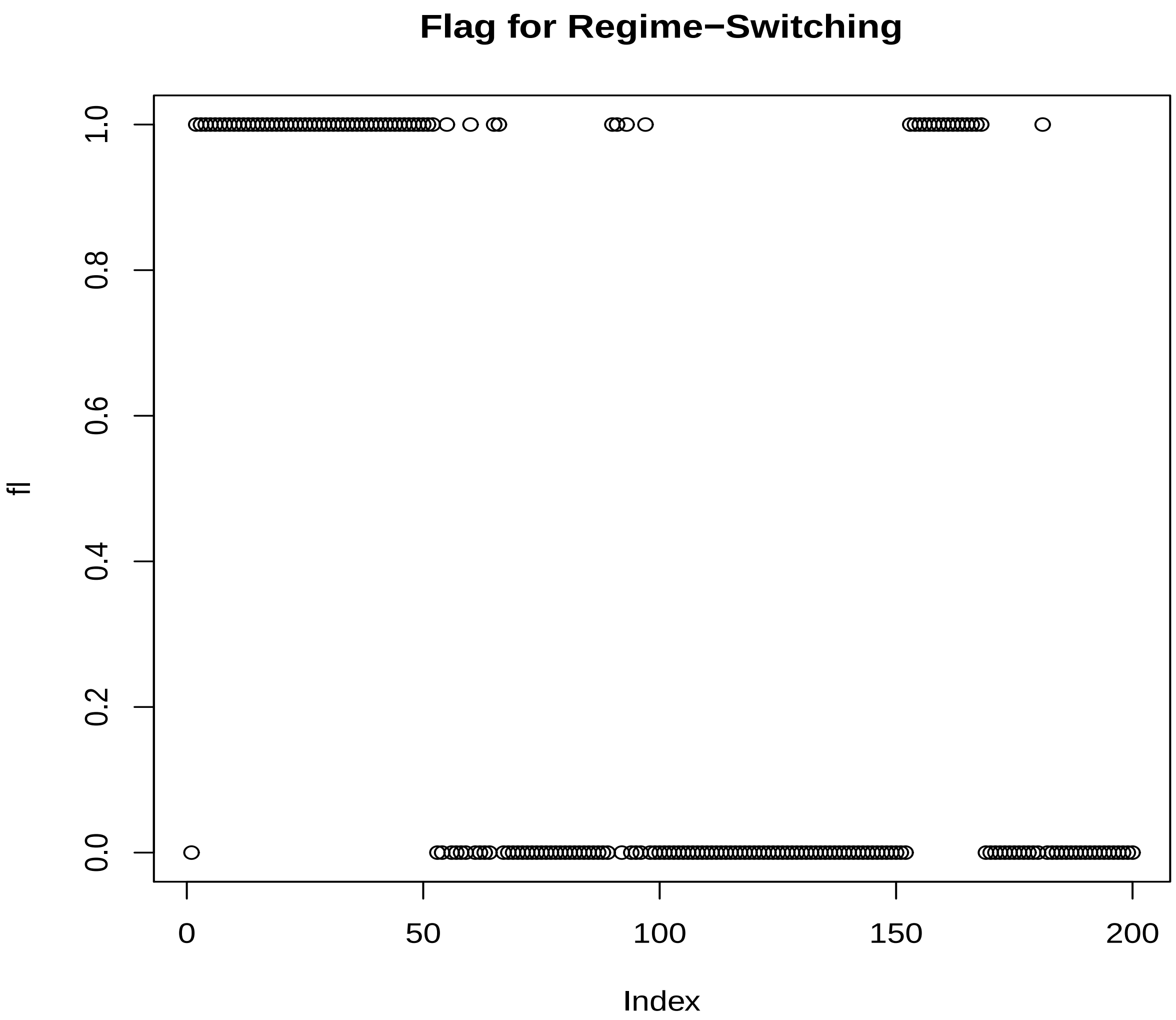
```
tau = 0
for (i in 2:n) {
if (y[i-1]<tau) {
y[i] = b0+b1*y[i-1] + e[i]
}
else {
y[i] = a0+a1*y[i-1] + e[i]
fl[i] = 1
}
}
```

Note that there is "if-else" inside the loop.

Plot I



Plot II



Flag for Regime−Switching
fl
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
Index

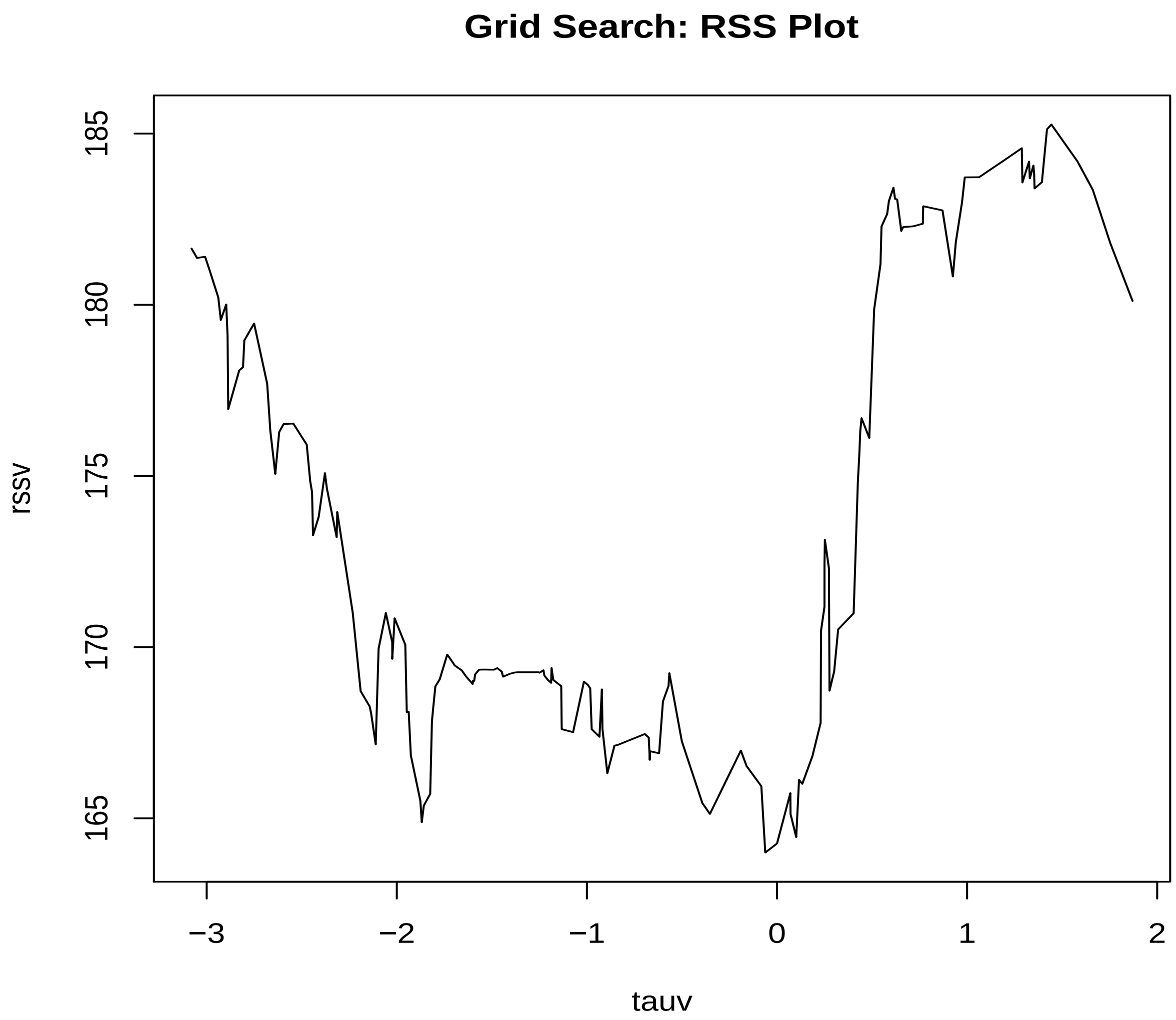
Warning about TAR

- You will get very inaccurate or misleading result of (9) if the dummy variable d_t has little variation
- That happens when the frequency of regime-switching is extremely low
- Plot II displays the flag (or indicator) for regime-switching. In this case, the chance of staying in either regime is reasonable. So TAR model makes sense.
- If regime switching seldom happens, then consider the traditional AR model.

Grid Search

```
o.y1 = y.1[order(y.1)]  # ordered first lag
min.ta = (0.1*n)
max.ta = (0.9*n)
for (ta in min.ta:max.ta) {
tau = o.y1[ta]           # potential threshold value
d = rep(0,n)
d[y.1>=tau] = 1
i.t = d*y.1
m.u = lm(y~y.1+d+i.t)
rssv[ta] = sum(resid(m.u)^2)
tauv[ta] = tau
}
plot(tauv, rssv, type = "l", main = "Grid Search: RSS Plot")
```

RSS Plot



Remarks

1. The RSS is minimized by a value which is close to the true value 0
2. However, the RSS plot is not as much "V-shaped" as the structural break model. Note that the RSS plot seems to indicate another threshold value close to -2, which actually does not exist.
3. You should exercise more caution when working with threshold model. It is always a good idea to provide the RSS plot.