Multiple changes in coefficients of autoregressive models

Miriam Marušiaková

Department of Statistics, Charles University, Prague

maruskay@gmail.com

Abstract

We deal with an F type test for detection of changes in parameters of an autoregressive model. We apply resampling methods to approximate the critical values of the test.

1. AR(p) model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta}_j + e_i \qquad t_{j-1} < i \le t_j, \quad j = 1, \dots, k+1$$

unknown change points $t_1, t_2, \dots, t_k$

convention $t_0 = p,\ t_{k+1} = n$

$$\boldsymbol{x}_i' = (1, y_{i-1}, y_{i-2}, \dots, y_{i-p}), \qquad i = p+1, \dots, n$$
$$\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp})$$

$$\beta_{j0} = (1 - \sum_{l=1}^{p} \beta_{jl})\mu_j, \quad \mu_j = Ey_i, \qquad t_{j-1} < i \le t_j$$

2. Assumptions

- Stationary AR(p) in segments
- i.i.d. errors e_i with
  $$Ee_i = 0, \quad 0 < Var(e_i) = \sigma^2 < \infty, \quad E|e_i|^4 < \infty$$
- $t_j = \lfloor n\lambda_j \rfloor,\ j = 1, \dots, k$
  $0 = \lambda_0 < \lambda_1 < \dots < \lambda_{k+1} = 1.$

3. Estimation of change points for known k

- The minimal length of a segment is $(n-p)\varepsilon$
  $$T_\varepsilon = \{(t_1, \dots, t_k) : t_{j+1} - t_j \ge n\varepsilon, \forall j = 0, \dots, k\}$$
- SSR for a given partition $(t_1, \dots, t_k)$
  $$S_n(t_1, \dots, t_k) \equiv \sum_{j=1}^{k+1} \min_{\boldsymbol{\beta}_j} \sum_{i=t_{j-1}+1}^{t_j} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_j)^2$$
- The change points are estimated as ([2, 7])
  $$(\hat{t}_1, \dots, \hat{t}_k) = \arg\min_{t_1, \dots, t_k \in T_\varepsilon} S_n(t_1, \dots, t_k)$$

4. F type test [1]

No change versus k changes (k fixed)

$$F_n^\varepsilon(k, q) = \max_{t_1, \dots, t_k \in T_\varepsilon} F_n(t_1, \dots, t_k, q)$$

$$F_n(t_1, \dots, t_k, q) = \frac{1}{kq} \frac{SSR_0 - SSR_k}{\tilde{\sigma}_n^2},$$

where

$$\tilde{\sigma}_n^2 = \frac{SSR_k}{n-q} \to_p \sigma^2 \quad \text{under } H_0 \text{ and } H_A$$

$$SSR_0 = \min_{\boldsymbol{\beta}} \sum_{i=p+1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2, SSR_k = S_n(t_1, \dots, t_k)$$

q - number of regressors

5. Equivalent expression of the test

$$\hat{e}_i = y_i - \boldsymbol{x}_i'\boldsymbol{C}_n^{-1} \sum_{i=p+1}^{n} \boldsymbol{x}_i y_i \overset{H_0}{=} e_i - \boldsymbol{x}_i'\boldsymbol{C}_n^{-1} \sum_{i=p+1}^{n} \boldsymbol{x}_i e_i$$

$$\boldsymbol{C}_{k,l} = \sum_{i=k+1}^{l} \boldsymbol{x}_i \boldsymbol{x}_i', \quad \boldsymbol{C}_{p,l} = \boldsymbol{C}_l, \quad k+1, l = p+1, \dots, n$$

$$\begin{aligned} &SSR_0 - SSR_k \\ &= \sum_{j=1}^{k+1} \left(\sum_{i=t_{j-1}+1}^{t_j} \boldsymbol{x}_i \hat{e}_i \right)' \boldsymbol{C}_{t_{j-1},t_j}^{-1} \left(\sum_{i=t_{j-1}+1}^{t_j} \boldsymbol{x}_i \hat{e}_i \right) \\ &\overset{H_0}{=} \Bigg\{ - \left(\sum_{i=p+1}^{n} \boldsymbol{x}_i e_i \right)' \boldsymbol{C}_n^{-1} \left(\sum_{i=p+1}^{n} \boldsymbol{x}_i e_i \right) \qquad (1) \\ &\quad + \sum_{j=1}^{k+1} \left(\sum_{i=t_{j-1}+1}^{t_j} \boldsymbol{x}_i e_i \right)' \boldsymbol{C}_{t_{j-1},t_j}^{-1} \left(\sum_{i=t_{j-1}+1}^{t_j} \boldsymbol{x}_i e_i \right) \Bigg\} \end{aligned}$$

6. Limit distribution under H_0

Using results from [3], one can show

$$F_n^\varepsilon(k, q) \to_d F^\varepsilon(k, q) = \sup_{\lambda_1, \dots, \lambda_k \in \Lambda_\varepsilon} F(\lambda_1, \dots, \lambda_k, q),$$

where

$$\begin{aligned} &F(\lambda_1, \dots, \lambda_k, q) \\ &= \frac{1}{kq} \sum_{j=1}^{k} \frac{\|\lambda_j \boldsymbol{W}(\lambda_{j+1}) - \lambda_{j+1} \boldsymbol{W}(\lambda_j)\|^2}{\lambda_j \lambda_{j+1}(\lambda_{j+1} - \lambda_j)}, \end{aligned}$$

$\boldsymbol{W}(t)$ is a vector of q independent standard Wiener processes and

$$\Lambda_\varepsilon = \{(\lambda_1, \dots, \lambda_k) : \lambda_{j+1} - \lambda_j \ge \varepsilon, \forall j = 0, \dots, k\}.$$

7. Bootstraping

- $\tilde{e}_{p+1}^*, \dots, \tilde{e}_n^*$ bootstrap sample with replacement from $\tilde{e}_{p+1}, \dots, \tilde{e}_n$
  $$\tilde{e}_i = y_i - \boldsymbol{x}_i'\boldsymbol{C}_{\hat{t}_{j-1},\hat{t}_j}^{-1} \sum_{i=\hat{t}_{j-1}+1}^{\hat{t}_j} \boldsymbol{x}_i y_i, \qquad \hat{t}_{j-1} < i \le \hat{t}_j$$
- Bootstrap version of $F_n^\varepsilon(k, q)$
  $$F_n^{\varepsilon *}(k, q) = \max_{t_1, \dots, t_k \in T_\varepsilon} F_n^*(t_1, \dots, t_k, q)$$
  $$F_n^*(t_1, \dots, t_k, q) = \frac{1}{kq} \frac{SSR_0^* - SSR_k^*}{\tilde{\sigma}_n^2},$$
  $SSR_0^* - SSR_k^*$ as in (1) with e_i replaced by $\tilde{e}_i^*$.

8. Theorem

Let the data follow H_0 or alternatives. Then, under given assumptions,

$$\sup_x \left\{ P\left(F_n^{\varepsilon *}(k, q) \le x | \boldsymbol{y}\right) - P\left(F^\varepsilon(k, q) \le x\right) \right\} \to_p 0$$

Proof - using key results from [4].

9. Simulation results [6, 7]

$$y_i = x_i \beta_j + e_i, \qquad t_{j-1} < i \le t_j, \quad j = 1, \dots, m+1$$
$$x_i = y_{i-1}$$

Graphs - inspiration from [5]

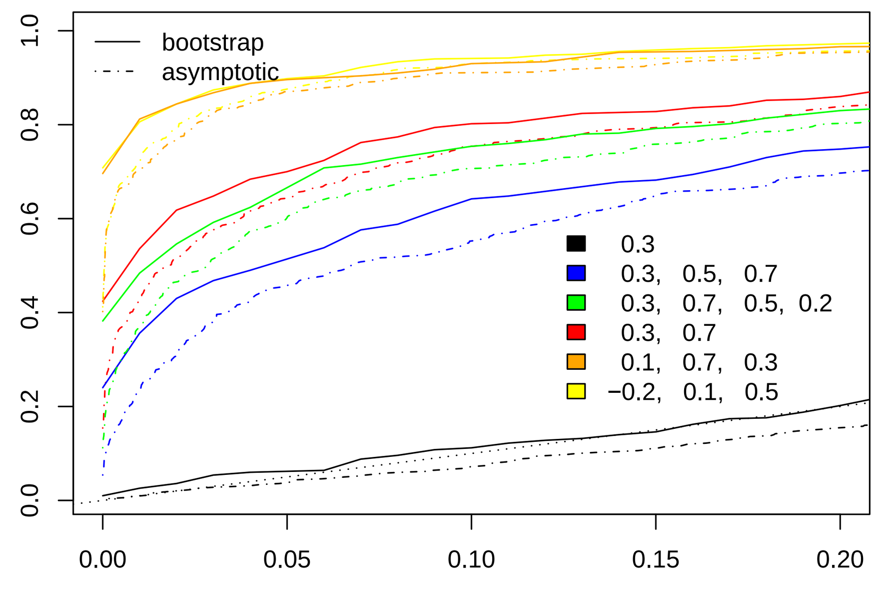

Figure 1: *Size-power-curves plots (SPC-plots) showing empirical distribution function of p-values of the test $F_n^\varepsilon(k, q)$ for the null hypothesis or some alternative with respect to the bootstrap distribution which was used to determine the critical values of the test. On the y-axis actual α-errors or 1 - β-errors for chosen quantiles on the x-axis. Here $n = 200$, $k = 2$, $q = 1$, $\varepsilon = 0.15$. For comparison p-values with respect to the asymptotic distribution $F^\varepsilon(k, q)$ calculated (dotdashed lines).*

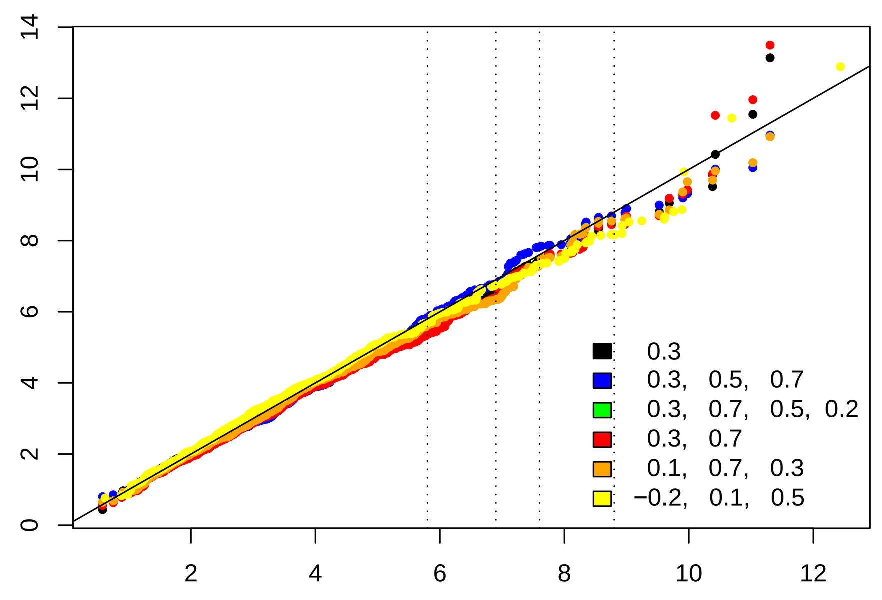

Figure 2: *QQ-plot of empirical conditional quantiles of $F_n^{\varepsilon *}(k, q)$ (1000 bootstrap samples), given one realisation $\boldsymbol{y}$, against empirical quantiles of $F_n^\varepsilon(k, q)$ under H_0 (1000 repetitions). Dotted lines are 90, 95, 97.5, 99 % empirical quantiles of $F_n^\varepsilon(k, q)$ under H_0. Here $n = 200$, $k = 2$, $q = 1$, $\varepsilon = 0.15$.*

References

[1] Bai J and Perron P. (1998), Econometrica **66**, 47 – 78.

[2] Bai J and Perron P. (2003), The Econometrics Journal **6**, 72 – 78.

[3] Hušková M, Prášková Z, Steinebach J. (2007), J. Statist. Plann. Inference **137**, 1243 – 1259.

[4] Hušková M, Kirch C, Prášková Z, Steinebach J. (2008), J. Statist. Plann. Inference **138**, 1697 – 1721.

[5] Kirch C (2006), PhD thesis, University of Cologne

[6] R Development Core Team (2007), ISBN 3-900051-07-0, URL http://www.R-project.org.

[7] Zeileis A, Leisch F, Hornik K and Kleiber C. (2002), Journal of Statistical Software **7**, 1 – 38.

Acknowledgement

The work was supported by GAČR 201/05/H007. The author would like to thank Prof. Marie Hušková for fruitful discussions and Department of Mathematics, Uppsala University for a nice working place and friendly environment.