Two contributions to change-point analysis

Marie Hušková

Charles University, Prague

Lancaster University, May 2015

Outline

Outline

Introduction

I. Robust procedures for detection of a change in regression

Robust procedures versus classical L_2 procedures

off-line and on-line procedures

II. Two-sample change point analysis

Motivated by real data – measurements of jump height and speed: 432 girls, 364 boys (6–19 years)

Outline

I. Robust procedures for detection of a change in regression

Regression model-off-line version:

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \boldsymbol{\delta}_n I\{i > k_0\} + e_i, \quad i = 1, \dots, n$$

$\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ – unknown parameters

$\boldsymbol{\delta}_n = (\delta_{n1}, \dots, \delta_{np})^T$ – unknown parameters

$\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ -observed regressors

k_0 – change-point – unknown

$e_1, \dots, e_n$ – random errors with some properties

Basic problems:

(a) $H_0 : k_0 = n, \quad H_1 : k_0 < n$

(b) Estimator of k_0 and other parameters

Typically max-likelihood type test under the assumption that the error are i.i.d. $N(0, \sigma^2)$ are derived and it is checked whether the test have a good performance under weaker assumptions (usually asymptotic behavior is studied).

It is well-known that such procedures are sensitive w.r.t. non-normality of errors (heavy tailed or there are outliers), therefore robust procedures were developed. Typically related to the so called M-estimators developed by P. Huber.

Rcall: M-estimator of $\boldsymbol{\beta}$ under H_0 is defined as a minimizer of

$$\sum_{i=1}^{n} \rho(Y_i - \mathbf{x}_i^T \mathbf{b})$$

w.r.t. $\mathbf{b}$, ρ is a convex loss function, denote $\psi = \rho'$– called score function

$\boldsymbol{\beta}_n(\psi)$ – minimizer

Typical choices of $\psi(\cdot)$:

- $\psi(x) = x$, $x \in R$– L_2 norm
- $\psi(x) = sign(x)$, $x \in R$– L_1 norm
- Huber function, $K > 0$

$$\begin{aligned}\psi(x) &= x,\ |x| > K\\ &= K\,sign(x),\ |x| \le K\end{aligned}$$

- $\psi_\beta(x) = \beta I\{x > 0\} - (1-\beta)I\{x \le 0\}$ - β-quantile, $\beta \in (0,1)$
- score function related likelihood ratio

Test statistics:

$$T_n(\psi) = \sup_{0<t<1} \left\{ \frac{1}{n} \mathbf{S}^T_{\lfloor (n+1)t \rfloor}(\psi) \left(\widehat{\mathbf{\Sigma}}_n(\psi) \right)^{-1} \mathbf{S}_{\lfloor (n+1)t \rfloor}(\psi) \right\}$$

$$\mathbf{S}_k(\psi) = \sum_{i=1}^{k} \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_n(\psi)), \quad k = 1, \ldots, n$$

$\widehat{\mathbf{\Sigma}}_n(\psi)$ – an estimator of $\mathbf{\Sigma}(\psi)$

$$\mathbf{\Sigma}(\psi) = \lim_{n \to \infty} var \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi(e_i) \right\}$$

Basic property under the null hypothesis: If $\frac{1}{n} \sum_{i=1}^{k} \mathbf{x}_i \mathbf{x}_i^T \approx \frac{k}{n} \mathbf{\Sigma}$ then as $n \to \infty$

$$T_n(\psi) \to^d \sup_{0<t<1} \left\{ \sum_{j=1}^{p} B_j^2(t) \right\}$$

$B_1(t), \ldots, B_p(t)$ – independent Brownian bridges

Simulations

the R-package sandwich (see Zeileis).
$n = 100, 200, 400$
$\beta_0 = 1.0, 1.0)$
$\mathbf{x}_i = (1, x_i)$, x_i – AR (1)
$k_0 = n/2$
$\alpha = 0.05$, $p = 2$
regressors to be i.i.d. N(0,1); the errors were generated as AR(1) with heavy tailed innovations coming either from the Student distribution t1 and t2; for various values of the coefficient :

δ		(0, 0)	(0.5, 0)	(1, 0)	(0, 0.5)	(0, 1)	(0.5, 0.5)	(1, 1)
ρ	n				L_2			
0	100	1.7 (5)	4 (12)	18 (36)	4 (12)	16 (35)	8 (20)	34 (56)
	200	1.5 (5)	13 (26)	53 (70)	13 (26)	55 (71)	29 (45)	79 (88)
	400	2.0 (5)	29 (40)	83 (89)	30 (41)	83 (88)	56 (66)	94 (97)
0.2	100	2.3 (5)	3 (9)	10 (23)	4 (12)	16 (35)	7 (17)	26 (49)
	200	2.1 (5)	7 (15)	34 (49)	13 (24)	52 (67)	22 (34)	69 (80)
	400	2.1 (5)	17 (26)	64 (73)	30 (42)	83 (88)	47 (58)	92 (94)
0.4	100	3.1 (5)	4 (7)	7 (14)	5 (10)	14 (27)	7 (12)	19 (35)
	200	2.7 (5)	6 (10)	20 (28)	13 (19)	49 (58)	18 (25)	61 (70)
	400	3.6 (5)	11 (15)	43 (51)	27 (33)	80 (84)	37 (44)	88 (91)
ρ	n				Huber			
0	100	2.4 (5)	12 (22)	48 (65)	9 (19)	34 (53)	20 (33)	66 (83)
	200	2.7 (5)	38 (48)	96 (98)	35 (46)	92 (95)	68 (77)	100 (100)
	400	4.1 (5)	79 (82)	100 (100)	76 (78)	100 (100)	98 (98)	100 (100)
0.2	100	1.5 (5)	5 (14)	23 (43)	8 (19)	31 (54)	12 (28)	50 (75)
	200	2.2 (5)	19 (29)	72 (82)	31 (44)	87 (93)	50 (62)	98 (99)
	400	3.2 (5)	48 (56)	99 (99)	70 (77)	100 (100)	89 (92)	100 (100)
0.4	100	1.5 (5)	3 (9)	11 (24)	6 (15)	23 (45)	8 (20)	34 (59)
	200	3.1 (5)	10 (15)	41 (51)	24 (32)	78 (84)	35 (44)	89 (94)
	400	4.2 (5)	24 (27)	81 (83)	57 (59)	99 (99)	72 (75)	100 (100)
ρ	n				L_1			
0	100	3.6 (5)	14 (20)	50 (60)	11 (16)	38 (47)	22 (30)	69 (77)
	200	3.0 (5)	38 (45)	95 (97)	35 (42)	91 (94)	67 (72)	100 (100)
	400	4.7 (5)	78 (79)	100 (100)	74 (74)	100 (100)	96 (97)	100 (100)
0.2	100	2.9 (5)	8 (12)	29 (40)	10 (14)	33 (43)	15 (22)	56 (66)
	200	3.7 (5)	22 (26)	78 (82)	31 (36)	86 (89)	50 (57)	98 (99)
	400	4.2 (5)	53 (55)	99 (99)	67 (70)	100 (100)	89 (90)	100 (100)
0.4	100	2.7 (5)	6 (9)	17 (24)	9 (13)	24 (35)	10 (17)	39 (51)
	200	3.2 (5)	13 (18)	47 (56)	23 (30)	73 (79)	34 (41)	89 (92)
	400	4.1 (5)	27 (30)	86 (87)	51 (54)	99 (99)	72 (74)	100 (100)

Table 5: Empirical power of the test in % with the size-corrected power in the parentheses, AR(1) errors with t_2 innovations, flat-top kernel

$\boldsymbol{\delta}$	(0, 0)	(1/2, 0)	(1, 0)	(0, 1/2)	(0, 1)	(1/2, 1/2)	(1, 1)
n				L_2			
100	1.6	1.9	2.3	2.0	3.1	2.4	4.4
200	0.9	1.8	3.3	1.8	4.0	2.4	7.6
400	1.3	2.0	3.9	2.0	4.7	2.8	8.2
n				Huber			
100	2.3	7.0	24.6	6.7	19.1	11.5	39.9
200	3.2	22.3	73.1	20.4	64.8	39.1	92.4
400	4.1	49.0	98.0	46.6	97.4	81.0	100.0
n				L_1			
100	3.8	11.6	38.9	9.5	27.6	17.4	52.8
200	4.5	30.7	88.5	27.7	78.8	52.5	97.7
400	4.8	64.6	99.9	60.6	99.6	91.5	100.0

Table 6: Empirical power of the test (in %), i.i.d. errors $\sim t_1$, flat-top kernel

$\boldsymbol{\delta}$		(0, 0)	(1/2, 0)	(1, 0)	(0, 1/2)	(0, 1)	(1/2, 1/2)	(1, 1)
ρ	n				*Huber*			
0	100	2.0	21.3	77.0	14.0	60.0	35.4	95.3
	200	3.3	69.6	99.6	62.3	100.0	94.4	100.0
	400	4.6	98.4	100.0	97.5	100.0	100.0	100.0
0.2	100	2.1	11.5	56.5	12.5	55.4	24.2	87.7
	200	2.5	47.9	98.9	60.1	100.0	88.2	100.0
	400	4.3	86.7	100.0	96.9	100.0	99.9	100.0
0.4	100	1.4	6.0	30.3	12.0	53.6	19.8	77.2
	200	2.9	25.6	88.8	54.3	99.6	74.5	100.0
	400	4.0	62.2	99.9	94.4	100.0	99.4	100.0

Table 7: Empirical power of the test (in %), AR(1) errors with the Bernoulli innovations, flat-top kernel

On-line procedures
Robust monitoring for CAPM for high-frequency portfolio betas

$$\mathbf{r}_i(s) = \boldsymbol{\alpha}_i + \boldsymbol{\beta}_i r_{i,M}(s) + \mathbf{e}_i(s), \quad i \in \mathbb{Z}, \quad s \in [0,1],$$

$\mathbf{r}_i(s) = (r_{i,1}(s), \ldots, r_{i,d}(s))^T$ – d-dimensional vector of (functional) log-returns at (say) "day" i and "intra-day time" s,

$r_{i,M}(s)$ – the log-return of the market portfolio at day i and time s,

$\mathbf{e}_i(s) = (e_{i,1}(s), \ldots, e_{i,d}(s))^T$ – d-dimensional (functional) error terms

$\boldsymbol{\alpha}_i$'s and $\boldsymbol{\beta}_i$'s are d-dimensional unknown parameters

$\boldsymbol{\beta}_i$'s are the parameters of interest, usually called the "portfolio betas"

We assume a training sample of size m with no instabilities is available, i.e.,

$$\boldsymbol{\alpha}_1 = \ldots = \boldsymbol{\alpha}_m =: \boldsymbol{\alpha}_0 = (\alpha_1^0, \ldots, \alpha_d^0)^T,$$

$$\boldsymbol{\beta}_1 = \ldots = \boldsymbol{\beta}_m =: \boldsymbol{\beta}_0 = (\beta_1^0, \ldots, \beta_d^0)^T,$$

$\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ – unknown parameters

Null hypothesis

$$H_0: \ \boldsymbol{\beta}_1 = \ldots = \boldsymbol{\beta}_m = \boldsymbol{\beta}_{m+1} = \ldots$$

of no "change versus" the alternative

$$H_A: \ \boldsymbol{\beta}_1 = \ldots = \boldsymbol{\beta}_{m+k^*} \neq \boldsymbol{\beta}_{m+k^*+1} = \ldots$$

a "structural break" at an unknown change-point $k^* = k_m^*$.

$$r_{i,j}(s) = \alpha_j^0 + \beta_j^0 r_{i,M}(s) + (\alpha_j^1 + \beta_j^1 r_{i,M}(s))\delta_m I\{i > m + k^*\} + e_{i,j}(s), \ \ j = 1, \ldots, d \tag{1}$$

Define

$$\boldsymbol{\psi}(\widehat{\mathbf{e}}_i(s_\nu)) = (\psi_1(\widehat{e}_{i,1}(s_\nu)), \dots, \psi_d(\widehat{e}_{i,d}(s_\nu)))^T$$

$$\widehat{\mathbf{e}}_i(s_\nu) = (\widehat{e}_{i,1}(s_\nu), \dots, \widehat{e}_{i,d}(s_\nu))^T,$$
$$\widehat{e}_{i,j}(s_\nu) = r_{i,j}(s_\nu) - \widehat{\alpha}_{jm} - \widehat{\beta}_{jm} r_{i,M}(s_\nu).$$

$s_\nu = \nu/n$, $\nu = 1, \dots, n$, $n = n(m)$

test statistic based on the first $m+k$ (functional) observations:

$$\widehat{Q}(k,m) = \Big(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \frac{1}{n} \sum_{\nu=1}^{n} r_{i,M}(s_\nu) \boldsymbol{\psi}(\widehat{\mathbf{e}}_i(s_\nu))\Big)^T (\widehat{\Sigma}_m)^{-1}$$
$$\Big(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \frac{1}{n} \sum_{\nu=1}^{n} r_{i,M}(s_\nu) \boldsymbol{\psi}(\widehat{\mathbf{e}}_i(s_\nu))\Big)$$

$\widehat{\Sigma}_m$ is an estimator of the asymptotic variance (matrix)

$$\Sigma = \lim_m var\Big\{ \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \int_0^1 r_{i,M}(s)\psi(\mathbf{e}_i(s))ds \Big\}$$

based on the first m observations.

The null hypothesis is rejected if

$$\widehat{Q}(k,m)/q_\gamma(k/m) \geq c$$

for properly chosen c.

For a vector-valued random variable $\mathbf{X}$ define

$$||\mathbf{X}||_p = \big(E|\mathbf{X}|^p\big)^{1/p}, \quad p \geq 1,$$

the L_p-norm of $\mathbf{X}$, where $|\mathbf{X}|$ denotes the Euclidean norm of $\mathbf{X}$.

Assumptions

(B.1) For any $i \in \mathbb{Z}$, $r_{i,M}(\cdot) = h(\boldsymbol{\xi}_i(\cdot), \boldsymbol{\xi}_{i-1}(\cdot), \ldots)$, where $h(\cdot)$ is a measurable function, $\{\boldsymbol{\xi}_i(\cdot)\}$ is a sequence of i.i.d. random functions, and $\sup_{s\in[0,1]} E|r_{i,M}(s)|^3 < \infty$.

(B.2) For any $i \in \mathbb{Z}$, $\mathbf{e}_i(\cdot) = \mathbf{g}(\boldsymbol{\zeta}_i(\cdot), \boldsymbol{\zeta}_{i-1}(\cdot), \ldots)$, where $\mathbf{g}(\cdot)$ is a measurable function, $\{\boldsymbol{\zeta}_i(\cdot)\}$ is a sequence of i.i.d. random functions having some further properties to be specified later.

(B.3) The sequences $\{\boldsymbol{\xi}_i(\cdot)\}$ and $\{\boldsymbol{\zeta}_i(\cdot)\}$ are independent.

(B.4) For all $i \in \mathbb{Z}$,

$$\sup_{s\in[0,1]} \sum_{L=1}^{\infty} \|r_{i,M}(s) - r_{iM}^{(L)}(s)\|_2 < \infty,$$

$$r_{iM}^{(L)}(\cdot) = h(\xi_i(\cdot), \xi_{i-1}(\cdot), \dots, \xi_{i-L+1}(\cdot), \xi_{i-L}^{(L)}(\cdot), \xi_{i-L-1}^{(L)}(\cdot), \dots),$$

with $\xi_{i-L}^{(L)}(\cdot), \xi_{i-L-1}^{(L)}(\cdot), \dots$ being i.i.d. with the same distribution as $\xi_0(\cdot)$ and independent of $\{\xi_i(\cdot)\}$.

(B.5) With $\boldsymbol{\psi}(\mathbf{e}_i(\cdot)) = (\psi_1(e_{i,1}(\cdot)), \dots, \psi_d(e_{i,d}(\cdot)))^T$, for all $i \in \mathbb{Z}$, it holds that

$$\sup_{s\in[0,1]} \sup_{|\mathbf{a}|\le a_0} \sum_{L=1}^{\infty} \|\boldsymbol{\psi}(\mathbf{e}_i(s) - \mathbf{a}) - \boldsymbol{\psi}(\mathbf{e}_i^{(L)}(s) - \mathbf{a})\|_2 < \infty$$

for some $a_0 > 0$, where

$$\mathbf{e}_i^{(L)}(\cdot) = \mathbf{g}(\zeta_i(\cdot), \zeta_{i-1}(\cdot), \dots, \zeta_{i-L+1}(\cdot), \zeta_{i-L}^{(L)}(\cdot), \zeta_{i-L-1}^{(L)}(\cdot), \dots),$$

with $\zeta_{i-L}^{(L)}(\cdot), \zeta_{i-L-1}^{(L)}(\cdot), \dots$ being i.i.d. with the same distribution as $\zeta_0(\cdot)$ and independent of $\{\zeta_i(\cdot)\}$.

(B.6) We let $n = n(m) \to \infty$ as $m \to \infty$.

(B.7) For all $i \in \mathbb{Z}$, $j = 1, \ldots, d$, with $s_\nu = \nu/n$ as above and $n = n(m) \to \infty$,

a)
$\lim_m (\log m) \frac{1}{n} \sum_{\nu=1}^{n} \sup_{h \in [0,1/n]} ||r_{i,M}(s_\nu) - r_{i,M}(s_\nu - h)||_2 = 0$

and

b)
$\lim_m (\log m) \frac{1}{n} \sum_{\nu=1}^{n} \sup_{h \in [0,1/n]} \|\psi_j(e_{i,j}(s_\nu)) - \psi_j(e_{i,j}(s_\nu - h))\|_2 = 0$.

Application

Sectors: Boeing (BA), Bank of America (BAC), Microsoft (MSFT), AT&T (T), and Exxon Mobile (XOM)

market portfolio, the S&P 100 index itself

The intra-day behavior of the process $\{\mathbf{r}_i(s) : s \in [0; 1]; i \in \mathbb{Z}\}$ is defined at time s as the difference between the log-prices of the stocks at time s and $s + 15$ min, is thus sampled every 15 minutes during any trading day i.

The process $r_{iM}(\cdot)$ is defined analogously.

Historical training period January 29, 2001 and consists of 120 trading days (the portfolio betas appear reasonably stable).

The monitoring horizon for the closed-end procedure was selected as 360 days, corresponding to $T = 3$. This covers the 9/11/2001 event.

shows L_2 estimates of portfolio betas based on moving windows of 10 trading days for each company throughout the historical and monitoring periods. The solid black vertical line marks the end of the historical period (120 days), whereas the dashed black line marks the last day, when the estimate is not influenced by the observations from the monitoring period. The grey lines refer in the same way to the 9/11 event.
The BAC and T estimates seem to be stable throughout the whole period, whereas there is a small temporary influence of the 9/11 event on MSFT and a very big one on BA. Finally there seems to be a shift in the portfolio beta of XOM right after the end of the training period. We come back to these observations later on.

$\widehat{Q}(k,m)/(c_{0.25}(0.05)\, q_\gamma(k/m))$, for the L_2 (dashed line), Huber (solid line) and L_1 (dotted line)

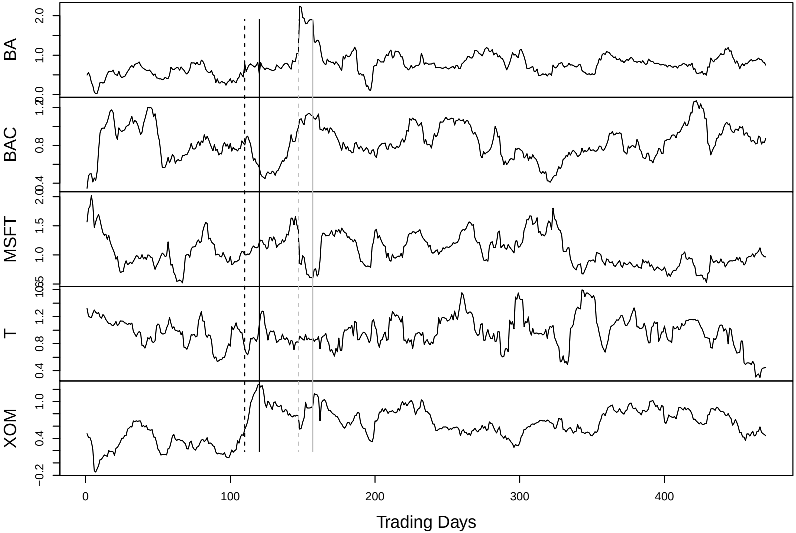
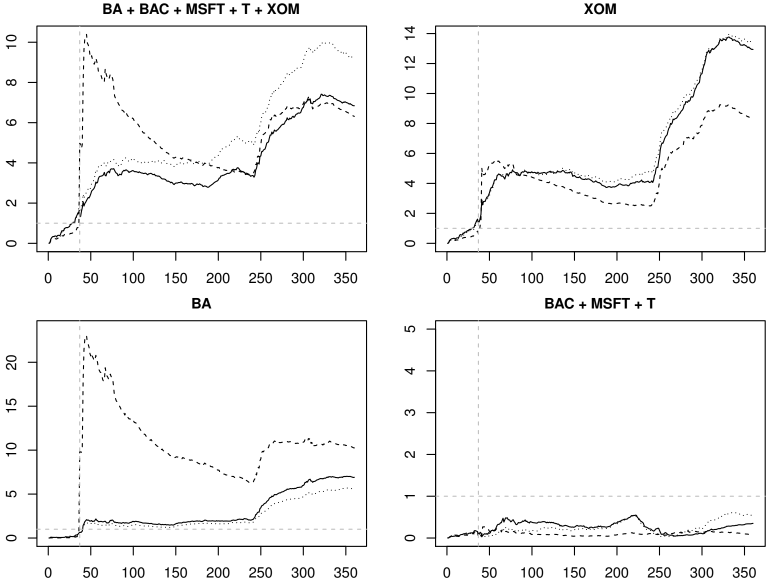


Figure: L_2 estimates of portfolio beta based on moving windows of 10


BA + BAC + MSFT + T + XOM
0
2
4
6
8
10
0 50 100 150 200 250 300 350
XOM
0
2
4
6
8
10
12
14
0 50 100 150 200 250 300 350
BA
0
5
10
15
20
0 50 100 150 200 250 300 350
BAC + MSFT + T
0
1
2
3
4
5
0 50 100 150 200 250 300 350

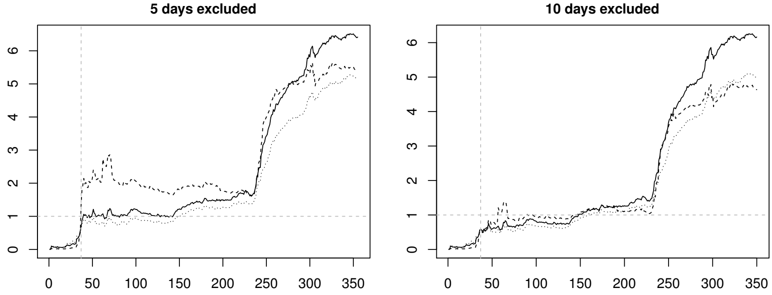


Figure: Boeing stock, normalized test statistics for L_2 (dashed line), Huber (solid line) and L_1 (dotted line) monitoring procedures. 5 or 10 days excluded from the monitoring after the 9/11.

Two contributions to change-point analysis

Marie Hušková

Outline

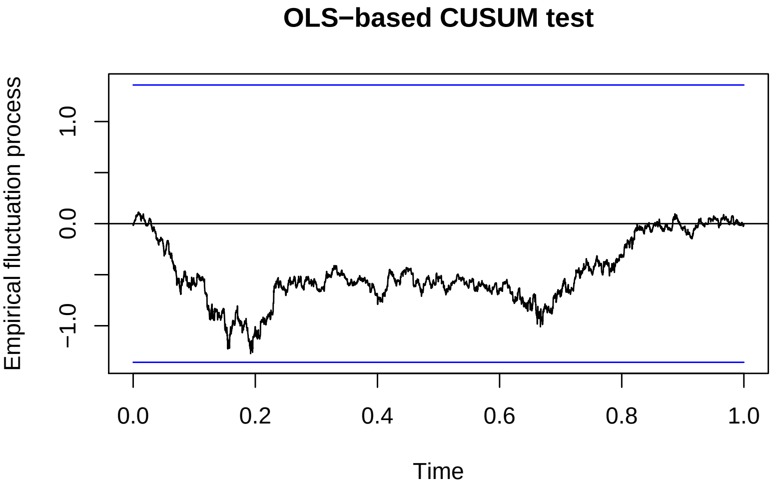
Introduction

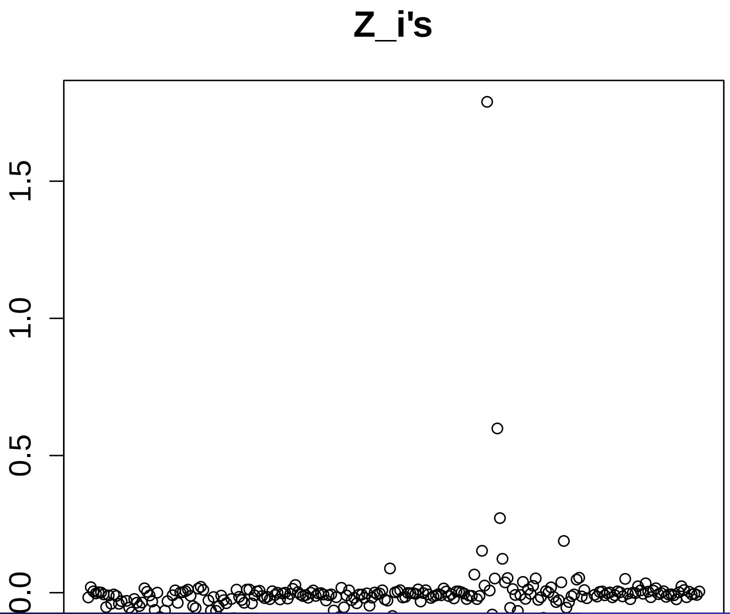
Detection of a change in regression

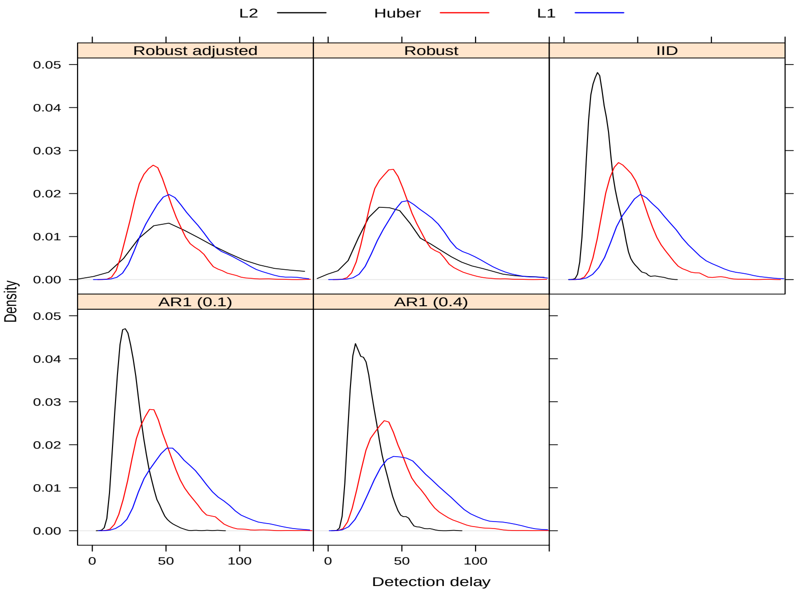
Retrospective analysis

		Hist. period	Just before 11/9	Including 11/9		After 11/9
XOM		1-120	1-156	1-240	1-480	121-480
	L2	0	111	118	118	0
	Huber	0	109	109	118	362
	L1	32	109	109	118	362
Other indicies						
	Huber					
	BA	0	0	112	151	0
	BAC	0	0	0	0	0
	MSFT	0	0	0	343	343
	T	67	67	67	0	0
	L2					
	BA	0	0	0	0	0
	BAC	0	0	0	0	0
	MSFT	0	0	0	343	343
	T	41	41	0	0	0

XOM indicates that retroepctive test based
on historical data does not detect any change.
the change occurs quite close to the end of historical period
If 156 observations a changes indicated at 111st


OLS−based CUSUM test
Empirical fluctuation process
1.0
0.0
−1.0
0.0
0.2
0.4
0.6
0.8
1.0
Time


Z_i's
1.5
1.0
0.5
0.0


L2
Huber
L1
Robust adjusted
Robust
IID
AR1 (0.1)
AR1 (0.4)
Density
0.00
0.01
0.02
0.03
0.04
0.05
0
50
100
Detection delay