Časové řady I

Ondřej Vencálek
Univerzita Palackého v Olomouci
ondrej.vencalek@upol.cz

seminář pro VŠB-TUO
2015-03-13, Ostrava

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Nové kreativní týmy v prioritách vědeckého bádání

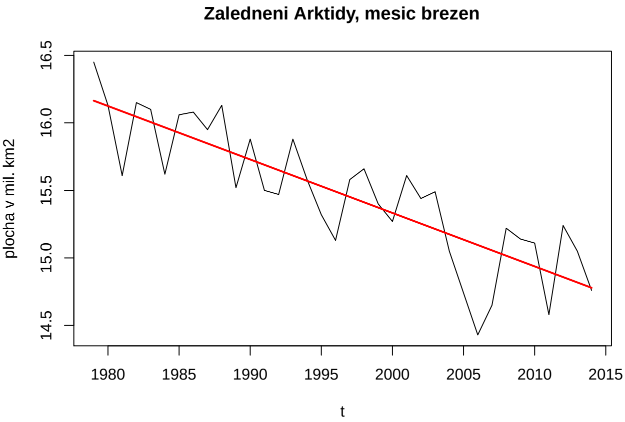
CZ.1.07/2.3.00/30.0055

Tento projekt je spolufinancován z ESF a státního rozpočtu ČR.

Dekompoziční přístup – o čem bude řeč ...

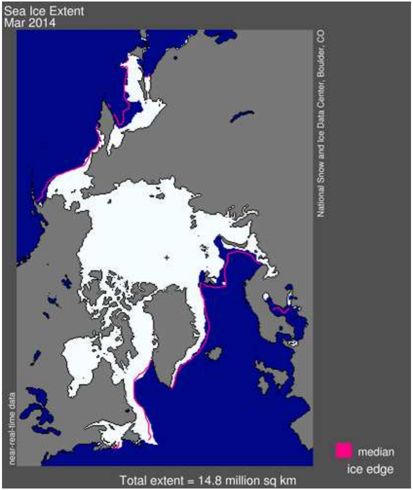
- **Trendová složka**
  - Popis trendu matematickými křivkami
    lineární, kvadratický, exponenciální, modifikovaný exponenciální, logistický, zobecněný logistický trend
  - Adaptivní přístupy k trendové složce
    – *metoda klouzavých průměrů*
    – *exponenciální vyrovnávání*
- **Sezónní složka**
  - Regresní přístup k sezónnosti
  - Holtova-Wintersova metoda
- **Náhodná složka**
  - Testy náhodnosti

Lineární trend



$$Tr(t) = \beta_0 + \beta_1 t, \qquad \text{pro } t = 1, \ldots, n$$

Příklad – trend zalednění Arktidy – obrázek



Příklad – trend zalednění Arktidy

Data: Monthly Sea Ice Extent – veřejně přístupná na stránkách National Snow & Ice Data Center `http://nsidc.org/`

`http://nsidc.org/data/docs/noaa/g02135_seaice_index/#monthly_extent_image`
`ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/Mar/N_03_area.txt`

year	mo	data_type	region	extent	area
1979	3	Goddard	N	16.45	13.13
1980	3	Goddard	N	16.13	12.92
1981	3	Goddard	N	15.61	12.62
1982	3	Goddard	N	16.15	12.99
1983	3	Goddard	N	16.10	12.84
1984	3	Goddard	N	15.62	12.48
1985	3	Goddard	N	16.06	12.66
1986	3	Goddard	N	16.08	12.65
1987	3	Goddard	N	15.95	12.75
1988	3	Goddard	N	16.13	13.84
1989	3	Goddard	N	15.52	13.14
1990	3	Goddard	N	15.88	13.44
1991	3	Goddard	N	15.50	13.35
1992	3	Goddard	N	15.47	13.41
1993	3	Goddard	N	15.88	13.71
1994	3	Goddard	N	15.58	13.47
1995	3	Goddard	N	15.32	13.28
1996	3	Goddard	N	15.13	12.83
1997	3	Goddard	N	15.58	13.24
1998	3	Goddard	N	15.66	13.50
1999	3	Goddard	N	15.40	13.47
2000	3	Goddard	N	15.27	13.10
2001	3	Goddard	N	15.61	13.57
2002	3	Goddard	N	15.44	13.36
2003	3	Goddard	N	15.49	13.36
2004	3	Goddard	N	15.05	12.93
2005	3	Goddard	N	14.74	12.67
2006	3	Goddard	N	14.43	12.44
2007	3	Goddard	N	14.65	12.49
2008	3	Goddard	N	15.22	13.20
2009	3	Goddard	N	15.14	13.08
2010	3	Goddard	N	15.11	13.19
2011	3	Goddard	N	14.58	12.48
2012	3	Goddard	N	15.24	13.08
2013	3	Goddard	N	15.05	13.10
2014	3	NRTSI-G	N	14.76	12.52

Lineární trend – trocha teorie

$$\begin{aligned} Y(t) &= Tr(t) + E(t), \qquad \text{pro } t = 1, \dots, n \\ Tr(t) &= \beta_0 + \beta_1 t \\ E(t) &\sim N(0, \sigma^2), \text{ nezávislé} \end{aligned}$$

Odhad (vektorového) parametru $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ metodou nejmenších čtverců, tj. minimalizujeme $\sum_{t=1}^{n}(Y(t) - (\beta_0 + \beta_1 t))^2$

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{pmatrix}$$

$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$,
$s^2 = \frac{1}{n-2}\sum_{t=1}^{n}(Y(t) - b_0 - b_1 t)^2$
Vlastnosti odhadu:

- $E\boldsymbol{b} = \boldsymbol{\beta}$
- $Var\boldsymbol{b} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$
- $\boldsymbol{b}$ má normální rozdělení

Predikce hodnoty $Y(T)$

- Bodový odhad:

$$\begin{aligned} Y(T) &= Tr(T) + E(T) = \beta_0 + \beta_1 T + E(T) \\ \widehat{Y}(T) &= \widehat{Tr}(T) + \widehat{E}(T) \\ \widehat{Y}(T) &= b_0 + b_1 T \end{aligned}$$

- Intervalový odhad:
  $Y(T) - (b_0 + b_1 T) = (\beta_0 + \beta_1 T + E(T)) - (b_0 + b_1 T) \sim N\left(0, \sigma^2 + (1, T)Var(\boldsymbol{b})(1, T)'\right)$

$$b_0 + b_1 T \pm t_{n-2}\left(1 - \frac{\alpha}{2}\right) s\sqrt{1 + (1, T)(\boldsymbol{X}'\boldsymbol{X})^{-1}(1, T)'},$$

  kde $t_{n-2}\left(1 - \frac{\alpha}{2}\right)$ je $(1 - \frac{\alpha}{2})$-kvantil Studentova rozdělení o $n-2$ stupních volnosti (viz následující poznámka).

Připomenutí: pojem *kvantil*

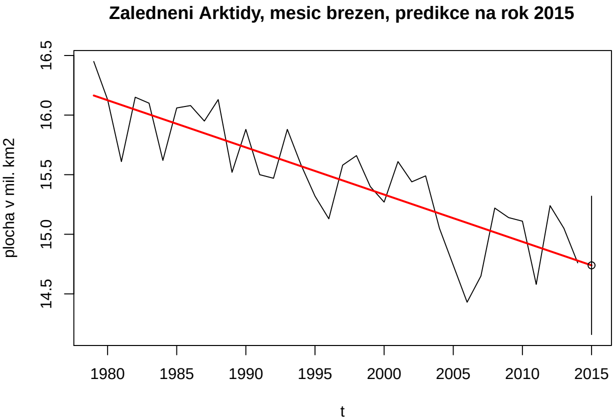
(pro spojitá rozdělení)

Máme dáno

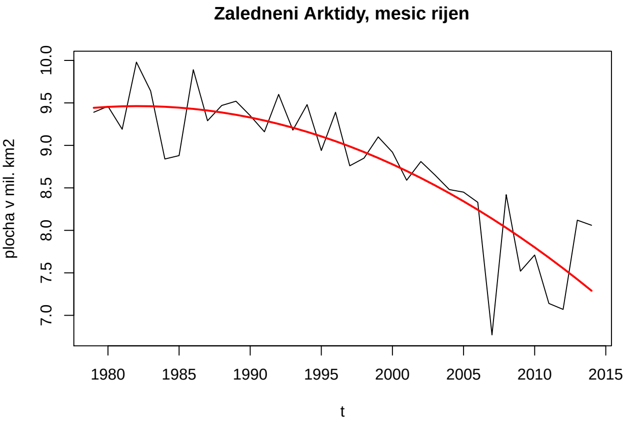
- rozdělení náhodné veličiny X
  (např. Studentovo rozdělení o 34 stupních volnosti)
- $\alpha \in (0, 1)$

Ptáme se, pro jaké x platí $P(X < x) = \alpha$.
Tuto hodnotu nazýváme α-kvantilem daného rozdělení.

Lineární trend: predikce

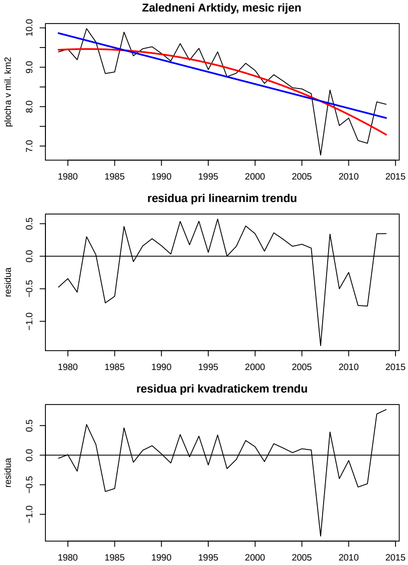


Kvadratický trend



$$Tr(t) = \beta_0 + \beta_1 t + \beta_2 t^2, \qquad \text{pro } t = 1, \ldots, n$$

Lineární nebo kvadratický trend?



Lineární nebo kvadratický trend?

$$Tr(t) = \beta_0 + \beta_1 t + \beta_2 t^2,$$

Testujeme hypotézu H: $\beta_2 = 0$ (proti oboustranné alternativě)

```
> model2 = lm(y~t+I(t^2))
> summary(model2)
...
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.366e+03  2.872e+03  -2.913  0.00638 **
t            8.451e+00  2.877e+00   2.937  0.00599 **
I(t^2)      -2.132e-03  7.205e-04  -2.959  0.00567 **
```

$P - value = 0.00567 < 0.05$, tedy zamítáme hypotézu nulovosti paremtru β_2, volíme kvadratický trend.

Kvadratický trend – trocha teorie

$$
\begin{aligned}
Y(t) &= Tr(t) + E(t), \qquad \text{pro } t = 1, \ldots, n \\
Tr(t) &= \beta_0 + \beta_1 t + \beta_2 t^2 \\
E(t) &\sim N(0, \sigma^2), \ \text{nezávislé}
\end{aligned}
$$

Odhad (vektorového) parametru $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ metodou nejmenších čtverců, tj. minimalizujeme $\sum_{t=1}^{n}(Y(t) - (\beta_0 + \beta_1 t + \beta_2 t^2))^2$

$$
\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{pmatrix}
$$

$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$,
$s^2 = \frac{1}{n-3}\sum_{t=1}^{n}(Y(t) - b_0 - b_1 t - b_2 t^2)^2$
Vlastnosti odhadu:

- $E\boldsymbol{b} = \boldsymbol{\beta}$
- $Var\boldsymbol{b} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$
- $\boldsymbol{b}$ má normální rozdělení

Predikce hodnoty $Y(T)$

- Bodový odhad:

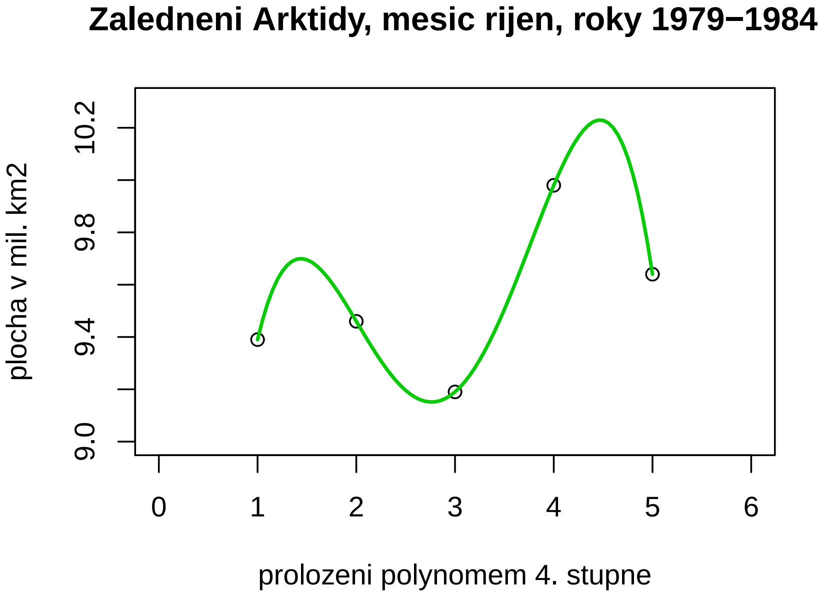
$$\begin{aligned} Y(T) &= Tr(T) + E(T) = \beta_0 + \beta_1 T + \beta_2 T^2 + E(T) \\ \widehat{Y}(T) &= \widehat{Tr}(T) + \widehat{E}(T) \\ \widehat{Y}(T) &= b_0 + b_1 T + b_2 T^2 \end{aligned}$$

- Intervalový odhad: $Y(T) - (b_0 + b_1 T + b_2 T^2) \sim N\left(0, \sigma^2 + (1, T, T^2) Var(\boldsymbol{b})(1, T, T^2)'\right)$

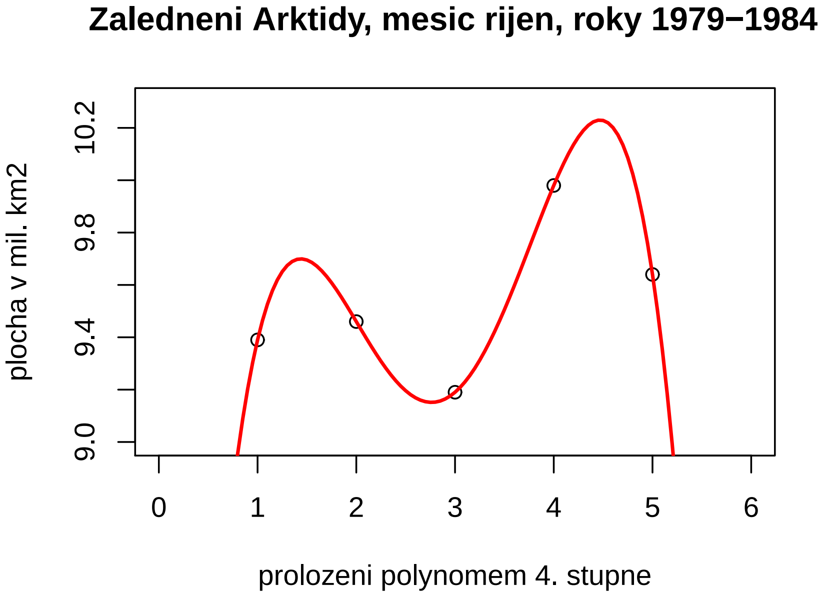
$$b_0 + b_1 T + b_2 T^2 \pm t_{n-3}\left(1 - \frac{\alpha}{2}\right) s \sqrt{1 + (1, T, T^2)(\boldsymbol{X}'\boldsymbol{X})^{-1}(1, T, T^2)'},$$

kde $t_{n-3}\left(1 - \frac{\alpha}{2}\right)$ je $(1 - \frac{\alpha}{2})$-kvantil Studentova rozdělení o $n - 3$ stupních volnosti.

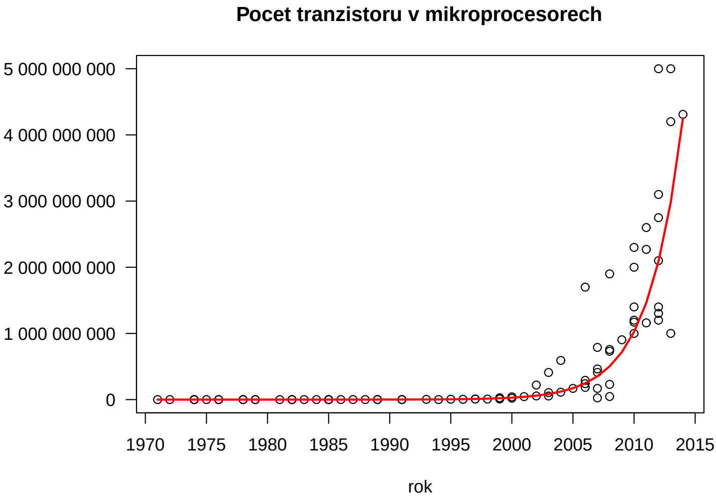
Varování před polynomy vyššího stupně



Varování před polynomy vyššího stupně



Exponenciální trend



$$Tr(t) = \alpha \cdot \beta^t, \qquad \text{pro } t = 1, \ldots, n$$

Exponenciální trend – příklad (Mooreův zákon)

Moorův zákon je empirické pravidlo, které roku 1965 vyslovil chemik a spoluzakladatel firmy Intel Gordon Moore. Původní znění bylo: „*počet tranzistorů, které mohou být umístěny na integrovaný obvod se při zachování stejné ceny zhruba každých 18 měsíců zdvojnásobí.*“ Takovýto růst se nazývá exponenciální.

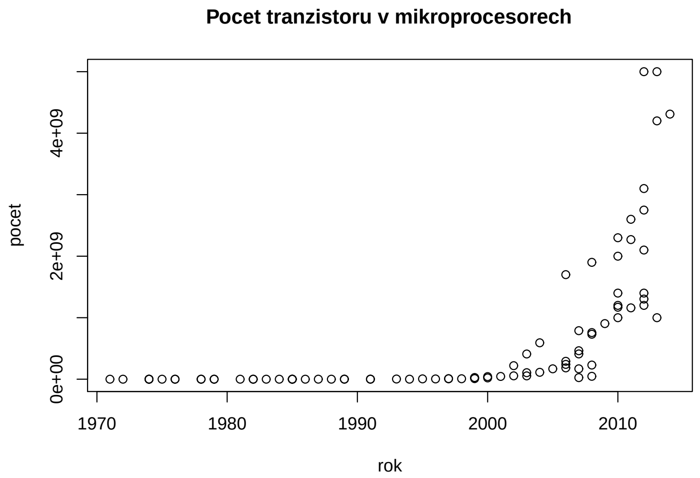
Složitost dnešních procesorů se poměřuje především počtem tranzistorů v nich zapojených. Rychlost růstu počtu tranzistorů na plošné jednotce se časem zpomalila a nyní se jejich **počet zdvojnásobuje přibližně jednou za dva roky**.

`http://cs.wikipedia.org/wiki/Mooreův_zákon`

Mooreův zákon – data

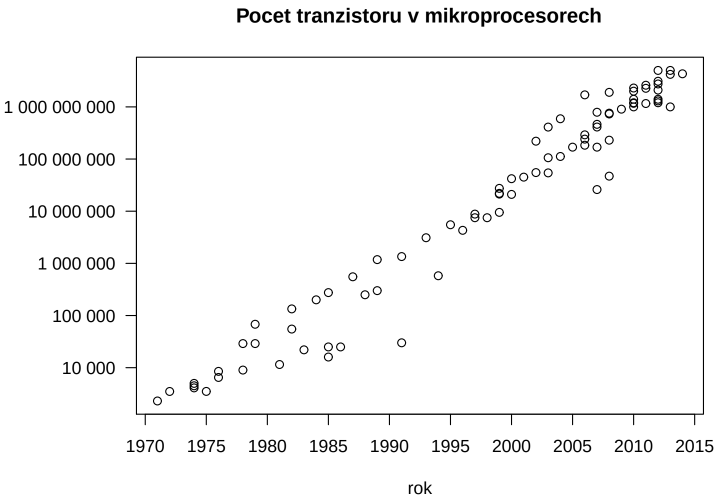
	Processor	Transistor.count	Date.of.introduct.
1	Intel 4004	2300	1971
2	Intel 8008	3500	1972
3	MOS Technology 6502	3510.00	1975
4	Motorola 6800	4100	1974
5	Intel 8080	4500	1974
6	RCA 1802	5000	1974
7	Intel 8085	6500	1976
8	Zilog Z80	8500	1976
9	Motorola 6809	9000	1978
10	Intel 8086	29000	1978
⋮	⋮	⋮	⋮
80	12-core POWER8	4200000000	2013
81	15-core Xeon Ivy Bridge-EX	4310000000	2014
82	62-core Xeon Phi	5000000000	2012
83	Apple A7 ("mobile SoC")	1000000000	2013
84	Xbox One main SoC	5000000000	2013

Graf I



Pocet tranzistoru v mikroprocesorech
pocet
0e+00
2e+09
4e+09
1970
1980
1990
2000
2010
rok

Graf II – po logaritmické transformaci



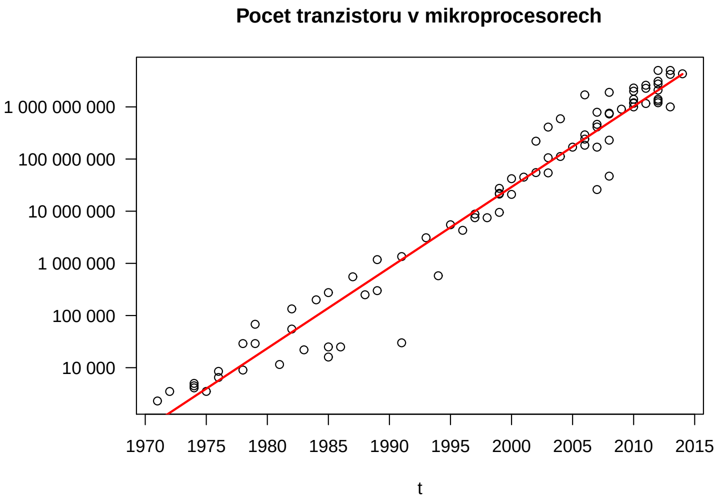
Exponenciální trend

$$
\begin{aligned}
Tr(t) &= \alpha \cdot \beta^{t} \\
\log(Tr(t)) &= \log(\alpha) + t\log(\beta) \\
\log(Tr(t)) &= \alpha_0 + t\beta_0
\end{aligned}
$$

Při označení

- $\alpha_0 := \log(\alpha)$,
- $\beta_0 := \log(\beta)$.

Lineární chování logaritmované veličiny



Jaký logaritmus mám použít? (To je jedno!)

$$
\begin{aligned}
\log_2(Tr(t)) &= \alpha_0 + \beta_0 t \\
&: \log_{10} 2 \\
\log_{10}(Tr(t)) &= \alpha_0^* + \beta_0^* t \\
&: \ln 10 \\
\ln(Tr(t)) &= \alpha_0^{**} + \beta_0^{**} t
\end{aligned}
$$

Odlogaritmování

- Při použití logaritmu o základu 2:

$$\begin{aligned} \log_2(Tr(t)) &= \alpha_0 + \beta_0 t \\ Tr(t) &= 2^{\alpha_0+\beta_0 t} = 2^{\alpha_0} \cdot (2^{\beta_0})^t = \alpha \cdot \beta^t \\ \text{tedy} \quad & \alpha = 2^{\alpha_0}, \beta = 2^{\beta_0} \end{aligned}$$

- Při použití logaritmu o základu 10:

$$\begin{aligned} \log_{10}(Tr(t)) &= \alpha_0^* + \beta_0^* t \\ Tr(t) &= 10^{\alpha_0^*+\beta_0^* t} = 10^{\alpha_0^*} \cdot (10^{\beta_0^*})^t = \alpha \cdot \beta^t \\ \text{tedy} \quad & \alpha = 10^{\alpha_0^*}, \beta = 10^{\beta_0^*} \end{aligned}$$

- Při použití logaritmu o základu e:

$$\begin{aligned} \ln(Tr(t)) &= \alpha_0^{**} + \beta_0^{**} t \\ Tr(t) &= e^{\alpha_0^{**}+\beta_0^{**} t} = e^{\alpha_0^{**}} \cdot (e^{\beta_0^{**}})^t = \alpha \cdot \beta^t \\ \text{tedy} \quad & \alpha = e^{\alpha_0^{**}}, \beta = e^{\beta_0^{**}} \end{aligned}$$

Význam parametru β

$$
\begin{aligned}
Tr(t) &= \alpha \cdot \beta^t \\
Tr(t+1) &= \alpha \cdot \beta^{t+1} \\
\frac{Tr(t+1)}{Tr(t)} &= \frac{\alpha \cdot \beta^{t+1}}{\alpha \cdot \beta^t} = \beta
\end{aligned}
$$

Hodnota sledované veličiny se za jednotku času β-násobí.
(podíl je konstantní v čase – tempo růstu je stálé)

Za jak dlouho se počet tranzistorů zdvojnásobí?

Hledáme takovou hodnotu Δ, pro kterou platí

$$\frac{Tr(t+\Delta)}{Tr(t)} = 2$$

Upravme

$$\frac{Tr(t+\Delta)}{Tr(t)} = \frac{\alpha\beta^{t+\Delta}}{\alpha\beta^{t}} = \beta^{\Delta} = 2$$

Řešme (logaritmováním)

$$\Delta \log_2 \beta = 1$$

Odtud

$$\Delta = 1/\log_2 \beta$$

Exponenciální růst počtu tranzistorů v mikroprocesorech

$$Tr(t) = 668,5 \cdot 1,428^{t},$$

kde za t dosazujeme $1, \ldots, 44$ pro roky $1971, \ldots, 2014$.

$$\Delta = 1,95$$

Naše analýza potvrzuje zdvojnásobení počtu tranzistorů za dobu přibližně 2 let.

Omezení modelu

Dne 13. září 1995 Gordon Moore v rozhovoru uvedl, že toto (jeho) pravidlo nemůže fungovat v neomezeném měřítku: „*Tak to nemůže pokračovat napořád. Povaha exponenciál je taková, že je tlačíme mimo limity a následně nastane pohroma*“. Také poznamenal, že tranzistory eventuálně dosáhnou limitu miniaturizace (zmenšování) na atomární úrovni ...

`http://cs.wikipedia.org/wiki/Mooreův_zákon`

Exponenciální růst – šachová legenda

Sissa ben Dahir

$$1 + 2 + 4 + 8 + \ldots$$

$$1 + 2^1 + 2^2 + 2^3 + \ldots + 2^{63} = 2^{64} - 1 \doteq 18\,000\,000\,000\,000\,000\,000$$

Exponenciální růst – šachová legenda

Sissa ben Dahir

$$1 + 2 + 4 + 8 + \ldots$$

$$1 + 2^1 + 2^2 + 2^3 + \ldots + 2^{63} = 2^{64} - 1 \doteq 18\,000\,000\,000\,000\,000\,000$$

Poznámka k odhadu parametrů

Místo minimalizace výrazu

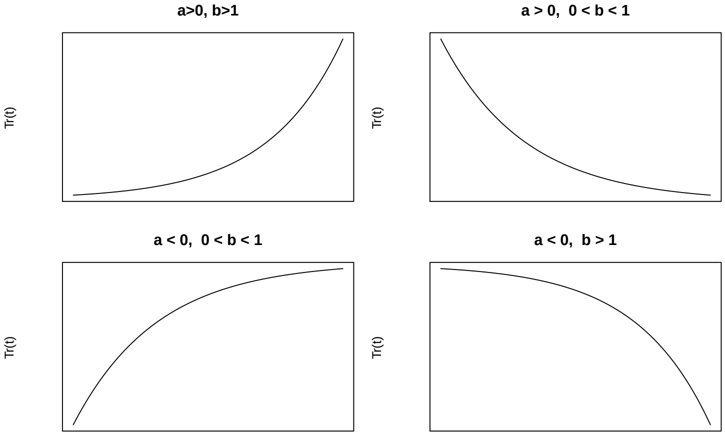
$$\sum_t (\log y_t - \log \alpha - t \log \beta)^2$$

se někdy doporučuje uvažovat metodu nejmenších vážených čtverců, tj. minimalizujeme

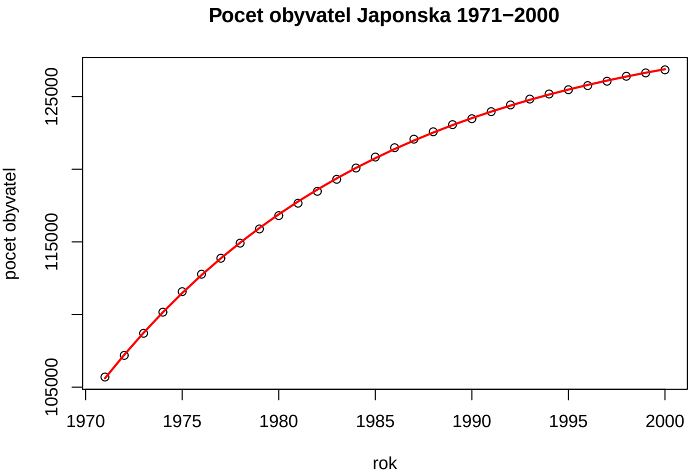
$$\sum_t w_t (\log y_t - \log \alpha - t \log \beta)^2$$

Typicky se volí $w_t = y_t^2$.

$\alpha > 0, \beta > 1 \ldots$



Modifikovaný exponenciální trend



$$Tr(t) = \gamma + \alpha \cdot \beta^t, \qquad \text{pro } t = 1, \ldots, n$$

Odhad parametrů

Logaritmování nám nepomůže.
Možná řešení:

- metoda postupných součtů
- metoda vybraných bodů
- metody nelineární regrese (doporučeno, R-ko: `nls()`)
- ...

Metoda postupných součtů

Rozdělíme řadu na tři stejně dlouhé úseky délky m (pokud počet pozorování není dělitelný třemi, jedno nebo dvě první vynecháme)

$$\begin{aligned}\sum_{t=1}^{m} y_t &\sim \sum_{t=1}^{m} Tr(t) = (\gamma + \alpha\beta^1) + (\gamma + \alpha\beta^2) + \ldots + (\gamma + \alpha\beta^m) = \\ &= \sum_{t=1}^{m} (\gamma + \alpha\beta^t) = m\gamma + \alpha \sum_{t=1}^{m} \beta^t = m\gamma + \alpha \frac{\beta(\beta^m - 1)}{\beta - 1}\end{aligned}$$

Podobně

$$\begin{aligned}\sum_{t=m+1}^{2m} y_t &\sim m\gamma + \alpha \frac{\beta^{m+1}(\beta^m - 1)}{\beta - 1} \\ \sum_{t=2m+1}^{3m} y_t &\sim m\gamma + \alpha \frac{\beta^{2m+1}(\beta^m - 1)}{\beta - 1}\end{aligned}$$

Metoda postupných součtů

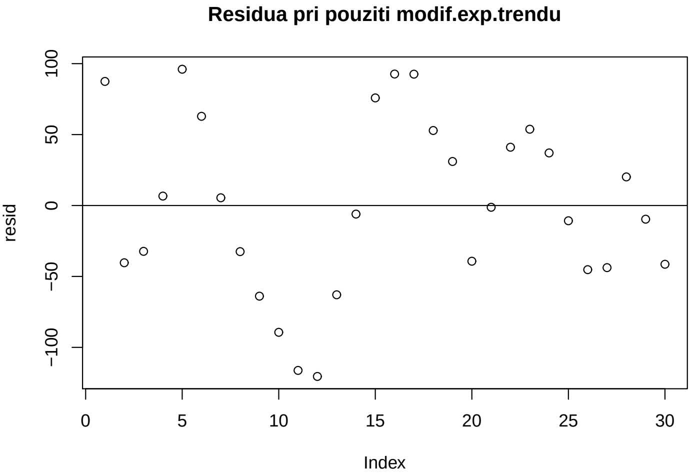
$$
\begin{aligned}
\widehat{\beta} &= \left(\frac{\sum_3 y_t - \sum_2 y_t}{\sum_2 y_t - \sum_1 y_t} \right)^{1/m} \\
\widehat{\alpha} &= \frac{\widehat{\beta} - 1}{\widehat{\beta}(\widehat{\beta}^m - 1)^2} \left(\sum_2 y_t - \sum_1 y_t\right) \\
\widehat{\gamma} &= \frac{1}{m} \left[\sum_1 y_t - \widehat{\alpha}\widehat{\beta}(\widehat{\beta}^m - 1)/(\widehat{\beta} - 1) \right]
\end{aligned}
$$

Nelineární regrese - metoda nejmenších vážených čtverců

R: `nls()`
Nutno zadat počáteční odhady parametrů.
K jejich získání využijeme interpretace parametrů:

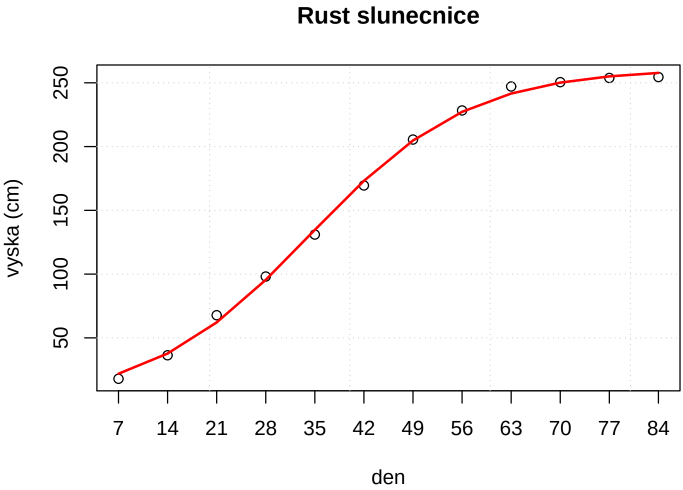
- $\gamma = \lim_{t\to\infty} \gamma + \alpha\beta^t = \lim_{t\to\infty} Tr(t)$ neboť $(0 < beta < 1)$
- $\alpha = \gamma - Tr(0)$
- β : průměrný podíl $\frac{Tr(t+1)-\gamma}{Tr(t)-\gamma}$

Residua



cyklická složka (proměnlivé délky i amplitudy)

Logistický trend



$$Tr(t) = \frac{\alpha}{1 + e^{-\gamma(t-\delta)}}, \qquad \text{pro } t = 1, \dots, n$$

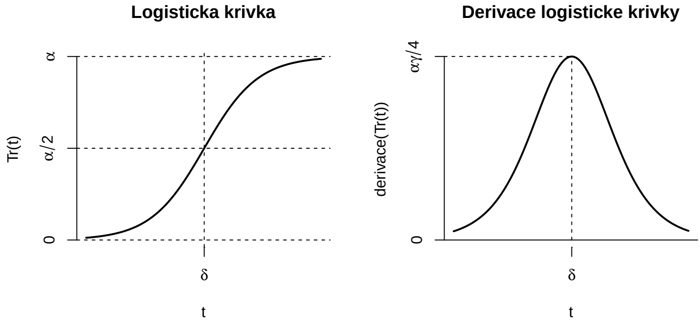
Logistický trend – další možné parametrizace

$$
\begin{aligned}
Tr(t) &= \frac{\alpha}{1 + e^{-\gamma(t-\delta)}} \\
Tr(t) &= \frac{\alpha}{1 + e^{\delta^* + \gamma^* t}} \\
Tr(t) &= \frac{\alpha}{1 + \delta^{**} e^{\gamma^* t}} \\
Tr(t) &= \frac{\alpha}{1 + \delta^{**} \gamma^{**t}}
\end{aligned}
$$

Při označení

- $\gamma^* = -\gamma$
- $\delta^* = \gamma\delta$
- $\gamma^{**} = e^{\gamma^*}$
- $\delta^{**} = e^{\delta^*}$

Logistický trend – interpretace parametrů



- α … asymptota
- δ … čas nejrychlejšího růstu
- γ … konstanta určující měřítko x-ové osy

Logistický trend – odvození

$$\frac{\partial Tr(t)}{\partial t} \propto Tr(t) \cdot [\alpha - Tr(t)]$$

Více viz *Vencálek, O: Stochastické modelování epidemií (DP)*

`https://is.cuni.cz/webapps/zzp/detail/44266/`

Zobecněný logistický trend – Richardsova křivka

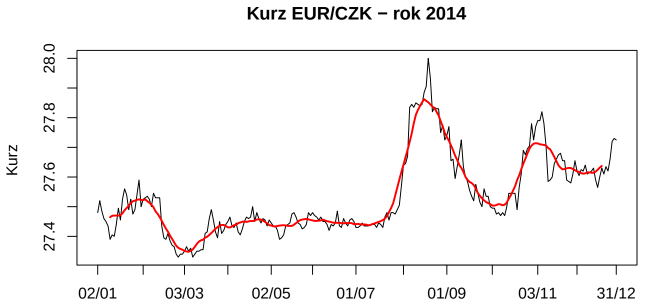
$$Tr(t) = \alpha \left[1 - (1-\omega)e^{-\gamma(t-\delta)}\right]^{\frac{1}{1-\omega}}$$

Rychlost růstu:

- $\omega < 2$ ... na počátku rychlý vzestup, pozvolný pokles na konci
- $\omega = 2$ ... stejně rychlý vzestup jako pokles (logistická křivka)
- $\omega > 2$ ... na počátku pozvolný nárůst, rychlý pokles na konci

Terminologická poznámka:
$\omega \to 1$... Gomperzova křivka, $\omega = 0$... von Bertalanffyho křivka
Růstové křivky, S-křivky, sigmoidální křivky

Metoda klouzavých průměrů



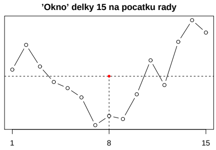
$$Tr(t) = \beta_0(t) + \beta_1(t)\tau + \ldots + \beta_r(t)\tau^r, \qquad \text{pro } \tau = t-m, \ldots, t+m$$

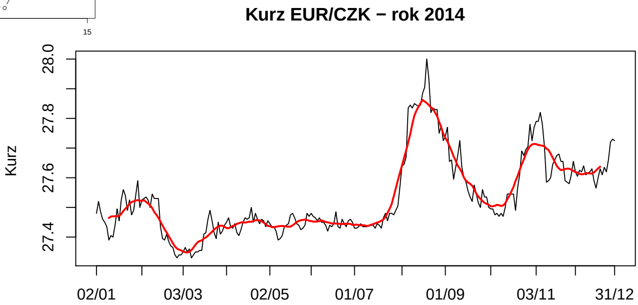
Kurz EUR/CZK

Data:

https://www.cnb.cz/cs/financni_trhy/devizovy_trh/kurzy_devizoveho_trhu/vybrane_form.jsp

Jednoduché (prosté) klouzavé průměry při délce okna 15

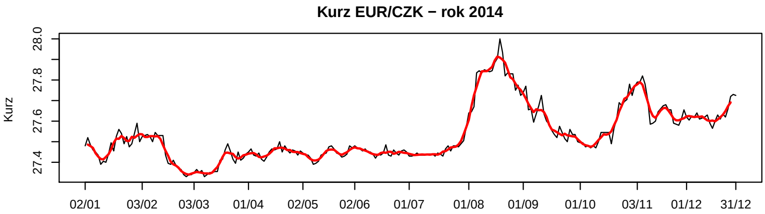


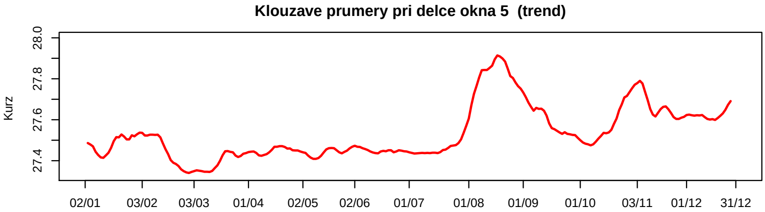


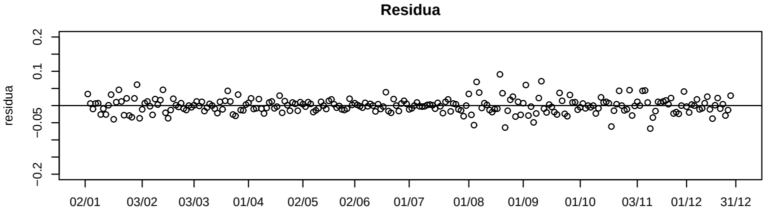
$m = 7$, tj. délka okna $2m + 1 = 15$,
$r = 0$, tj. lokálně konstantní trend,
pak $\widehat{y}_t = \frac{1}{2m+1} \sum_{\tau=-m}^{m} y_{t+\tau}$ pro $t = m + 1, \ldots, n - m$.

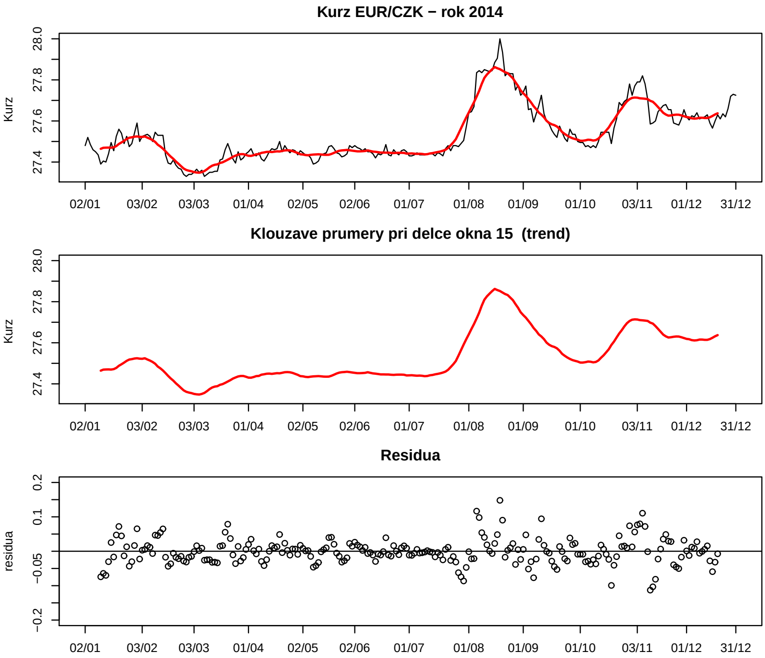
Otázky k zodpovězení

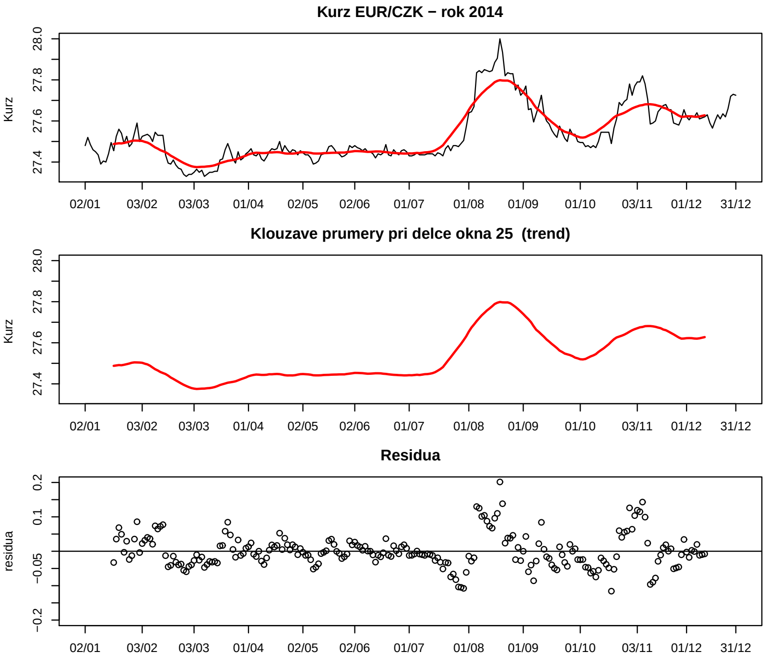
- Jak volit délku okna, tj. jak volit m? (délka okna je $2m+1$)
- Může být délka okna sudá?
- Jak volit řád polynomu, kterým řadu lokálně prokládáme, tj. jak volit r?
- Jak se vypořádat s počátečním a s koncovým úsekem řady a jak postupovat při predikci?


Kurz EUR/CZK − rok 2014
Kurz
27.4
27.6
27.8
28.0
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12


Klouzave prumery pri delce okna 5 (trend)
Kurz
27.4
27.6
27.8
28.0
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12

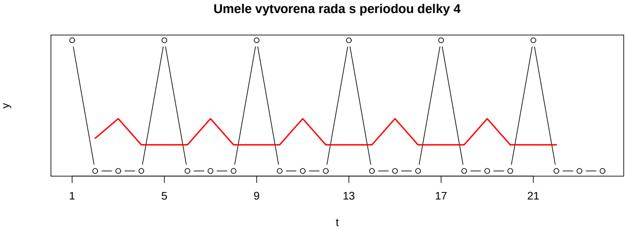

Residua
residua
−0.2
−0.05
0.1
0.2
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12


Kurz EUR/CZK − rok 2014
Kurz
28.0
27.8
27.6
27.4
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12
Klouzave prumery pri delce okna 15 (trend)
Kurz
28.0
27.8
27.6
27.4
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12
Residua
residua
0.2
0.1
−0.05
−0.2
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12



Kurz EUR/CZK − rok 2014
Kurz
28.0
27.8
27.6
27.4
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12
Klouzave prumery pri delce okna 25 (trend)
Kurz
28.0
27.8
27.6
27.4
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12
Residua
residua
0.2
0.1
−0.05
−0.2
02/01
03/02
03/03
01/04
02/05
02/06
01/07
01/08
01/09
01/10
03/11
01/12
31/12

Kdy potřebujeme sudou délku okna?



Obrázek: vyrovnání časové řady s *konstantním* trendem a s délkou periody 4 pomocí klouz. průměrů při délce okna 5.

- Sudá délka periody $\rightarrow$ sudá délka okna
- Centrované klouzavé průměry

Centrované klouzavé průměry

- Střed „prvního“ okna délky $2m$ je v čase $(1 + 2m)/2 = m + 0.5$,
- tj. $\widehat{y}_{m+0.5} = (y_1 + y_2 + \ldots + y_{2m})/(2m)$,
- Střed „druhého“ okna délky $2m$ je v čase $(2 + 2m + 1)/2 = m + 1.5$,
- tj. $\widehat{y}_{m+1.5} = (y_2 + \ldots + y_{2m} + y_{2m+1})/(2m)$,
- $\widehat{y}_{m+1} = (\widehat{y}_{m+0.5} + \widehat{y}_{m+1.5})/2 =$ $\frac{1}{2m} y_1 + \frac{1}{m} y_2 + \frac{1}{m} y_3 + \ldots + \frac{1}{m} y_{2m} + \frac{1}{2m} y_{2m+1}$.

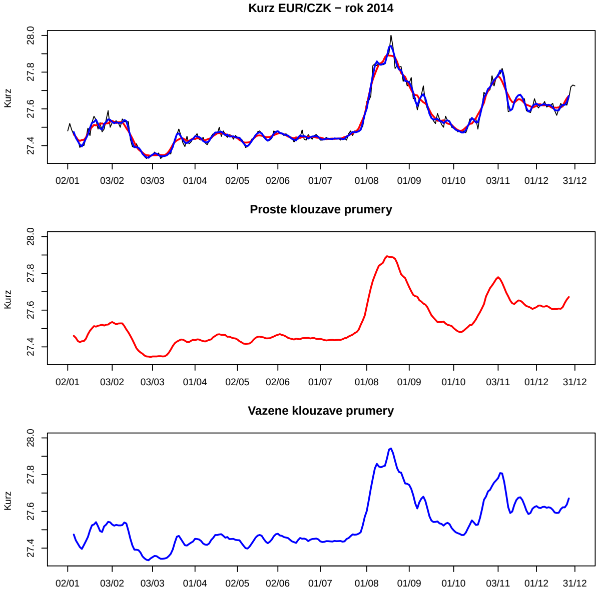
Vážené klouzavé průměry

Dá se ukázat, že proložení polynomu vyššího stupně ($r > 1$) je ekvivalentní výpočtu váženého průměru pozorování v okně. Váhy, které přitom dáváme jednotlivým pozorováním, je třeba dopočítat.

Příklad:
$m = 3, r = 2$, tj. prokládáme parabolu (polynom 2. studně) v okně délky $2m + 1 = 7$. Pak

$$\widehat{y}_t = \frac{-2}{21} y_{t-3} + \frac{3}{21} y_{t-2} + \frac{6}{21} y_{t-1} + \frac{7}{21} y_t + \frac{6}{21} y_{t+1} + \frac{3}{21} y_{t+2} + \frac{-2}{21} y_{t+3}$$

Porovnání prostých a vážených průměrů při stejné délce okna 7



Jak se vypořádat s počátečním a s koncovým úsekem řady a jak postupovat při predikci?

„Zajedeme“ s oknem na začátek či na konec řady a tam proložíme příslušný polynom. (jako bychom zapomněli na všechna další porozování kromě posledních (prvních) $2m+1$, na nichž prokládáme příslušný polynom, který pak můžeme použít k predikci.

Pozor: čím vyšší stupeň polynomu, tím blíže jsou vyhlazené hodnoty skutečně pozorovaným hodnotám, ale pozor, predikce mohou být dosti nestabilní (viz varování týkající se predikce pomocí polynomu).

Volba řádu použitého polynomu r

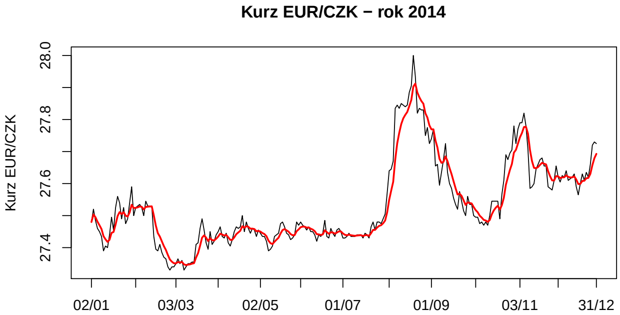
- *Cipra* popisuje „objektivní kritérium volby řádu klouzavých průměrů“ (str. 49-51),
- dodává však: „V poslední době se začínají také hledat numericky jednoduché metody, které by určily řád klouzavých průměrů jako hodnotu minimalizující vhodně zkonstruované kritérium“.

Poučení:

- Zkusme různé hodnoty pro r (řád polynomu): $r = 0, 1, 2, 3$ a dívejme se, jak dobré budeme mít predikce. To můžeme zhodnotit např. pomocí
  - střední čtvercové chyby (MSE = Mean Squared Error) $\sum_{t=1}^{n}(y_t - \widehat{y}_t)^2/n$,
  - střední absolutní odchylky (MAD = Mean Absolute Deviation) $\sum_{t=1}^{n}|y_t - \widehat{y}_t|/n$,

  či jiné „ztrátové funkce“ odrážející „cenu chyby v předpovědi“. Vyberemu tu hodnotu r, pro kterou jsou predikce nejlepší (mají nejnižší hodnotu některé z výše uvedených ztrátových funkcí).

Exponenciální vyrovnávání



$$Tr(t) = \beta_0(t) \text{ JEDNODUCHÉ EXP.VYROV.}$$
$$Tr(t) = \beta_0(t) + \beta_1(t)\tau \quad \text{pro } \tau = -1, -2, \ldots \text{ DVOJITÉ EXP.V.}$$

Exponenciální vyrovnávání – určení parametrů

Mějme dán pevně

- čas t, v němž se snažíme odhadnout hodnotu (lokál.) trendu,
- vyrovnávací konstantu $\alpha \in (0, 1)$.

Parametry $\beta_0(t), \beta_1(t)$ popisující lokální trend určíme z rovnic

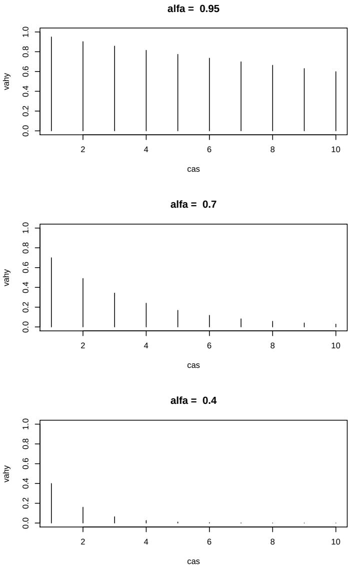
- Jednoduché exp. vyrovnávání

$$\min_{\beta_0(t)} \sum_{\tau=0}^{\infty} [y_{t-\tau} - \beta_0(t)]^2 \alpha^{\tau}$$

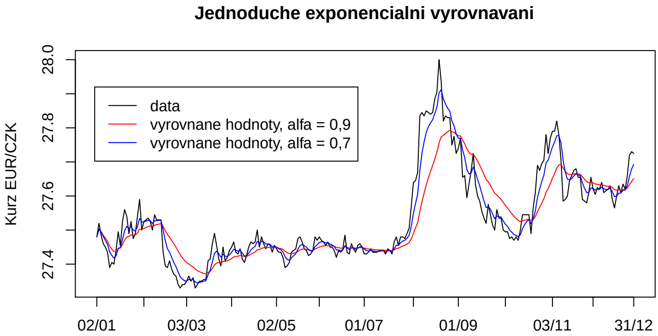
- Dvojité exp. vyrovnávání

$$\min_{\beta_0(t), \beta_1(t)} \sum_{\tau=0}^{\infty} \left[y_{t-\tau} - (\beta_0(t) + \beta_1(t)(-\tau))\right]^2 \alpha^{\tau}$$

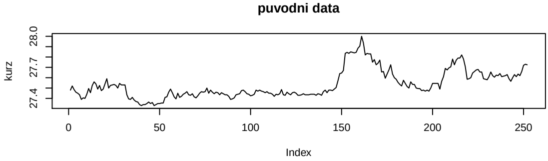
Exponenciální vyrovnávání – váhy

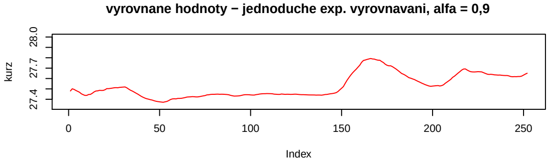


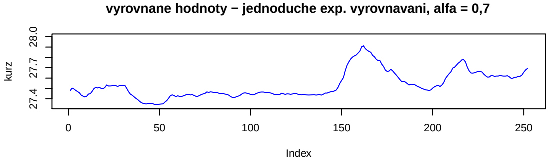
Jednoduché exponenciální vyrovnávání



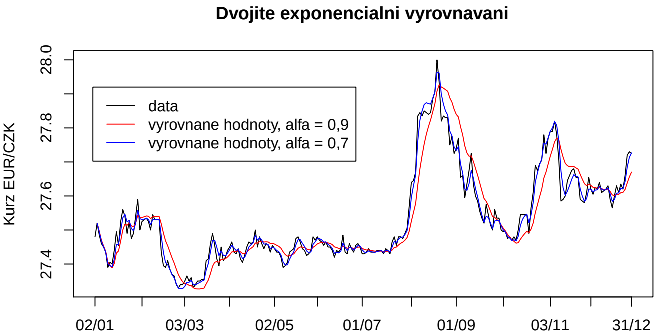
Jednoduché exponenciální vyrovnávání



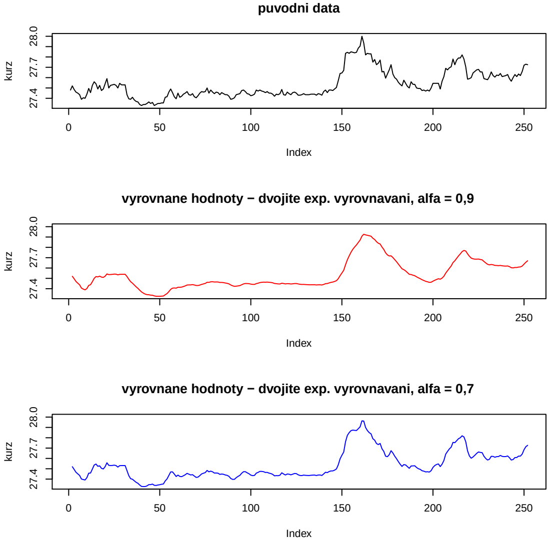




Dvojité exponenciální vyrovnávání



Dvojité exponenciální vyrovnávání



Jednoduché exponenciální vyrovnávání – rekurentní vztah pro odhad hodnot $\beta_0(t)$

$$\min_{\beta_0(t)} \sum_{\tau=0}^{\infty} [y_{t-\tau} - \beta_0(t)]^2 \alpha^\tau$$

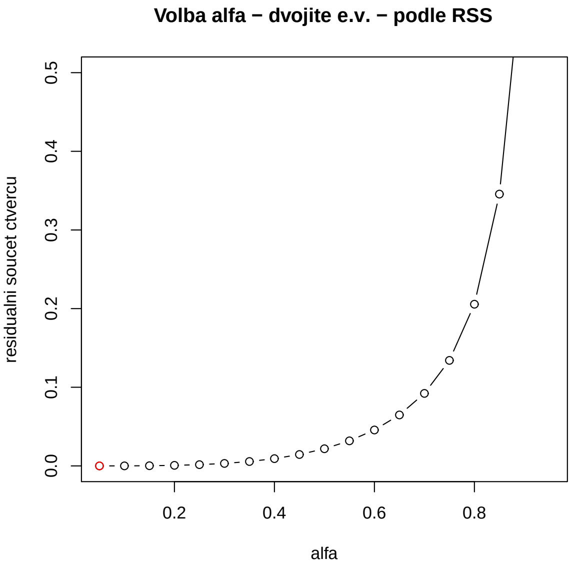
$$\frac{\partial}{\partial \beta_0(t)} \sum_{\tau=0}^{\infty} [y_{t-\tau} - \beta_0(t)]^2 \alpha^\tau = -2 \sum_{\tau=0}^{\infty} [y_{t-\tau} - \beta_0(t)] \alpha^\tau = 0$$

$$\sum_{\tau=0}^{\infty} y_{t-\tau} \alpha^\tau = \beta_0(t) \sum_{\tau=0}^{\infty} \alpha^\tau = \beta_0(t) \frac{1}{1-\alpha}$$

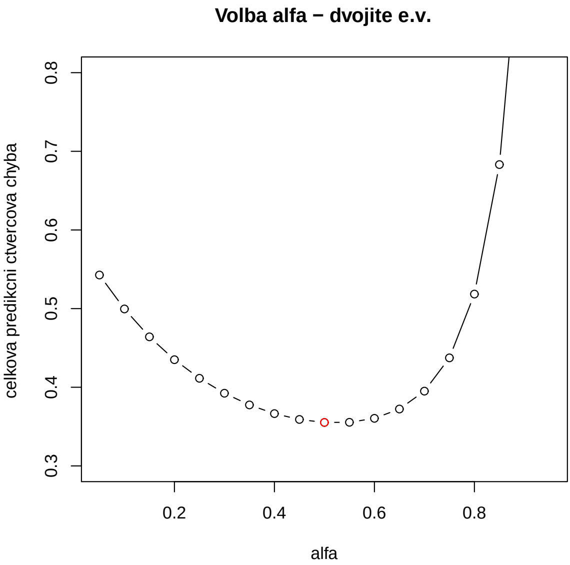
$$\begin{aligned} \beta_0(t) &= (1-\alpha) \sum_{\tau=0}^{\infty} y_{t-\tau} \alpha^\tau \\ &= (1-\alpha) y_t + (1-\alpha) \sum_{\tau=1}^{\infty} y_{t-\tau} \alpha^\tau \\ &= (1-\alpha) y_t + \alpha \beta_0(t-1) \end{aligned}$$

Poznámka: $\beta_0(t) = \widehat{y}_t$ (vyrovnaná hodnota)
Poznámka: Problém určení $\widehat{y}_0$

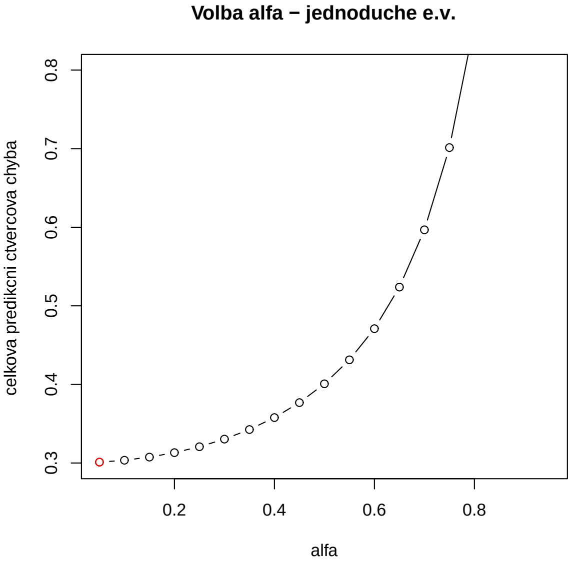
Dvojité exponenciální vyrovnávání – volba α



Dvojité exponenciální vyrovnávání – volba α



Jednoduché exponenciální vyrovnávání – volba α



Srovnání s cvičenou opicí

Co dělá cvičená opice: predikuje zítřejší hodnotu hodnotou z dneška.
Celková čtvercová chyba predikce:

- exp. třídění s optimalizovanou volbou α: 0,301
- cvičená opice: 0,300

Závěr: cvičená opice v tomto případě vítězí nad exponenciálním vyrovnáváním.
Poučení: ne vždy platí, že čím sofistikovanější metoda, tím lepší výsledek.
Poznámka: kdo umí v predikci kurzu výrazně porazit cvičenou opici, zbohatne.

Teorie efektivních trhů

1965 – Eugene Fama (NC 2013)
Tzv. slabá forma teorie efektivních trhů říká, že na základě technické analýzy minulých statistických tvarů nelze ceny akcií na burze předvídat. „Grafařské techniky“ musejí tedy nevyhnutelně selhat.