

Statistické „sledování“ jakosti, procesní.

SPC (Shewhartovy – Statistical Process Control Charts) diagramy

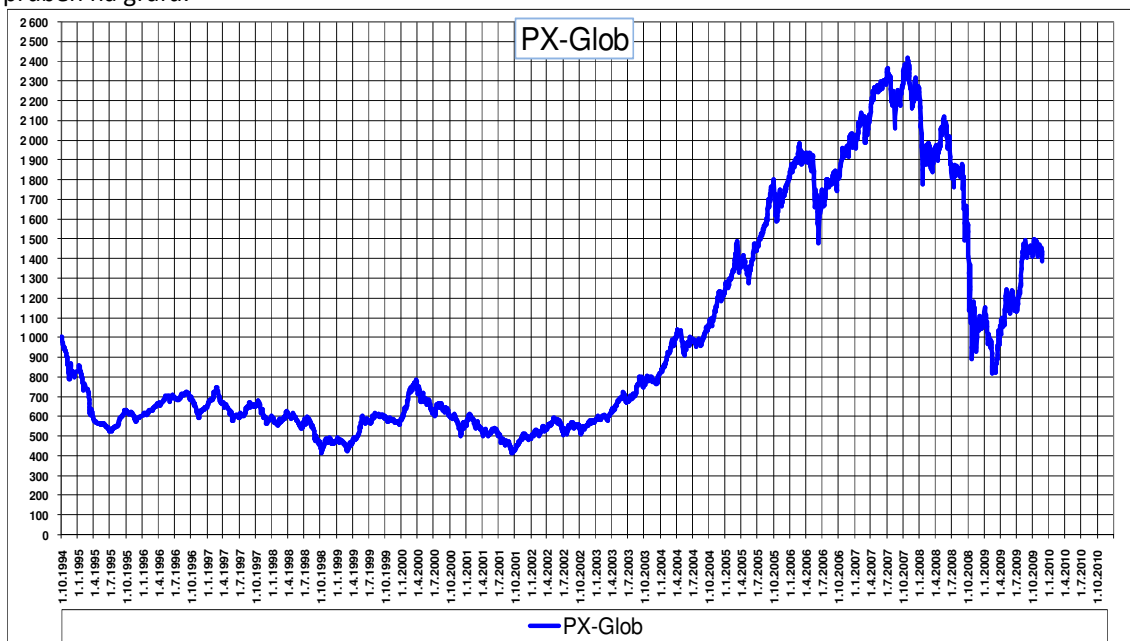
Příklad využití pro sledování kapitálového trhu

Jedním z informativních zdrojů kapitálového trhu jsou indexy trhu. Pražská burza používá dva, jedním z nich je index PX a druhý PX Glob. Dynamika druhého bude předmětem následující (zkrácené) studie. Jeho složení:

Báze indexu PX-GLOB ke dni 12.11.2009

Název	ISIN	Redukovaný počet CP	Redukovaná tržní kapitalizace [mil. Kč]	Váha [%]
ČEZ	CZ0005112300	287 249 111,0	256 226,2	23,75
ERSTE GROUP BANK	AT0000652011	317 925 086,0	253 545,3	23,50
KOMERČNÍ BANKA	CZ0008019106	38 009 852,0	141 396,6	13,11
TELEFÓNICA O2 C.R.	CZ0009093209	322 089 890,0	139 787,0	12,96
VIG	AT0000908504	128 000 000,0	124 812,8	11,57
NWR	NL0006282204	264 065 749,0	44 099,0	4,09
CETV	BMG200452024	50 524 273,0	25 919,0	2,40
UNIPETROL	CZ0009091500	181 334 764,0	25 930,9	2,40
PRAŽSKÁ ENERGETIKA	CZ0005078154	2 553 831,0	17 851,3	1,65
PHILIP MORRIS ČR	CS0008418869	1 913 698,0	16 591,8	1,54
SM PLYNÁRENSKÁ	CZ0005084459	1 472 285,0	8 015,1	0,74
VGP	BE0003878957	18 583 050,0	6 764,2	0,63
PEGAS NONWOVENS	LU0275164910	9 229 400,0	3 857,9	0,36
VČ PLYNÁRENSKÁ	CZ0005092551	633 996,0	3 487,0	0,32
ORCO	LU0122624777	10 943 866,0	2 117,6	0,20
ECM	LU0259919230	6 868 092,0	2 169,6	0,20
PRAŽSKÉ SLUŽBY	CZ0009055158	1 556 125,0	1 271,4	0,12
ENERGOAQUA	CS0008419750	701 000,0	956,9	0,09
AAA	NL0006033375	67 757 875,0	939,8	0,09
RM-S HOLDING	CS0008416251	784 035,0	862,4	0,08
TOMA	CZ0005088559	1 477 266,0	837,6	0,08
SPOLEK CH.HUT.VÝR.	CZ0005092858	3 878 816,0	775,8	0,07
LÉČ.LÁZNĚ JÁCHYMOV	CS0008446753	418 530,0	472,9	0,04
ČESKÁ NÁMOR.PLAVBA	CZ0008413556	187 310,0	92,1	0,01
JČ PAPIRNY VĚTRNÍ	CZ0005005850	2 444 052,0	25,7	0,00
Celkem			1 078 805,9	100,00

Data je možné získat na: <http://www.pse.cz/dokument.aspx?k=Statisticke-Soubory> , historický průběh na grafu:



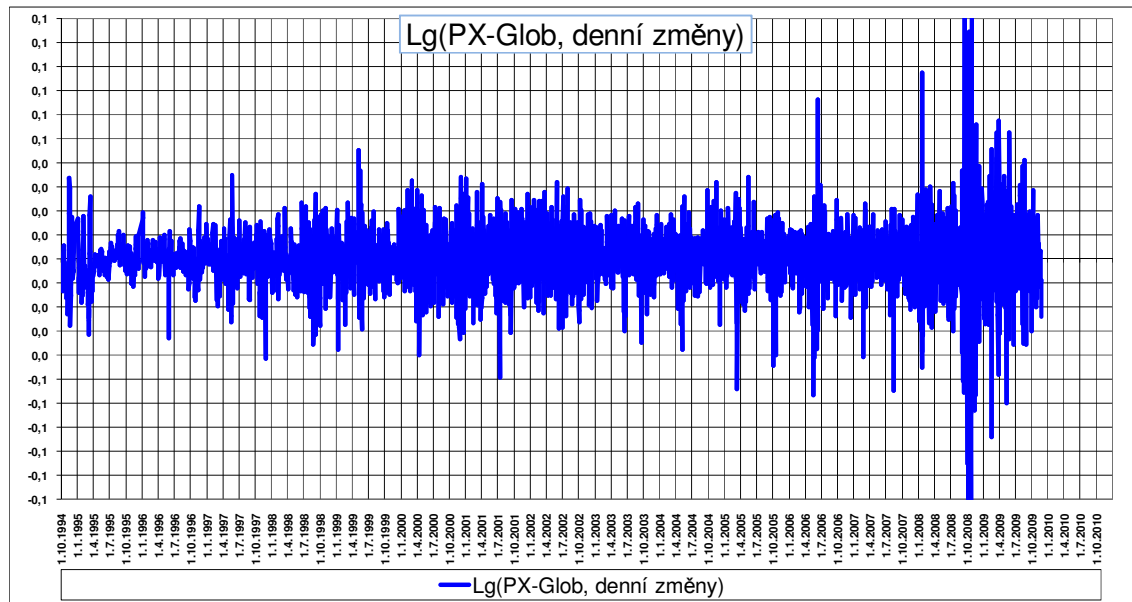
Pro sledování dynamiky trhu je významná hodnota $w_t = \frac{x_t}{x_{t-1}}$, kde:

w_t je mezi-denní výnos ve dni t , čas je měřen v kотаčních dnech a

x_t je hodnota indexu ve dni t .

Z důvodů početních a z důvodů částečné stacionarizace se používá logaritmus tohoto výnosu, tedy:

$L_t = \lg\left(\frac{x_t}{x_{t-1}}\right)$. Jeho průběh:



Obvykle se předpokládá, že tento „logaritmický výnos“ má normální rozdělení. Tato hypotéza byla ověřována pomocí shody modelové a empirické distribuční funkce. Shoda byla ověřována pro tři typy rozdělení:

1. Normální rozdělení, popsané hustotou $f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ a distribuční funkcí

$$F_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

2. Cauchyho rozdělení, popsané hustotou $f_{x_0,d}(x) = \frac{1}{\pi d \left(1 + \left(\frac{x-x_0}{d}\right)^2\right)}$ a distribuční funkcí

$$F_{x_0,d}(x) = \frac{1}{\pi} \arctg\left(\frac{x-x_0}{d}\right) + \frac{1}{2}.$$

3. Laplaceovo rozdělení s hustotou $f_{\mu,b}(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$ a distribuční funkcí

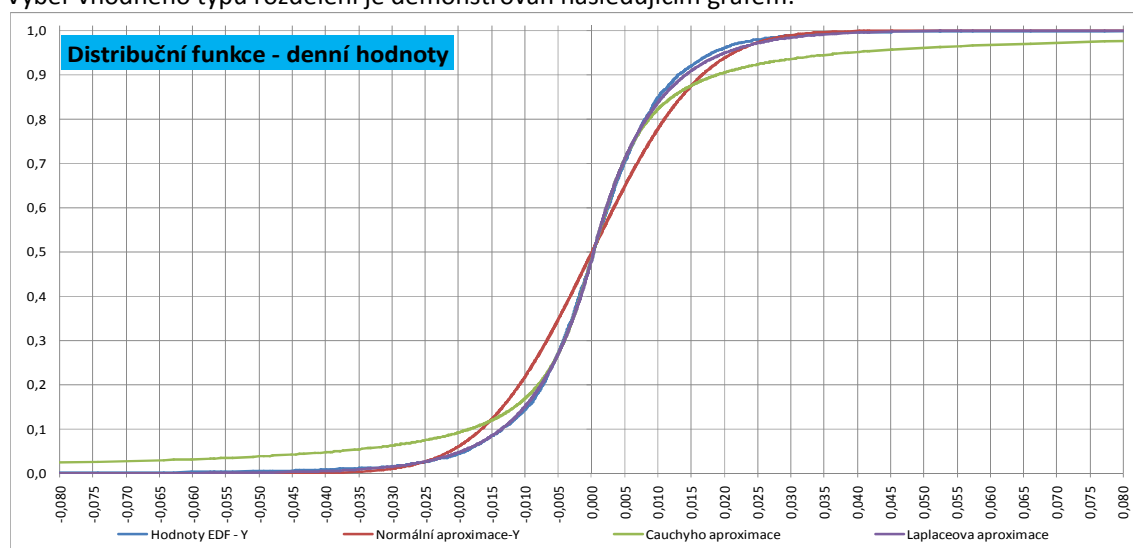
$$F_{\mu,b}(x) = \frac{1}{2} \left(1 + \text{sign}(x-\mu) \left(1 - e^{-\frac{|x-\mu|}{b}} \right) \right).$$

Námět: Navrhněte metody získání parametrických bodových odhadů všech třech rozdělení.

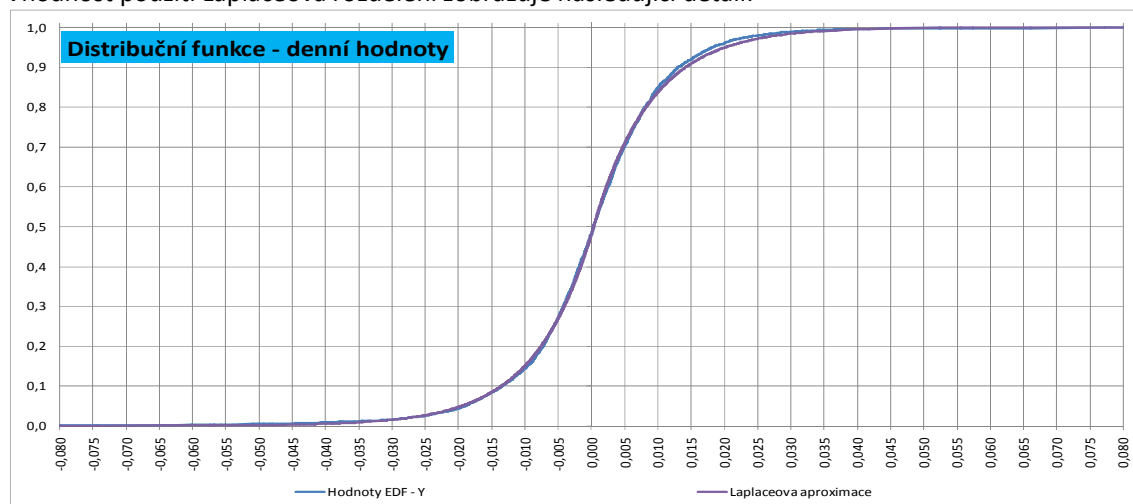
Použitá data zahrnovala časový horizont od 30.9.1994 do 26.11.2009. Základní „popisné parametry“ jsou shrnuty v následující tabulce:

	PX-Glob	Lg(PX-Glob, denní změny)
Maximum	2 414,9	0,1102
Průměr	992,0	0,0001
Medián	688,5	0,0003
Minimum	412,1	-0,1568
StD	559,9	0,0130
Q-rozpětí	808,1	0,0121
Prům. odchyl.	474,7	0,0086
Počet	3 779	3 778

Výběr vhodného typu rozdělení je demonstrován následujícím grafem:

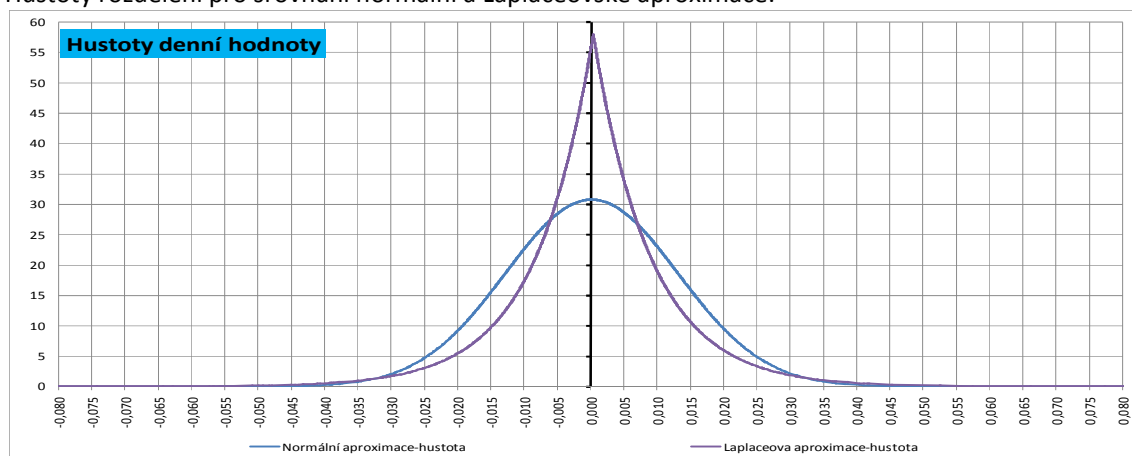


Vhodnost použití Laplaceova rozdělení zobrazuje následující detail:



Pro korektnější (ne však tak názorné) je vhodné využít některých testů shody pro distribuční funkce např. založené na Kolmogorov-Smirnovovských statistikách.

Hustoty rozdělení pro srovnání normální a Laplaceovské aproximace:



Diskuse: Pokud má náhodná proměnná hustotu $f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$ pak má náhodná

proměnná $|x - \mu|$ hustotu $f(x) = \frac{1}{b} e^{-\frac{x}{b}}; x \geq 0; b = \tau$ a distribuční funkci $F(x) = 1 - e^{-\frac{x}{b}}; x \geq 0$,

tedy běžné exponenciální rozdělení. Proto bude následně tato „transformovaná“ náhodná veličina $(|x - \mu|)$ využita k testování „běžného“ a „neběžného“ chování, a to z důvodu jednoduchého pravděpodobnostního popisu (modelu). Pak je nestranným a vydatným odhadem parametru b průměr z pozorování transformovaných dat.

Pro konstrukci příslušných karet budeme potřebovat rozdělení průměru, maxima a minima z náhodného výběru pevného rozsahu n . Za takový rozsah budeme považovat **jeden týden, tedy pět registrací** transformace indexu PX-Glob. Základní „popisné parametry“ jsou opět shrnuty v následující tabulce:

Den	Týdenní průměr, abs(Lg(PX-Glob, denní změny)-median)	Týdenní StD, abs(Lg(PX-Glob, denní změny)-median)	Týdenní R, abs(Lg(PX-Glob, denní změny)-median)	Týdenní MIN, abs(Lg(PX-Glob, denní změny)-median)	Týdenní MAX, abs(Lg(PX-Glob, denní změny)-median)	Roční zpětný průměr z týdenních průměrů	Roční zpětný průměr z týdenních StD	Roční zpětný průměr z týdenních R
Minimum	0,0007	0,0005	0,0012	0,0000	0,0014	0,0031	0,0022	0,0054
Průměr	0,0086	0,0060	0,0146	0,0024	0,0170	0,0085	0,0059	0,0145
Medián	0,0074	0,0049	0,0121	0,0015	0,0142	0,0080	0,0056	0,0137
Maximum	0,0847	0,0605	0,1513	0,0578	0,1571	0,0217	0,0145	0,0360
StD	0,0070	0,0049	0,0123	0,0034	0,0135	0,0037	0,0024	0,0060
Q-rozpětí	0,0056	0,0043	0,0106	0,0025	0,0114	0,0024	0,0016	0,0040
Počet	757	757	757	757	757	706	706	706

Nejprve rozdělení součtu transformovaných náhodných proměnných:

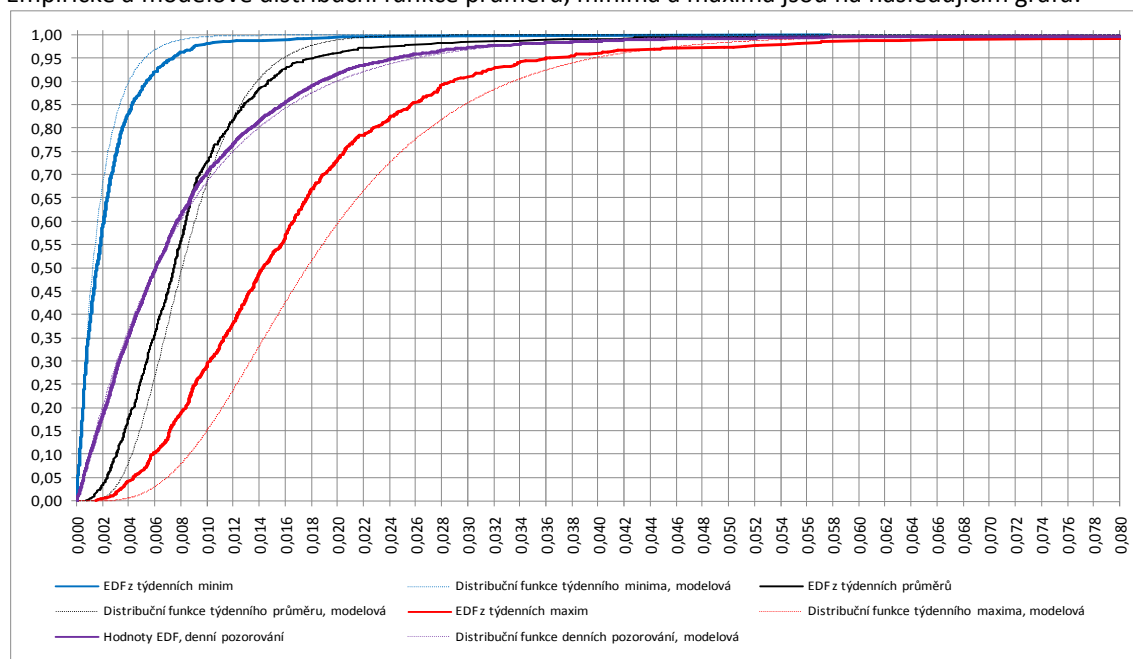
Hustota $f_{\sum_{i=1}^n x_i}(x) = \frac{x^{n-1} e^{-\frac{x}{\tau}}}{\tau^{n-1} \Gamma(n)}$ a distribuční funkce $F_{\sum_{i=1}^n x_i}(x) = \frac{\int_0^x z^{n-1} e^{-\frac{z}{\tau}} dz}{\tau^{n-1} \Gamma(n)} = G\left(x; n, \frac{1}{\tau}\right)$, tedy klasické gama rozdělení.

A odtud **rozdělení průměru** $f_{\frac{1}{n} \sum_{i=1}^n x_i}(x) = n f_{\sum_{i=1}^n x_i}(nx)$ a $F_{\frac{1}{n} \sum_{i=1}^n x_i}(x) = G\left(nx; n, \frac{1}{\tau}\right)$ (poznámka: pozor konvence pro označování gama rozdělení je neustálená, někdy bývá užit $G\left(x; n, \frac{1}{\tau}\right) \equiv G(x; n, \tau)$). Tomu odpovídá p-quantil $x_p = \frac{1}{n} G^{-1}\left(p; n, \frac{1}{\tau}\right)$.

Rozdělení minima $F_{\min\{x_1, \dots, x_n\}}(x) = 1 - (1 - F_x(x))^n = 1 - e^{-\frac{nx}{\tau}}$ je opět exponenciální rozdělení (toto je speciální případ vlastnosti „zachování minima“ pro Weibullovo rozdělení) s parametrem $\frac{\tau}{n}$. Kvantil rozdělení minima bude $x_p = \frac{-\tau}{n} \lg(1-p)$.

Rozdělení maxima $F_{\max\{x_1, \dots, x_n\}}(x) = F_x^n(x) = \left(1 - e^{-\frac{x}{\tau}}\right)^n$ už není exponenciální rozdělení. Kvantil rozdělení maxima bude $x_p = -\tau \lg\left(1 - p^{\frac{1}{n}}\right)$.

Empirické a modelové distribuční funkce průměru, minima a maxima jsou na následujícím grafu:



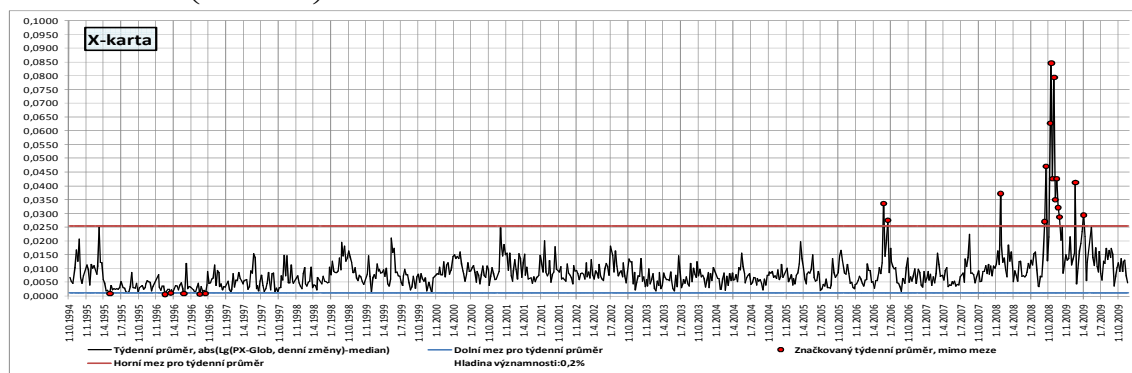
Z tohoto grafu je patrné, že modelová a empirická distribuční funkce denních pozorování vykazují přijatelnou shodu u distribučních funkcí týdenních průměrů, minim a maxim tomu už tak není. Je to dáno tím, že model byl konstruován na denních pozorováních a je tedy odhadem s nějakými vlastnostmi (zde kvalita shody empirické a modelové distribuční funkce). Tyto vlastnosti už však nebývají přenosné na různé operace s modelem (přesněji s jeho odhadem). Přesto v této situaci nedokonalého modelu (a v té jsme v statistice téměř vždy) můžeme zvolený typ modelu (a jeho parametrické naplnění) považovat za model standardní situace a „vyhledávat situace“ mimo tento model = nestandardní.

Konstrukce X-karty

X-karta bude zahrnovat průběh týdenních průměrů transformovaných pozorování v konfrontaci s mezemi pro zvolenou *pravděpodobnost planého poplachu* p = *hladina významnosti*:

$$HMZ_x = \frac{1}{n} G^{-1} \left(1 - p/2; n, \frac{1}{\hat{\tau}} \right)$$

$$DMZ_x = \frac{1}{n} G^{-1} \left(p/2; n, \frac{1}{\hat{\tau}} \right)$$



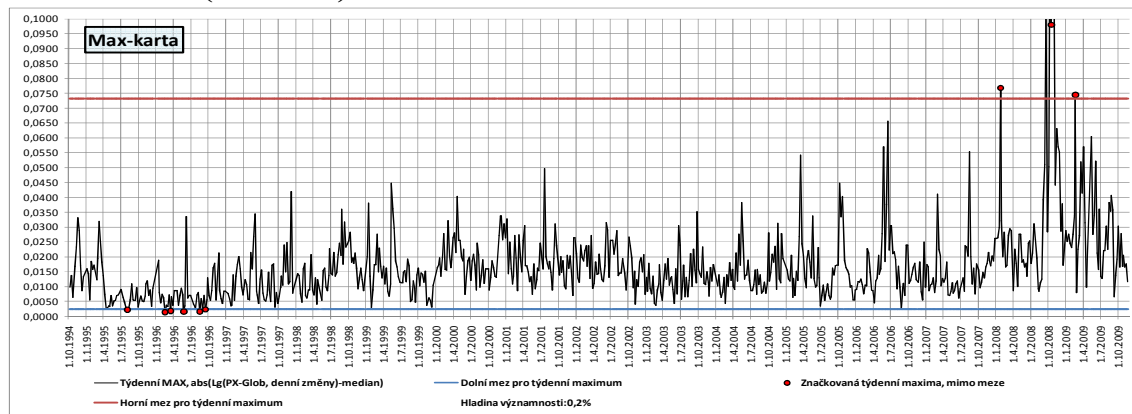
Alternativy ke kartám testujícím variabilitu

MAX-karta

Tato karta testuje týdenní maxima, bude zahrnovat průběh týdenních maxim, meze pro ně budou:

$$HMZ_{\max} = -\hat{\tau} \lg \left(1 - (1 - p/2)^n \right)$$

$$DMZ_{\max} = -\hat{\tau} \lg \left(1 - (p/2)^n \right)$$

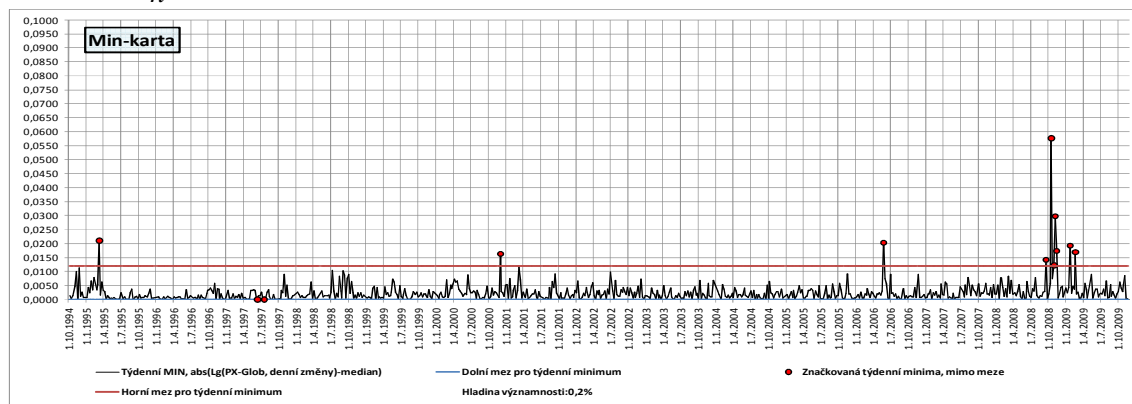


MIN-karta

Tato karta testuje týdenní minima, bude zahrnovat průběh týdenních minim, meze pro ně budou:

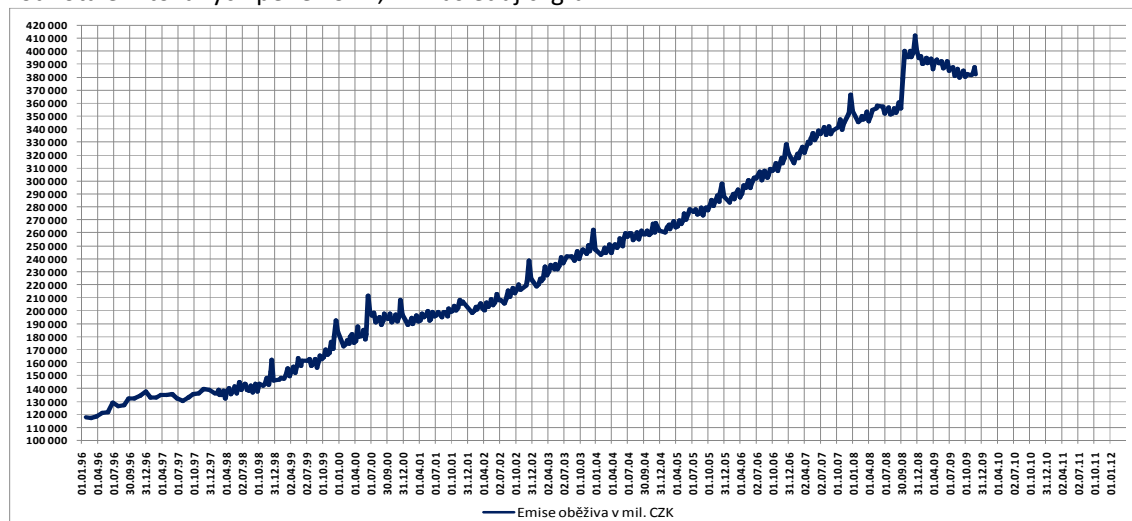
$$HMZ_{\min} = \frac{-\hat{\tau}}{n} \lg(1 - (1 - p/2))$$

$$DMZ_{\min} = \frac{-\hat{\tau}}{n} \lg(1 - p/2)$$



Interpretace

Všechny tři „karty“ (diagramy) ukazují nestandardní chování pražského akciového trhu od počátku října 2008. To bylo způsobeno dílčím „RUNem“ na české akciové fondy pod vlivem „informací“ o okolní finanční krizi. Podílové a „účastnické“ listy byly „proměňovány“ v peníze. Tomu odpovídá i hodnota emitovaných peněz ČNB, viz následující graf.

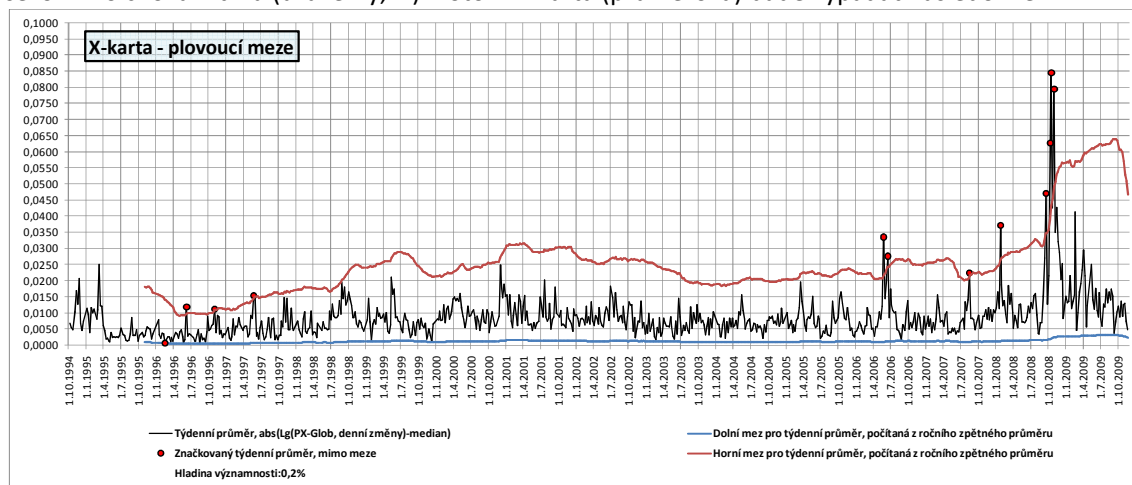


Data, zdroj: http://www.cnb.cz/cs/o_cnb/hospodareni/dekadni_balance/dekadni_balance_archiv/

Zajímavá je v tomto kontextu i „informační politika“ ČNB. Čísla byla zveřejňována, průběhy ne (to ostatně mimo měnově-politické predikce ČNB nečiní [až na malé výjimky] i u jiných ukazatelů). Veškeré orgány a osoby ČNB se zdržely jakýchkoliv komentářů, tím ČNB přispěla k prakticky „neviditelnému“ odeznění a minimalizace efektů tzv. finančnímu segmentu krize v ČR.

Modifikace

Uvedené typy karet vycházely z odhadu parametru τ z celé známé historie. Pokud chceme respektovat časovou proměnnost českého akciového trhu, můžeme využít „plovoucích verzí“ když za základ vezmeme odhad τ z právě uplynulých posledních 52 týdnů (roku) – např. pro dílčí eliminaci sezónního chování trhu (uzávěrky, ...). Potom X-karta (průměrová) bude vypadat následovně:“



Námět: Sestrojte ostatní karty.

Námět: Odvodte rozdělení výběrové směrodatné odchylky $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ pro náhodný

výběr z exponenciální proměnné s hustotou $f(x) = \frac{1}{\tau} e^{-\frac{x}{\tau}}; x \geq 0$ a sestrojte S-kartu.

Přílohy: EXCElovské soubory: VSM-11-0.XLS, VSM-11-1.XLS, VSM-11-2.XLS, VSM-11-3.XLS, VSM-11-4.XLS.

Doporučená a zdrojová literatura:

Jiří Reif	Metody matematické statistiky, ZČU v Plzni 2004
Jaroslav Hátle, Jiří Likeš	Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974.
C. Radhakrishna Rao	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
Mendenhall. W.	Introduction to Probability and Statistics. PWS-KENT, Publishing Company, Boston 1987.
Groebner, D.F., Shannon P. W.	Business Statistics. A Decision-Making Approach. Merrill Publishing Company, Columbus Ohio, 1989.
Douglas C. Montgomery	Introduction to Statistical Quality Control, 5th Edition. John Wiley & Sons, August 2004.
	http://www.qualityamerica.com/