Přírůstkové statistické modely časových řad

Arnoštka Netrvalová, Hana Mašková [1]

Abstrakt

Tento příspěvek je součástí rozsáhlejších prací na téma předvídání kurzů. Pokud pro kurz $x_t = \Sigma_{i=1}^{t} \Delta_i$ předpokládáme, že proměnné Δ_i jsou stejně rozdělené a nezávislé, jedná se o „náhodnou procházku“. Problém je v tom, že proměnné Δ_i nejsou zcela nezávislé, resp. musíme uvažovat slabou závislost, kterou se pokusíme modelovat pomocí přírůstků. Je tedy zřejmé, že kurz je nestacionární proces, zatímco „přírůstky“ se chovají stacionárně. Podíváme-li se na distribuční funkci přírůstků, zjistíme, že nemají normální rozdělení.
Distribuční funkci ve většině případů nepotřebujeme znát celou a úloha se tak redukuje na odhad několika málo bodů, resp. na odhad pravděpodobnosti alternativního jevu.

Klíčová slova

stacionární časová řada, bodový odhad bodu distribuční funkce

1. Úvod

Předpokládejme dostatečně dlouhé pozorování časové řady v „ekvidistantním“ vzorkování: $x_1, x_2, ..., x_{i-1}, x_i, x_{i+1}, ...x_n$. Dále budeme předpokládat, že se jedná o kladné náhodné proměnné. Z pozorovaných hodnot odvodíme nové náhodné proměnné:

a) rozdíl: $\xi_\tau = x_i - x_{i-\tau}$,

b) podíl: $\eta_\tau = \dfrac{x_i}{x_{i-\tau}}$,

c) logaritmus podílu: $\lambda_\tau = lg\left(\dfrac{x_i}{x_{i-\tau}}\right)$.

Pozn.: Pokud máme k dispozici dostatečně dlouhé časové řady, můžeme si dovolit „vytvořit“ řady zpožděné o více kroků, tedy dvoudenní, třídenní atd. změny: $x_i = x_{i-\tau} + \Delta_\tau$, kde Δ_τ je např. 5-denní změna, resp. $x_i = k_\tau \cdot x_{i-\tau}$ pro podíly a $\lg x_i = \lg(k_\tau) + \lg(x_{i-\tau})$ pro logaritmy podílů.

Pomocí odhadů pravděpodobnostních charakteristik těchto náhodných proměnných (u nichž předpokládáme jejich stacionární chování) můžeme odvodit distribuční funkci náhodné proměnné x_i:

$$F_{x_i}(x) = P\{x_i < x\} = P\{x_i - x_{i-\tau} < x - x_{i-\tau}\} = F_\xi(x - x_{i-\tau}) \equiv F_{x_i,\xi}(x / x_{i-\tau}) \quad (1)$$

$$F_{x_i}(x) = P\{x_i < x\} = P\left\{\frac{x_i}{x_{i-\tau}} < \frac{x}{x_{i-\tau}}\right\} = F_\eta\left(\frac{x}{x_{i-\tau}}\right) \equiv F_{x_i,\eta}(x / x_{i-\tau}) \quad (2)$$

[1] Ing. Arnoštka Netrvalová, Ing. Hana Mašková, Západočeská univerzita v Plzni, FAV – KIV, Univerzitní 22, 306 14 Plzeň, netrvalo@kiv.zcu.cz, maskova@kiv.zcu.cz.

$$F_{x_i}(x) = P\{x_i < x\} = P\left\{\frac{x_i}{x_{i-\tau}} < \frac{x}{x_{i-\tau}}\right\} = P\left\{\lg\frac{x_i}{x_{i-\tau}} < \lg\frac{x}{x_{i-\tau}}\right\} =$$

$$= F_\lambda\left(\lg\frac{x}{x_{i-\tau}}\right) \quad \equiv \quad F_{x_i,\lambda}(x / x_{i-\tau}) \qquad (3)$$

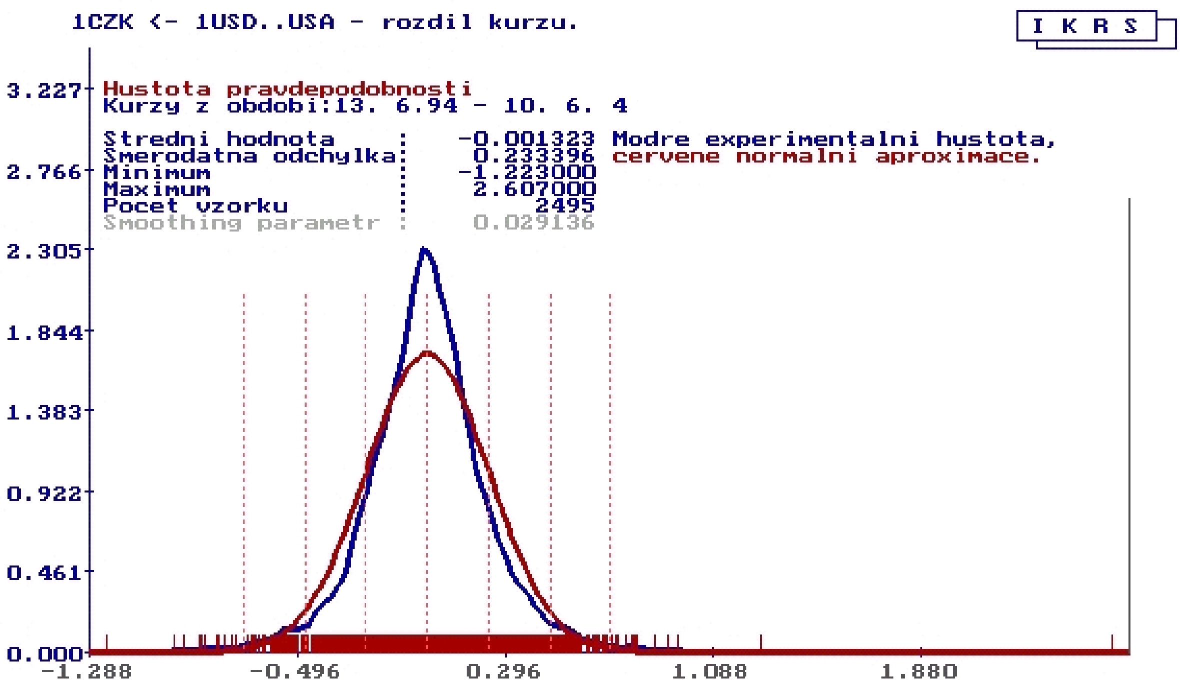
Samozřejmě, volba uvedených „spojovacích“ náhodných proměnných je dána historicky a výpočetní jednoduchostí. Z toho vůbec neplyne, že neexistují jiné „spojovací“ náhodné proměnné, možná i lepší než uvedené.

K uvedeným odhadům bodů distribuční funkce $F_{x_i}(x)$ nám pak stačí bodové odhady hodnot distribučních funkcí $F_\xi(x),\ F_\eta(x),\ F_\lambda(x)$. Pro pevně daná x pak stačí sledovat alternativní jev „kolikrát“ pozorovaná hodnota $(x_i - x_{i-\tau}),\ (x_i / x_{i-\tau}),\ \lg(x_i / x_{i-\tau})$ nepřekročí zadanou hodnotu $(x - x_{i-\tau}),\ (x / x_{i-\tau}),\ \lg(x / x_{i-\tau})$.

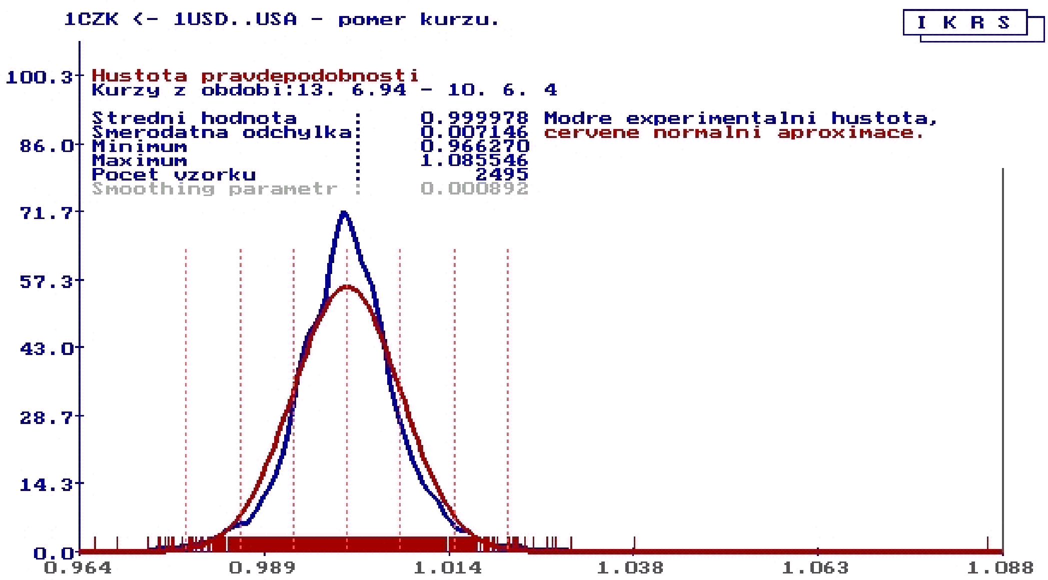
Statistickým prostředkem je pak odhad parametru alternativního rozdělení.

2. Motivace a důvody

Podívejme se na následující obrázky hustot přírůstků kurzů:



Obr.č. 1: Hustota pravděpodobnosti pro rozdíl kurzů



Obr.č.2: Hustota pravděpodobnosti pro podíl kurzů

Jak je z obrázků 1 a 2 zřejmé, skutečné hustoty „zvolených spojovacích“ náhodných proměnných jsou špičatější než normální rozdělení. Také lze konstatovat, že hustoty jsou spíše hustotami s těžkými konci než hustotami popsatelnými gaussovským rozdělením. Tuto skutečnost, i se všemi jejími problémy, je nezbytné respektovat při odhadování. Samozřejmě, uvedená fakta by bylo možné prokázat exaktnějšími metodami.

Při testování rizika, zda kurz v následujícím období neopustí danou oblast (jednostranný nebo dvoustranný interval), potřebujeme (pouze) kvalitní odhady nejvýše dvou (konečného, malého počtu) bodů distribuční funkce nebo kvantilů. Nepotřebujeme tedy běžně celou distribuční funkci (empirickou distribuční funkci včetně „boxu“ ve kterém leží [3]). Jedná se tedy o získání odhadu pravděpodobnosti alternativního jevu „nepřekročení hladiny *x*“.

Může se též jednat o řešení inverzního problému „nalezení hladiny *x*“ k dané pravděpodobnosti. Dále se budeme zabývat řešením první úlohy. Druhá úloha je řešitelná pomocí první, např. metodou půlení intervalu (vyhledávání dané hladiny) s předem danou přesností.

3. Požadavky na bodový odhad a metoda nalezení

Nejprve si uvedeme potřebné pojmy:

- pravděpodobnost, že kurz překročí danou hladinu *x*: $P\{x_i > x\}$,
- pravděpodobnost, že kurz nepřekročí danou hladinu *x*: $P\{x_i < x\}$.

Pravděpodobnost, že kurz „zůstane v mezích“:

$$P\{x_D < x_i < x_H\} = P\{x_i < x_H\} - P\{x_i < x_D\} = F_{x_i}(x_H) - F_{x_i}(x_D), \tag{4}$$

kde x_H a x_D jsou horní a dolní meze. Distribuční funkce uvedené ve vztahu (4) jsou závislé na čase. Abychom se závislosti zbavili, použijeme přírůstky: $P\{x_{i+1} < x\} = P\{x_{i+1} - x_i < x - x_i\}$, kde označíme $\xi_1 = x_{i+1} - x_i$. Odtud dostáváme: $P\{\xi_1 < x - x_i\} = F_{\xi_1}(x - x_i)$ – tato distribuční funkce je již „přijatelně“ stacionární.

Dále uvažujme empirickou distribuční funkci:

$$\overline{F}_{\xi}(x) = \frac{\text{počet pozorování, kdy } \xi \text{ nepřekročí } x}{\text{počet všech pozorování}}. \tag{5}$$

Takto definovaná distribuční funkce má hodnotu 1 za posledním pozorováním a hodnotu 0 před prvním pozorováním, tj. pozorované minimum a maximum „považuje“ za meze, což samozřejmě nemusí být pravda – tzv. problém nulových frekvencí.

Proto pro náš odhad budeme požadovat následující:
- odhad musí být nenulový i pro hladinu nižší než minimální pozorování,
- odhad musí být nejednotkový i pro hladinu vyšší než maximální pozorování.

Jednou z možností řešení je aposteriorní bayesovský odhad [2] pravděpodobnosti daného alternativního jevu – nepřekročení nebo naopak překročení dané hladiny.

Označíme z_n počet pozitivních pozorování daného jevu z celkem n (nezávislých) pozorování, potom:

$$P\{z_n = k / p\} = \binom{n}{k} p^k (1-p)^{n-k} .$$

Dále budeme mít k dispozici apriorní hustotu $f(p)$, budeme přepokládat, že se jedná o beta-hustotu:

$$f(p) = \frac{(a+b-1)!}{(a-1)!(b-1)!} p^{a-1}(1-p)^{b-1}$$

pro a, b přirozená (jedná se o zjednodušení). Jde o Jeffreysovu hustotu k danému problému [2]. Pro $a = b = 1$ se jedná o rovnoměrné rozdělení na intervalu (0,1), tedy o využití principu neurčitosti. Potom:

$$P\{z_n = k, p\} = P\{z_n = k / p\} \cdot f(p) = \binom{n}{k} p^k (1-p)^{n-k} f(p),$$

kde při pevném p se jedná o pravděpodobnost a při pevném k se jedná o hustotu pravděpodobnosti.

Pak dostáváme pro aposteriorní hustotu „parametru“ p:

$$f(p / z_n = k) = \frac{P\{z_n = k, p\}}{P\{z_n = k\}} = \frac{\binom{n}{k} p^k (1-p)^{n-k} f(p)}{\binom{n}{k} \int_0^1 x^k (1-x)^{n-k} f(x) dx} =$$

$$= \frac{(n+a+b-1)!}{(k+a-1)!(n-k+b-1)!} p^{k+a-1} (1-p)^{n-k+b-1} .$$

Podobným výpočtem získáme i optimální aposteriorní bayesovský odhad p za předpokladu $z_n = k$:

$$E\{p / z_n = k\} = \int_0^1 p \frac{(n+a+b-1)!}{(k+a-1)!(n-k+b-1)!} p^{k+a-1} (1-p)^{n-k+b-1} dp = \frac{k+a}{n+a+b} \equiv \hat{p} . \quad (6)$$

Výraz pro takový odhad můžeme psát i ve tvaru:

$$\hat{p}(k) = \frac{k+a}{n+a+b} = \frac{\frac{k}{n} + \frac{a}{n}}{1 + \frac{a}{n} + \frac{b}{n}}, \quad (7)$$

kde náhodná proměnná k/n má pro $n > 9 / p(1-p)$ asymptoticky normální rozdělení: $N\left(p, \frac{p(1-p)}{n}\right)$ [1]. Toho lze využít pro intervalový odhad [1].

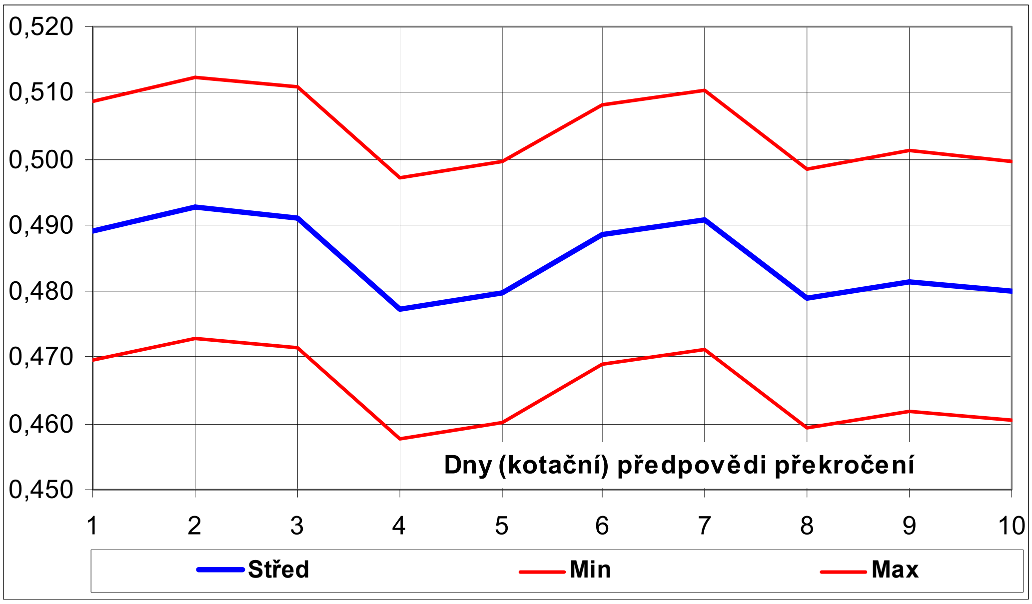
Je zřejmé, že odhad (6) splňuje výše požadované podmínky nenulovosti a nejednotlivosti, protože:

$$\hat{p}(0) = \frac{a}{n+a+b} > 0\,;\ \ \hat{p}(1) = \frac{n+a}{n+a+b} < 1\,.$$

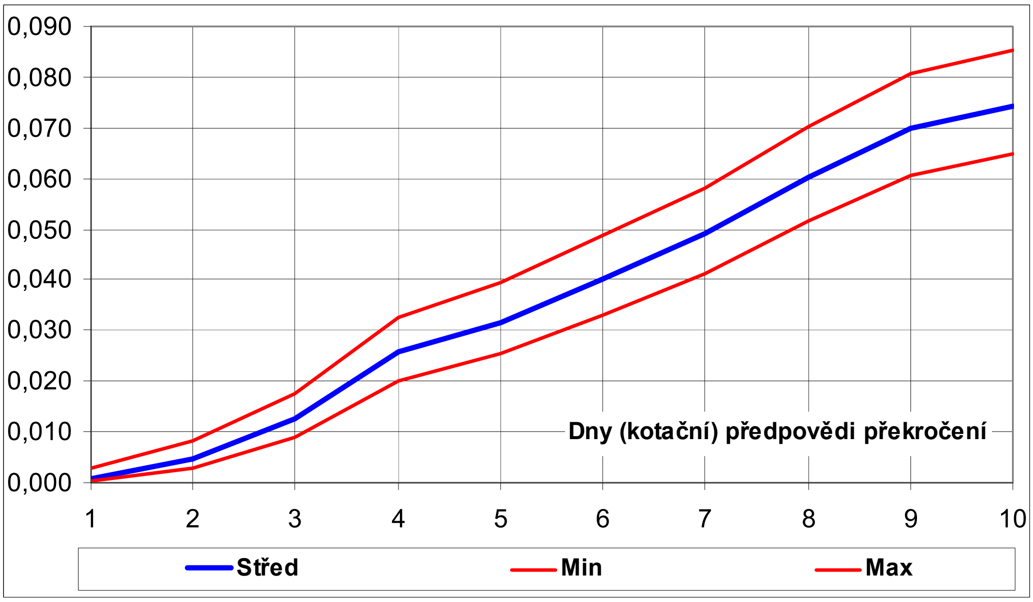
V praxi bude výhodné použít principu neurčitosti $a = b = 1$, pak $\hat{p}(k) = \frac{k+1}{n+2}$. Pro takový odhad platí:

$$\hat{p}(0) = \frac{1}{n+2} \xrightarrow{n\to\infty} 0\,;\ \ \hat{p}(1) = \frac{n+1}{n+2} \xrightarrow{n\to\infty} 1\,.$$

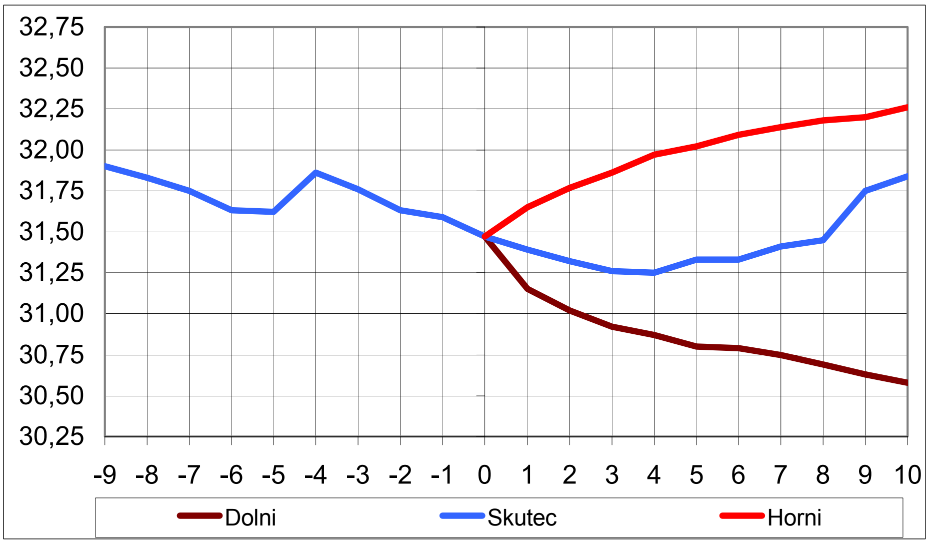
4. Příklad využití



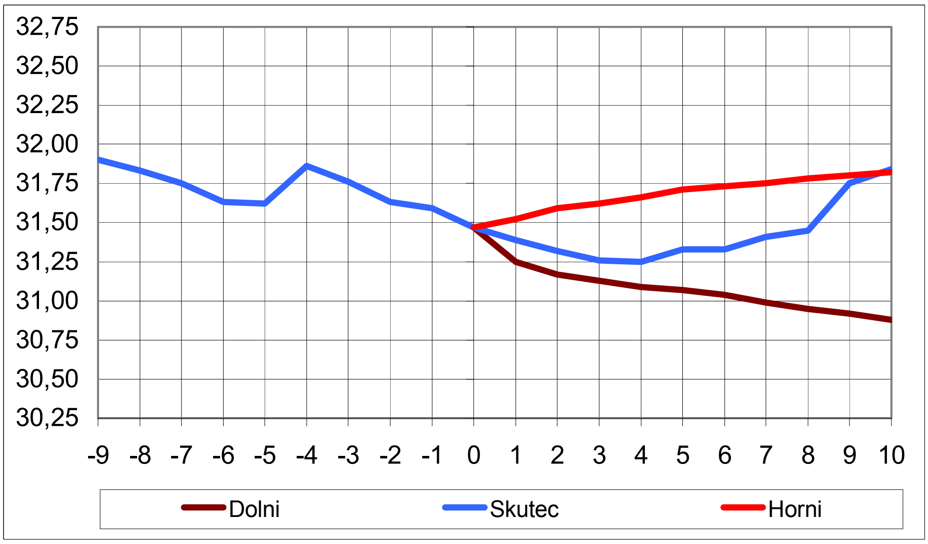
Obr.č. 3: Pravděpodobnost toho, že kurz USD v následujících dnech překročí současných kurz o 0,001 Kč



Obr.č. 4: Pravděpodobnost toho, že kurz USD v následujících dnech překročí současných kurz o 1,001 Kč



Obr.č. 5: Kurz CZK:EUR, hladina významnosti 5 %, pozorování 1.1.1999-16.6.2004 (celkem 1360)



Obr.č. 6: Kurz CZK:EUR, hladina významnosti 20 %, pozorování 1.1.1999-16.6.2004, bod 0 = 2.6.2004

Literatura

[1] HÁTLE, J., LIKEŠ, J.: *Základy počtu pravděpodobnosti a matematické statistiky*. SNTL/ALFA, Praha, 1974.

[2] HUŠKOVÁ, M.: *Bayesovské metody*. Skripta, Univerzita Karlova, Praha, 1985.

[3] RÉNYI, A.: *Teorie pravděpodobnosti*. ACADEMIA, Praha, 1972.

Summary – Incremental statistical models of time series

This contribution is a part of works dealing with exchange rate forecasting. If we suppose the exchange rate $x_t = \Sigma_{i=1}^{t} \Delta_i$, where Δ_i are iid variables, then it is a “random walk“. The problem is that the variables Δ_i are not exactly independent and so we have to consider a weak dependency, which will be modelled by means of increments.

In most cases, however, we do not need to know all the distribution function of increments, so the task reduces to the estimation of several points, which is the probability estimation of an alternative event.