

## K odhadu volatility finančních řad při oceňování derivátů<sup>#</sup>

Michal Černý\*

### 1. Úvod

K odhadu reálných hodnot (*fair prices, fair values*) klasických evropských opcí (a některých typů neklasických opcí) se nejčastěji užívá Black-Scholesův model. Tento model je řešením stochastické rovnice, jež popisuje strategii nákupů a prodejů podkladového instrumentu, pomocí které lze opci replikovat (tzv.  $\Delta$ -hedging). Klasický Black-Scholesův model je založen na řadě předpokladů; jmenujme alespoň ty nejdůležitější: (a) podkladová proměnná se chová jako lognormální náhodná procházka s driftem  $\mu$  a volatilitou  $\sigma$ , kde  $\sigma$  je (známá) konstanta, (b) bezriziková úroková sazba je známá konstanta, (c) bezrizikový výnos podkladového instrumentu je známá konstanta, (d) replikační strategii ( $\Delta$ -hedging) je možné realizovat ve spojitém čase, (e) držba podkladového instrumentu ani replikační strategie nenesou žádné transakční náklady, (f) na trhu neexistují arbitrážní příležitosti. Předpoklady je možné různým způsobem rozvolňovat a Black-Scholesův model zobecňovat: například tak, že místo předpokladu konstantní bezrizikové úrokové sazby se pracuje s modelem, kde bezriziková úroková sazba je známou funkcí času, anebo s modelem, kde je bezriziková úroková sazba stochastická.

Hodnoty opcí silně závisí na volatilitě podkladové proměnné. Klasický model předpokládá volatilitu konstantní (i když lze uvažovat i zobecněné modely, kde volatilita je funkcí času anebo funkcí podkladové proměnné). „Skutečná“ volatilita je snad vždy neznámá a je tudíž nutné ji odhadovat. Odhad mimo jiné komplikuje i skutečnost, že reálné časové řady vykazují volatilitu, jež není v čase konstantní – je třeba proto odhadovat *aktuální* volatilitu. Je známý fakt, že odhad volatility je při oceňování opcí Black-Scholesovým modelem patrně nejvážnější a nejdůležitější problém; je totiž běžné, že každý procentní bod volatility má obrovskou cenu.

Při odhadu je v principu možné uvažovat dva přístupy:

- (i) odhadovat volatilitu z kotací burzovních opcí; odhad volatility je řešením rovnice

$$BS(\sigma) = FV,$$

kde  $BS(\sigma)$  je oceňovací model opce (u klasických opcí nejčastěji tradiční Black-Scholesův vzorec), chápáný jako funkce v proměnné  $\sigma$ , a  $FV$  je bursovní kotace opce, anebo

- (ii) odhadovat volatilitu z historických dat.

V praxi převažuje způsob (ii). Často je třeba ocenit deriváty, jejichž ekvivalenty na organizovaných trzích neexistují; jako příklad lze uvést třeba zaměstnanecké akciové opce, které mívají splatnost kolem pěti let, zatímco na burzách se s opcemi s takto dlouhými

<sup>#</sup> Práce byla podpořena grantem GAČR č. 402/06/0190

\* RNDr. Ing. Michal Černý, Ph.D.; Katedra ekonometrie, Vysoká škola ekonomická v Praze, [cernym@vse.cz](mailto:cernym@vse.cz).

splatnostmi neobchoduje. V takovém případě nezbyvá než odhadnout volatilitu z historických dat.

Je otázkou, zdali reálné časové řady splňují předpoklad (a); odpověď zní, že s vysokou mírou pravděpodobnosti nikoliv. Je hlubokou (a patrně věčnou) otázkou, jaký je „skutečný“ proces, který finanční časové řady generuje; tím se však zde nebudeme zabývat. Poznamenejme jen, že Black-Scholesův model je populární mimo jiné i proto, že praxe ukazuje, že je vůči nesplnění předpokladu (a) poměrně robustní. Budeme se zabývat dílčí otázkou: je volatilita časové řady konstantní? Anebo existuje jeden nebo více okamžiků, ve kterých dochází ke změně volatility (tzv. *change point*)? A kdy tyto okamžiky nastávají?

Tyto otázky mají velký význam. Při odhadu volatility z historických dat se totiž otvírá přirozená otázka, *jak dlouhou historii uvažovat*. Často totiž máme k dispozici data za dlouhou dobu (třeba i několik desítek let denních, a dokonce i „intra-day“ dat), jež zahrnuje různá historická období, která se vyznačují nesrovnatelnými ekonomickými podmínkami. Situace v minulosti se může významně odlišovat od současnosti — v minulosti mohlo např. proběhnout období krizí, období růstu a prasknutí „bublin“ na akciových trzích apod. I prosté pozorování finančních časových řad potvrzuje, že jejich volatilita není v čase konstantní. Protože pro Black-Scholesův model potřebujeme odhadovat *okamžitou volatilitu*, obvykle se při odhadu volatility obvykle nebere v úvahu celá dlouhá řada dat, kterou máme k dispozici, nýbrž jen data za relativně krátké historické období. Někdy se užívá i různých metod s „exponenciálním zapomináním dějin“, kdy se historickým pozorováním přisuzují exponenciálně klesající váhy.

Wilmott ([8], str. 112) k odhadu volatility uvádí:

Nejčastějším odhadem volatility je jednoduše

$$\sqrt{\frac{1}{(M-1)\delta t} \sum_{i=1}^M (R_i - \bar{R})^2} \quad (1)$$

$(R_i = \frac{S_i - S_{i-1}}{S_{i-1}})$  je výnosnost za  $i$ -té období, uzavírací cena  $i$ -tého dne,  $M$  je počet období a  $\delta t$  je čas). Je-li  $\delta t$  dostatečně malé, je možné průměrnou výnosnost  $\bar{R}$  ignorovat. Pro malé  $\delta t$  lze užít i

$$\sqrt{\frac{1}{(M-1)\delta t} \sum_{i=2}^M (\ln S_i - \ln S_{i-1})^2} \quad (2)$$

Je velmi nepravděpodobné, že by volatilita byla konstantní, ať se podíváme na kterékoliv aktivum. Když se změní ekonomické podmínky, sezónnost apod., nevyhnutelně dojde ke změně volatility v čase. Chceme-li znát dnešní volatilitu, musíme při výpočtu užít alespoň nějaká historická data. Bohužel to znamená, že vůbec nemáme záruku, že odhadujeme volatilitu dnešní.

Typicky by se k výpočtu denní výnosnosti použily denní uzavírací ceny; výnosnosti za posledních 10, 30, 100, ... dní by se dosazovaly do uvedeného vzorečku. Anebo lze užít výnosnosti za kratší či delší období. Protože v odhadu volatility mají všechny výnosnosti stejnou váhu, projeví se jedna zvýšená výnosnost v odhadu volatility do doby, než uplyne 10 (nebo 30 nebo 100) dní. Pak se (odhadnutá) volatilita opět srovná, a je tak naprosto pochybná.

I přes oprávněnou kritiku různých heuristických přístupů se však praxe bez odhadu historické volatility obejde jen těžko. Není cílem tohoto článku hodnotit různé možnosti odhadu volatility; ze zkušenosti se ukazuje, že bychom jen těžko hledali jedinou univerzální metodu. Ukážeme však jeden možný postup, jak posuzovat délku historického

období  $M$ , které je rozumné při odhadu volatility brát v úvahu. Tento postup do jisté míry odbourává Wilmottovu námitku o „pochybném odhadu“ volatility.

## 2. Kdy došlo k poslední změně volatility?

Bud'te  $S_i$  uzavírací ceny sledovaného aktiva, např. akcie,  $i$ -tého dne ( $i = 0, 1, \dots, n$ ; přijmeme konvenci, že  $S_n$  je dnešní,  $S_{n-1}$  včerejší cena atd.). Nahlížejíme na ně jako na náhodné veličiny (a nikoliv jejich realizace), a předpokládáme dále, že denní logaritmické relativní změny

$$y_i = \ln \frac{S_i}{S_{i-1}}, \quad i = 1, 2, \dots, n$$

jsou nezávislé náhodné veličiny s normálním rozdělením se střední hodnotou  $\mu$ . Normalita je silný předpoklad; užijeme jej při odvození testu, a v třetí části se budeme zabývat situací, kdy nemáme normalitu zaručenu.

Na tomto místě poznamenejme, že k datům  $S_i$  přistupujeme z poněkud jiného úhlu pohledu, než nabízí tradiční ekonometrie. Často se chování řady  $S_i$  zkoumá např. pomocí (G)ARCH modelů. My postupujeme jiným způsobem: nezkoumáme  $S_i$ , nýbrž  $y_i$ , a v souladu s tradičními Black-Scholesovými předpoklady uvažujeme  $y_i$  normální nezávislé. Je otázkou, jakou informaci by tradiční instrumentárium pro analýzu časových řad *pro náš konkrétní problém* přineslo. (Autor si není vědom, že by byla známa verze Black-Scholesova modelu, která pracuje s nějakým obecnějším modelem podkladové proměnné, např. typu GARCH.)

Na první pohled je předpoklad nezávislosti  $y_i$  poněkud odvážný; přesto se ukazuje, že jej mnohé reálné finanční řady (s vysokou pravděpodobností) splňují. (Což je, bezpochyby, překvapivé.) Splňují jej (pravděpodobně) například i logaritmické relativní změny cen akcií Yahoo! na newyorské burze (ticker YHOO) v období 1996–2004. (Toto období zahrnuje jak nafouknutí, tak i následné splasknutí tzv. internetové bubliny v roce 2000. Tento příklad je více rozebrán v [4].) Splňují jej (pravděpodobně) i kurzy CZK/EUR, kterými se zabýváme ve čtvrté části.

Poznamenejme, že zde odvozené testy je možné, po vhodné modifikaci, užívat v situacích, kdy v  $y_i$  existují některé typy autoregrese; takový případ se v praxi může snadno vyskytnout. Blíže viz přehledovou práci [1]. Není snad ani třeba zdůrazňovat, že při práci s reálnými daty je vždy velmi důležité řádně prozkoumat, zdali data alespoň přibližně splňují předpoklady modelů a testů, které se k jejich analýze užívají.

Obecná metodika k detekci bodů změn, kterou se zde zabýváme, může být užitečná i při ekonometrické analýze časových řad. Bylo by jistě možné (a užitečné) konstruovat např. testy na existenci bodu změny v parametrech GARCH modelů, které se užívají při ekonometrické analýze časových řad. Jde však o jiný typ problému, než jakým se zabýváme v tomto textu.

Přistupme k odvození testu. Položme si otázku, zdali mají všechny veličiny  $y_i$  stejný rozptyl  $\sigma^2$

$$y_i \sim N(\mu, \sigma^2) \quad (H)$$

( $\mu$  a  $\sigma$  jsou neznámé parametry), anebo existuje okamžik  $\kappa$ , ve kterém dochází ke změně rozptylu

$$\begin{aligned} y_i &\sim N(\mu, \sigma_1^2) && \text{pro } i = 1, 2, \dots, \kappa \\ y_i &\sim N(\mu, \sigma_2^2) && \text{pro } i = \kappa+1, \kappa+2, \dots, n \end{aligned} \quad (A)$$

( $\mu, \sigma_1 > 0, \sigma_2 > 0$  a  $\kappa$  jsou neznámé parametry splňující  $\sigma_1 \neq \sigma_2$  a  $1 \leq \kappa < n$ ).

Jestliže  $y_i$  splňují model (H), nabízí se odhadovat volatilitu z celého datového souboru. Pokud však splňují model (A), je přirozené nejprve nalézt vhodný odhad  $\hat{\kappa}$  parametru  $\kappa$  (tzv. bod změny, *changepoint*), kdy dochází ke změně rozptylu. Okamžik  $\hat{\kappa}$  data rozdělí na dvě části, část „před změnou“ ( $i = 1, 2, \dots, \hat{\kappa}$ ) a část „po změně“ ( $i = \hat{\kappa}+1, \hat{\kappa}+2, \dots, n$ ). Při odhadu současné volatility je nyní přirozené část dat „před změnou“ ignorovat a volatilitu odhadovat pouze z dat „po změně“.

Zkonstruuje test hypotézy o existenci změny rozptylu v datech pomocí podílu věrohodností; konstrukce tohoto testu je dobře známa např. ze statistické kontroly jakosti (viz [1]). Nulovou hypotézou budiž (H) a alternativou (A).

Za platnosti (H) má sdružené rozdělení  $y_i$  tvar  $f_H = \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{y_i - \mu}{\sigma}\right)$  a za platnosti (A)

má tvar

$$f_A = \prod_{i=1}^{\kappa} \frac{1}{\sigma_1} \phi\left(\frac{y_i - \mu}{\sigma_1}\right) \cdot \prod_{i=\kappa+1}^n \frac{1}{\sigma_2} \phi\left(\frac{y_i - \mu}{\sigma_2}\right), \quad (3)$$

kde  $\phi(\cdot)$  je hustota  $N(0, 1)$ . Nejprve uvažme, že je parametr  $\kappa$  pevný; pak je logaritmus podílu věrohodností

$$\begin{aligned} L_{\kappa}(\mu, \sigma, \sigma_1, \sigma_2) &= \ln \frac{f_A}{f_H} = \ln \frac{\prod_{i=1}^{\kappa} \frac{1}{\sigma_1} \phi\left(\frac{y_i - \mu}{\sigma_1}\right) \cdot \prod_{i=\kappa+1}^n \frac{1}{\sigma_2} \phi\left(\frac{y_i - \mu}{\sigma_2}\right)}{\prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{y_i - \mu}{\sigma}\right)} = \\ &= \frac{1}{2} (n \ln \sigma^2 - \kappa \ln \sigma_1^2 - (n - \kappa) \ln \sigma_2^2) + \frac{\sum_{i=1}^{\kappa} (y_i - \mu)^2}{2\sigma^2} - \frac{\sum_{i=1}^{\kappa} (y_i - \mu)^2}{2\sigma_1^2} - \frac{\sum_{i=\kappa+1}^n (y_i - \mu)^2}{2\sigma_2^2}. \end{aligned}$$

Položíme-li derivace  $L_{\kappa}$  podle  $\mu, \sigma, \sigma_1$  a  $\sigma_2$  rovny nule, získáme maximálně věrohodné odhady

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \hat{\sigma} = \sqrt{\frac{1}{n} s_{1:n}^2}, \quad \hat{\sigma}_1 = \sqrt{\frac{1}{\kappa} s_{1:\kappa}^2}, \quad \hat{\sigma}_2 = \sqrt{\frac{1}{n - \kappa} s_{\kappa+1:n}^2},$$

kde jsme pro  $1 \leq i \leq j \leq n$  označili  $s_{i:j}^2 = \sum_{k=i}^j (y_k - \hat{\mu})^2$ . Dosazením do  $L_{\kappa}$  dostáváme

$$\begin{aligned} L_{\kappa}(\hat{\mu}, \hat{\sigma}, \hat{\sigma}_1, \hat{\sigma}_2) &= \frac{1}{2} \left( n \ln \frac{s_{1:n}^2}{n} - \kappa \ln \frac{s_{1:\kappa}^2}{\kappa} - (n - \kappa) \ln \frac{s_{\kappa+1:n}^2}{n - \kappa} \right) + \underbrace{\frac{s_{1:n}^2}{2} - \frac{s_{1:\kappa}^2}{2} - \frac{s_{\kappa+1:n}^2}{2}}_{=0} = \\ &= \frac{1}{2} \left[ \kappa \ln \left( \frac{\kappa}{n} \frac{s_{1:n}^2}{s_{1:\kappa}^2} \right) + (n - \kappa) \ln \left( \frac{n - \kappa}{n} \frac{s_{1:n}^2}{s_{\kappa+1:n}^2} \right) \right] =: \frac{1}{2} Z_{\kappa}^2. \end{aligned}$$

Opusťme nyní předpoklad pevného  $\kappa$ ; maximalizací  $L_\kappa$  přes  $\kappa$  získáme statistiku

$$\max_{1 \leq k < n} \frac{1}{2} Z_k^2. \quad (4)$$

Z technických důvodů je jednodušší pracovat s ekvivalentní statistikou

$$\mathbf{Z}_n := \max_{1 \leq k < n} \sqrt{Z_k^2},$$

která rozhoduje stejně jako (4) (má ale jiné kritické hodnoty). Hypotéza ( $H$ ) o neexistenci bodu změny se zamítne, jestliže  $\mathbf{Z}_n$  překročí příslušnou kritickou hodnotu  $Q_n^{[\alpha]}$  (na předepsané hladině  $\alpha$ ). Konstrukce kritických hodnot  $Q_n^{[\alpha]}$  je (obtížný) problém, kterému se budeme krátce věnovat ve třetí části.

Tvar statistiky  $\mathbf{Z}_n$  přímo nabízí jako odhad neznámého bodu změny

$$\hat{\kappa} := \arg \max_{1 \leq k < n} \sqrt{Z_k^2}. \quad (5)$$

Lze ukázat, že  $\hat{\kappa}$  je za platnosti ( $A$ ) konsistentním odhadem skutečného bodu změny  $\kappa$ ; dokonce je známo jeho (asymptotické) rozdělení, viz např. [1]. V našem kontextu je též přirozené pracovat s okamžikem

$$K = \max \left\{ k : \sqrt{Z_k^2} > Q_n^{[\alpha]} \right\}; \quad (6)$$

okamžik  $K$  je nejmladší okamžik v minulosti, kdy proces  $\left( \sqrt{Z_k^2} \right)_{k=1,2,\dots,n-1}$  opouští kritický obor  $Q_n^{[\alpha]}$ , takže v tomto smyslu lze období  $K+1, K+2, \dots, n$  chápat jako období, kde změna rozptylu (s vysokou pravděpodobností) není detekována.

Hypotézu ( $H$ ) jsme formulovali jako „neexistuje okamžik změny rozptylu“ a alternativu jako „existuje jeden okamžik změny rozptylu“. Je možné konstruovat testy s obecnějšími alternativami: „existuje  $q$  změn rozptylu“, kde  $q$  je známé, anebo „existuje  $q \leq q_{\max}$  změn“, kde  $q$  je neznámé a  $q_{\max}$  známé. V takovém obecnějším modelu se otvírají další otázky – kolik existuje bodů změn, kde se tyto změny nacházejí atd. Pro náš problém však nejsou tolik důležité; nás především zajímá, zdali došlo k aspoň jedné změně a kdy došlo k poslední změně. Zůstaneme u naší jednoduché alternativy; odvozený test totiž dobře funguje i v případě, kdy změn existuje více.

K odhadu délky období, z něž se odhaduje volatilita, se přímo nabízejí dvě metody.

**Algoritmus 1.** Na vstupu je  $(n; y_1, y_2, \dots, y_n)$ . Spočtíme některou z metod z třetí části  $Q_n^{[\alpha]}$  a  $\hat{\kappa}$  podle (5). Jestliže  $\sqrt{Z_{\hat{\kappa}}^2} \leq Q_n^{[\alpha]}$ , algoritmus končí (v datech již nebyla žádná změna rozptylu detekována); jinak polož  $y'_1 := y_{\hat{\kappa}+1}$ ,  $y'_2 := y_{\hat{\kappa}+2}$ , ...,  $y'_{n-\hat{\kappa}} := y_n$  a celý postup opakuj s  $(n-\hat{\kappa}; y'_1, y'_2, \dots, y'_{n-\hat{\kappa}})$ .

**Algoritmus 2.** Na vstupu je  $(n; y_1, y_2, \dots, y_n)$ . Spočtíme  $Q_n^{[\alpha]}$  některou z metod z třetí části,  $\hat{\kappa}$  podle (5) a  $K$  podle (6). Jestliže  $\sqrt{Z_{\hat{\kappa}}^2} \leq Q_n^{[\alpha]}$ , algoritmus končí (v datech již nebyla žádná změna rozptylu detekována); jinak polož  $y'_1 := y_{K+1}$ ,  $y'_2 := y_{K+2}$ , ...,  $y'_{n-K} := y_n$ , a celý postup opakuj s  $(n-K; y'_1, y'_2, \dots, y'_{n-K})$ .

Oba popsané postupy mají stále jednu nevýhodu. Chování  $\mathbf{Z}_n$  může být totiž významně ovlivněno dlouhou historií dat, což v naší situaci může být nepříjemné. Existuje přirozený způsob, jak s tímto problémem naložit: dívat se na nejprve na data nejmladší, a postupně řadu prodlužovat a přidávat data starší. Položme proto

$$z_i = y_{n-i+1}, \quad i = 1, 2, \dots, n \quad (7)$$

a postupně pro  $j = 4, 5, \dots, n$  (i když v praxi je patně rozumné začínat s  $j$  větším než 4) hledáme bod změny v  $z_1, z_2, \dots, z_j$ . Hledáme tak vlastně nejmladší bod změny rozptylu. Zformulujeme tento postup přesněji jako další algoritmus.

**Algoritmus 3.** Na vstupu je  $(n; y_1, y_2, \dots, y_n)$ ;  $z_i$  nechť mají stejný význam jako v (7). Polož  $j := 4$  a proveď následující:

*Krok 1.* Spočti  $\mathbf{Z}_n$  a  $\hat{\kappa}_j$  z posloupnosti dat  $z_1, z_2, \dots, z_j$  a spočti  $Q_j^{[a]}$  ( $\hat{\kappa}_j$  je odhad bodu změny v posloupnosti  $z_1, z_2, \dots, z_j$ ).

*Krok 2.* Je-li  $\mathbf{Z}_n > Q_j^{[a]}$ , na výstup zapiš  $y_{n-\hat{\kappa}+1}, y_{n-\hat{\kappa}+2}, \dots, y_n$  a skonči.

*Krok 3.* Je-li  $j = n$ , na výstup zapiš  $y_1, y_2, \dots, y_n$  a skonči.

*Krok 4.* Polož  $j := j + 1$  a pokračuj krokem 1.

V kroku 1 a 2 je možné místo  $\hat{\kappa}$  užít i  $K$  ve stejné roli, jakou hrálo v algoritmu 2.

Nabízejí se i další postupy; je možné se inspirovat například v oblasti on-line kontroly jakosti. Statistika kontroly jakosti zkoumá (mimo jiné) tento problém: nechť nám přicházejí nová a nová pozorování jisté veličiny, jejíž rozdělení známe. V jistém neznámém okamžiku dojde ke změně rozdělení (typicky: změni se některý parametr rozdělení, v našem případě rozptyl normálního rozdělení). Jak co nejdříve poté, co ke změně došlo, s dostatečnou mírou jistoty poznat, že ke změně došlo? Více o této problematice viz práci [2]. Zde jen poznamenejme, že naše situace je jednodušší – při práci s posloupnostmi  $z_1, z_2, \dots, z_j$  s rostoucím  $j$  postupujeme „proti směru času“, a tak rozhodovat o změně v rozptylu *co nejdříve po jejím výskytu* není nutné. Stačí, pokud změnu vůbec někdy poznáme.

Aby byl popis metody úplný, zbývá rozebrat možnosti výpočtu  $Q_j^{[a]}$ ; tomu věnujeme třetí část textu. Na závěr této části ještě doplníme několik poznámek. Ukázali jsme jen jeden, velmi jednoduchý způsob, jak zkonstruovat test hypotézy „neexistuje bod změny rozptylu“ a s jeho pomocí i bod změny odhadnout. Možností však existuje celá řada. Zmiňme zde alespoň jeden důležitý postup, který využívá bayesovský pohled na alternativní hypotézu. O bodu změny  $\kappa$  předpokládáme, že jde o náhodnou veličinu s rovnoměrným apriorním rozdělením. Rozdělení (3) by pak bylo rozdělení podmíněné daným  $\kappa$ . Běžnou konstrukcí podle věty o úplné pravděpodobnosti se získá nepodmíněné rozdělení  $y_1, y_2, \dots, y_n$  za platnosti alternativy a testové kritérium se odvodí podílem věrohodností. Výsledkem je statistika, jež má mírně odlišné vlastnosti: lépe než  $\mathbf{Z}_n$  detekuje změny, které se vyskytují „uvnitř“ datového souboru a o něco hůře detekuje změny, které se vyskytují „na kraji“.

Pracovali jsme v modelu, kde drift  $\mu$  je konstantní, a zkoumáme jen změnu volatility; je však samozřejmě možné konstruovat i statistiky k testování změny jen v driftu a statistiky k testování změny v driftu a/nebo ve volatilitě. K dispozici jsou i výsledky pro závislá data. Problematika bodů změn se v posledních letech intenzivně studuje, viz např. přehledovou práci [1].

### 3. Kritické hodnoty $Q_j^{[\alpha]}$

Explicitně vyjádřit rozdělení  $\mathbf{Z}_n$  je mimořádně obtížné; k dispozici jsou pouze aproximace a asymptotické výsledky.

**Asymptotika.** Asymptoticky s  $n \rightarrow \infty$  má za platnosti  $(H)$  veličina  $\mathbf{Z}_n$  degenerované rozdělení. Lze však ukázat [1], že veličina  $a_n \mathbf{Z}_n - b_n$ , kde  $a_n = \sqrt{2 \ln \ln n}$  a  $b_n = \sqrt{2 \ln \ln n} + \ln \sqrt{\ln \ln n} - \ln \sqrt{\pi}$ , má v limitě  $n \rightarrow \infty$  distribuční funkci extrémálního typu:

$$\Pr[a_n \mathbf{Z}_n - b_n > x] \rightarrow e^{-2e^{-x}}. \quad (8)$$

Odtud (pro velká  $n$ ) získáme kritické hodnoty ve tvaru

$$Q_n^{[\alpha]} = \frac{2 \ln \ln n + \ln \sqrt{\ln \ln n} - \ln \sqrt{\pi} - \ln \ln \frac{1}{\sqrt{1-\alpha}}}{\sqrt{2 \ln \ln n}}. \quad (9)$$

Zajímavé je, že (8) zůstává v platnosti i za podstatně slabších předpokladů než  $(H)$ ; vůbec není třeba předpokládat normalitu. Stačí, jsou-li  $y_i$  nezávislé, stejně rozdělené a pro některé  $d > 4$  je  $\mathbb{E}|y_i|^d < \infty$  (viz [1]). Takže (9) se vyplatí používat v situacích, kdy nemáme dostatečnou jistotu, že  $y_i$  jsou normální. Je však třeba upozornit na to, že vztah je použitelný pouze pro velká  $n$ , přičemž velmi záleží na zvolené hladině testu  $\alpha$ : pro  $\alpha = 5\%$  je vztah (9) použitelný přibližně již pro  $n \approx 30$ , zatímco pro  $\alpha = 1\%$  až přibližně od  $n \approx 100$ ; i tak ale (9) dává značně konzervativní kritické hodnoty.

**Simulace.** Pro malá  $n$  pomohou simulace; kritická hodnota nezávisí na rozptylu  $\sigma$  ani na střední hodnotě  $\mu$ , takže je nejjednodušší simulovat data z  $N(0, 1)$ . Předpoklad normality v  $(H)$  je však poměrně svazující; lze proto doporučit i simulaci z empirického rozdělení  $y_i$ .

**Permutace.** Velmi oblíbenou metodou, která využívá empirického rozdělení  $y_i$ , a osvědčuje se i v případě ne-normálních dat, je užití permutačního principu (bližší viz [5]). Je založen na pozorování, že splňuje-li posloupnost  $y_1, y_2, \dots, y_n$  model  $(H)$ , splňuje jej i posloupnost  $y_{P(1)}, y_{P(2)}, \dots, y_{P(n)}$ , kde  $P$  je libovolná permutace množiny  $\{1, 2, \dots, n\}$ . Naproti tomu, splňuje-li posloupnost  $y_1, y_2, \dots, y_n$  model  $(A)$ , pak zvolíme-li permutaci  $P$  náhodně, s vysokou pravděpodobností posloupnost  $y_{P(1)}, y_{P(2)}, \dots, y_{P(n)}$  model  $(A)$  splňovat nebude. Označme

$$\mathbf{Z}_n^{[P]} := \max_{1 \leq k \leq n} \left[ k \ln \left( \frac{k \cdot \sum_{i=1}^n (y_{P(i)} - \hat{\mu})^2}{n \cdot \sum_{i=1}^k (y_{P(i)} - \hat{\mu})^2} \right) + (n-k) \ln \left( \frac{(n-k) \cdot \sum_{i=1}^n (y_{P(i)} - \hat{\mu})^2}{n \cdot \sum_{i=k+1}^n (y_{P(i)} - \hat{\mu})^2} \right) \right],$$

kde  $P$ ; je vlastně hodnota statistiky  $\mathbf{Z}_n^{[P]}$ , kdybychom bychom ji počítali z dat  $y_{P(1)}, y_{P(2)}, \dots, y_{P(n)}$ . Necht' jsou nyní  $y_1, y_2, \dots, y_n$  pevná; permutační distribuční funkcí je funkce

$$F(x) = \Pr[\mathbf{Z}_n^{[P]} < x],$$

kde  $P$  je náhodná permutace. (Jde o rozdělení  $\mathbf{Z}_n^{[P]}$  při náhodné permutaci  $P$  podmíněné danými  $y_1, y_2, \dots, y_n$ .) Kvantil tohoto rozdělení, tzv. *permutační kvantil*, nyní poslouží jako odhad  $Q_n^{[\alpha]}$ .



Přesně spočítat permutační distribuční funkci  $F(x)$ , a tedy i hledaný kvantil, vyžaduje obrovské výpočetní úsilí (všech permutací je  $n!$ ); ukazuje se však, že velmi dobrou aproximací  $F(x)$  je empirická permutační distribuční funkce

$$\frac{\left| \left\{ j \in \{1, 2, \dots, N\} : \mathbf{Z}_n^{[P_j]} < x \right\} \right|}{N},$$

kde  $P_j, j = 1, 2, \dots, N$ , jsou náhodné permutace množiny  $\{1, 2, \dots, n\}$ , již pro  $N$  řádu  $10^4$ . Pro tuto aproximaci permutačního kvantilu vystačíme již s přijatelným výpočetním úsilím; stačí generovat  $N$  náhodných permutací, a z jejich empirické permutační distribuční funkce odhadnout hledaný kvantil.

Permutační rozdělení je rovněž vhodné v situacích, kdy nemáme zaručenu normalitu pozorovaných dat.

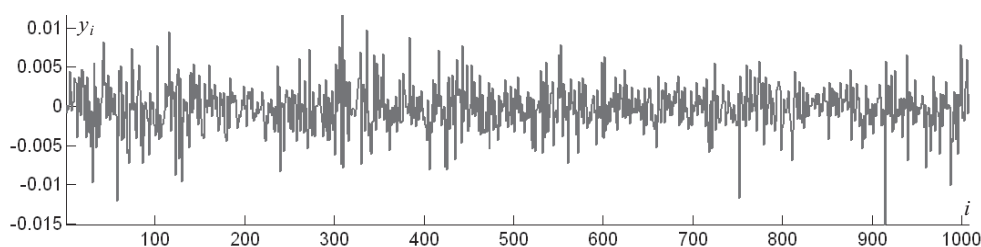
Výčet není zdaleka vyčerpávající; existují i další metody aproximace kritických hodnot. Následující tabulka shrnuje kritické hodnoty testu získané simulacemi pro normálně rozdělené náhodné veličiny.

	$n = 10$	$n = 50$	$n = 100$	$n = 300$	$n = 600$	$n = 1000$	$n = 2000$
$\alpha = 10\%$	2,56	2,83	2,90	3,02	3,06	3,08	3,12
$\alpha = 5\%$	2,83	3,09	3,16	3,27	3,32	3,34	3,36
$\alpha = 2,5\%$	3,09	3,32	3,39	3,49	3,54	3,53	3,59
$\alpha = 1\%$	3,40	3,59	3,67	3,76	3,81	3,83	3,80
$\alpha = 0,5\%$	3,64	3,75	3,83	3,95	4,02	4,02	4,05

## 4. Ilustrace

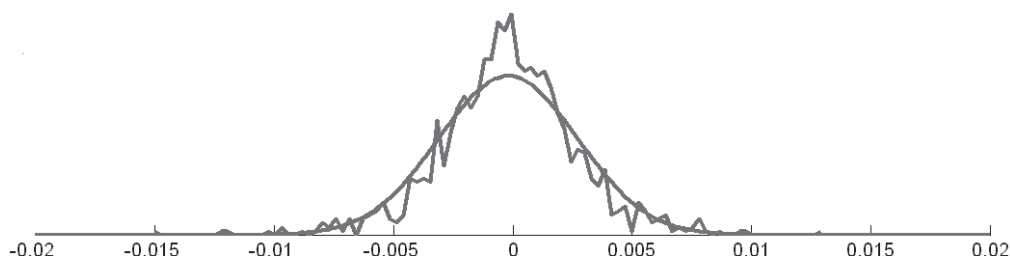
V této části demonstrujeme navrženou metodu na příkladu veličiny, jejíž volatilitu je v České republice třeba při oceňování derivátů odhadovat patrně nejčastěji: měnový kurz CZK/EUR. Denní data  $S_i$  jsou běžně dostupná (slovem *den* nadále rozumíme obchodní den). Pracujme s obdobím 2. 1. 2004–31. 12. 2007, což je pro naše účely dostatečně dlouhá historie; celkem jde o  $n = 1009$  pozorování ( $S_1$  je kurz k 2. 1. 2004 a  $S_{1009}$  je kurz k 31. 12. 2007). Výsledky by však byly stejné, i kdybychom pracovali s delším historickým obdobím. Následující obrázek zachycuje logaritmické relativní denní změny

$$y_i = \ln \frac{S_{i+1}}{S_i}, \quad i = 1, 2, \dots, 1008.$$





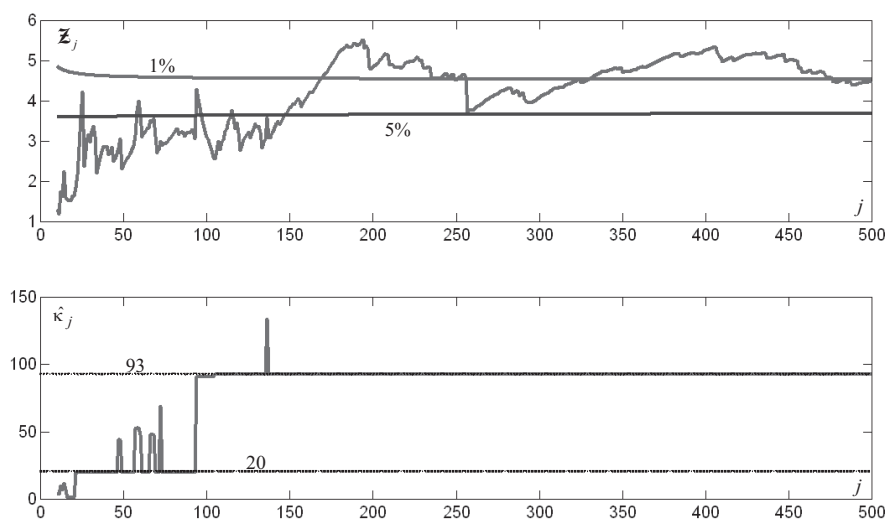
Není naším cílem podat detailní analýzu této časové řady; jen stručně konstatujeme, že na základě rozboru odhadu autokorelační funkce se ukazuje, že řada  $y_i$  autokorelaci nevykazuje. To je důležité, neboť náš test je na autokorelaci citlivý. Pro zajímavost uvádíme i empirické rozdělení  $y_i$  a jeho srovnání s normálním; to je důležité při rozhodnutí, kterou metodu užít ke konstrukci kritických hodnot.



Následující obrázek ilustruje algoritmus 3;  $j$  má zde stejný význam jako při popisu algoritmu. Jsme v situaci, kdy 31. 12. 2007 hledáme odhad nejmladšího okamžiku změny rozptylu; data za období počínaje odhadnutým okamžikem změny až po 31. 12. 2007 bychom pak použili k odhadu volatility. Horní graf následujícího obrázku ukazuje trajektorii  $\mathbf{z}_j$  (viz krok 1 algoritmu 3) postupně pro  $j = 10, 11, \dots, 500$ ; je zde zachycena i kritická hodnota spočtená na základě vztahu (9) na 1 % a 5 % hladině. Jak je z obrázku vidět, při  $\alpha = 5\%$  je kritická hodnota poprvé překročena pro  $j = 25$  (tedy: již při pohledu na 25 nejmladších hodnot  $y_i$  je detekována existence statisticky významné změny) a odhad bodu změny  $\hat{\kappa}_j$  z kroku 1 algoritmu 3 (dolní obrázek) je 20, což vede k závěru, že pro odhad volatility bychom uvažovali 20denní historii.

Pohled na dolní obrázek, kde je zachycen odhad bodu změny v závislosti na  $j$ , ukazuje, že existují dva „obory stability“ odhadu bodu změny: hledíme-li pouze na nedávnou historii, řekněme  $j = 22, 23, \dots, 90$ , nabývá  $\hat{\kappa}_j$  hodnoty 20 (až na několik výjimek). Při pohledu na delší historii, řekněme  $j \geq 100$ , je hodnota  $\hat{\kappa}_j$ , až na jednu výjimku, rovna 93.

Pokud bychom pracovali s  $\alpha = 1\%$ , kritická hodnota bude poprvé překročena pro  $j = 170$ , což nabízí k odhadu volatility užít historii  $\hat{\kappa}_{170} = 93$  dnů (viz dolní obrázek). Zvětšujeme-li  $j$  (tedy: hledíme dále do minulosti), zůstává odhad bodu změny  $\hat{\kappa}_j$  stabilní.



## Literatura

- [1] ANTOCH, J.; HUŠKOVÁ, M.; JARUŠKOVÁ, D. 2002. Off-line Statistical Process Control. In *Multivariate Total Quality Control*. Heidelberg : Physica-Verlag, 2002, s. 1–86.
- [2] ANTOCH, J.; JARUŠKOVÁ, D. 2002. On-line Statistical Process Control. In *Multivariate Total Quality Control*. Heidelberg : Physica-Verlag, 2002, s. 87–124.
- [3] CIPRA, T. 1986. *Analýza časových řad s aplikacemi v ekonomii*. Praha : SNTL/Alfa, 1986.
- [4] ČERNÝ, M. 2008. *Detekce a analýza změn v ekonometrických modelech* ( disertační práce ). VŠE Praha : 2008.
- [5] HUŠKOVÁ, M.; ANTOCH, J. 2003. *Detection of Structural Changes in Regression*. Tatra Mountains Mathematical Publications, vol. 26, no. 2 (2003), s. 201–215.
- [6] MÁLEK, J. 2005. *Dynamika úrokových měr a úrokové deriváty*. Praha : Ekopress, 2005.
- [7] TALEB, N. 1997. *Dynamic Hedging: Managing Vanilla and Exotic Options*. Wiley&Sons, 1997.
- [8] WILMOTT, P. 2005. *Paul Wilmott Introduces Quantitative Finance*. Wiley&Sons, 2005.

## On Estimation of Volatility of Financial Time Series for Pricing Derivatives

### Abstract

Estimation of volatility of financial time series plays a crucial role in pricing derivatives. Volatility is often estimated from historical data; however, it is well known that volatility varies in time. We propose a method to choose a suitable length of historical data to estimate contemporary volatility. The method is based on adaptation of a procedure used in statistical quality control – a hypothesis, that data contains a changepoint of volatility, is tested and if the test gives a positive answer, the changepoint is estimated. Then, a period of data where no changepoint is statistically significant is used to estimate contemporary volatility. The approach is illustrated on an analysis of CZK/EUR exchange rates.

**Keywords:** Derivative, Black-Scholes model, time series, volatility, changepoint.

**JEL classification:** G12, C12, C13.