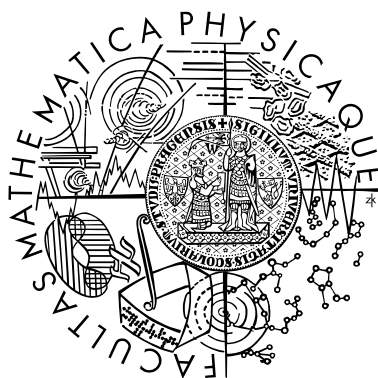


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Jakub Mikulka

Exponenciální vyrovnávání

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Tomáš Hanzák

Studijní program: Matematika, Obecná matematika

2008

Na tomto místě bych chtěl poděkovat vedoucímu své bakalářské práce Mgr. Tomáši Hanzákovi za výběr zajímavého tématu, trpělivost, ochotu a cenné rady při vzájemné spolupráci.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 5. 8. 2008

Jakub Mikulka

Obsah

1	Úvod	5
2	Exponenciální vyrovnávání	7
2.1	Brownova metoda	7
2.2	Holtova metoda	9
2.3	Vztah mezi Brownovou a Holtovou metodou	13
3	Citlivost Holtovy metody na změny vyrovnávacích konstant	20
3.1	Výpočet MSE.....	20
3.2	Výpočet ρ	26
4	Aplikace na Brownovu metodu	29
5	Numerické aplikace	36
5.1	Ověření vzorců pomocí simulace	36
5.2	Vyzkoušení vzorců na reálných časových řadách	37
6	Závěr	39
	Literatura	40

Abstrakt

Název práce: Exponenciální vyrovnávání

Autor: Jakub Mikulka

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Tomáš Hanzák

e-mail vedoucího: hanzak@karlin.mff.cuni.cz

Abstrakt: Práce se zabývá dvěma metodami exponenciálního vyrovnávání pro nesezónní časové řady s lokálně lineárním trendem: Holtově metodě a dvojitému exponenciálnímu vyrovnání (Brownově metodě). Je ukázáno, že Brownova metoda je speciálním případem Holtovy metody. Dále je uveden vztah procesu ARIMA(0, 2, 2) a Holtovy metody. Hlavní částí práce je teoretické odvození hodnoty MSE a autokorelačního koeficientu předpovědních chyb ϱ při použití Holtovy metody pro všechny kombinace jejích vyrovnávacích konstant za předpokladu generování řady procesem ARIMA(0, 2, 2) pro všechny hodnoty jeho parametrů. Odvozené teoretické vzorce jsou aplikovány též na Brownovu metodu. Odvozené vzorce jsou pomocí simulací ověřeny a vyzkoušeny na reálných časových řadách. Jsou formulovány praktické závěry týkající se obou metod.

Klíčová slova: autokorelační koeficient předpovědních chyb, Holtova metoda, dvojitě exponenciální vyrovnávání, MSE, vyrovnávací konstanty

Abstract

Title: Exponential smoothing

Author: Jakub Mikulka

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Tomáš Hanzák

Supervisor's e-mail address: hanzak@karlin.mff.cuni.cz

Abstract: The thesis deals with two exponential smoothing type methods for non-seasonal time series with local linear trend: Holt method and double exponential smoothing (Brown method). It is shown that Brown method is a special case of Holt method. The relationship between Holt method and ARIMA(0, 2, 2) process is presented. The main part of the thesis is a theoretical derivation of MSE and autocorrelation coefficient of forecasting errors ϱ when using Holt method with all combinations of smoothing constants and with the time series generated by ARIMA(0, 2, 2) process with all combinations of its parameters. Then the theoretically derived formulae are applied also to Brown method. The derived formulae are verified via simulations and tried on real time series. The practical conclusions related to both methods are formulated.

Keywords: autocorrelation coefficient of forecasting errors, double exponential smoothing, Holt method, MSE, smoothing constants

1 Úvod

Metody exponenciálního vyrovnávání jsou velice často a s úspěchem v praxi používané metody k analýze časových řad. V této práci se budeme zabývat výhradně nesezónními časovými řadami s lokálně lineárním trendem, pro něž jsou vhodné zejména dvě metody exponenciálního vyrovnávání: Holtova metoda a dvojité exponenciální vyrovnávání neboli Brownova metoda.

Specifickou a doposud poněkud opomíjenou kapitolou v problematice používání těchto metod v praxi je analýza citlivosti metod na volbu vyrovnávacích konstant, které slouží k nastavení toho, jak rychle se má každé nové pozorování zahrnout do směrnice trendu a úrovně vyrovnané řady. Nesprávná volba vyrovnávacích konstant přitom může vést k neuspokojivým výsledkům při použití jedné z metod.

V úvodních odstavcích 2.1 a 2.2 si představíme Holtovu i Brownovu metodu a zavedeme dále používané značení. Odvodíme zde vztah mezi Holtovou metodou a procesem ARIMA(0, 2, 2) i vícerozměrným procesem VAR(1).

V odstavci 2.3 si poté odvodíme vztah mezi Holtovou a Brownovou metodou, a sice, že Brownova metoda je speciálním případem Holtovy metody. V literatuře se proto často považuje za zbytečné Brownovu metodu vůbec používat. Holtova metoda je zřejmě flexibilnější, protože používá dvě na sobě nezávislé vyrovnávací konstanty, kdežto Brownova pouze jednu. Dále se proto budeme snažit odvodit, jak velké chyby se dopustíme, budeme-li používat Brownovu metodu, jejíž výhodou je jednodušší volba vyrovnávací konstanty. Na závěr odstavce si ukážeme, jak vypadají řady generované procesem ARIMA(0, 2, 2).

Odstavce 3.1 a 3.2 jsou hlavní částí celé práce. Budeme zde předpokládat, že časová řada byla generována procesem ARIMA(0, 2, 2) s různými kombinacemi hodnot parametrů procesu. Za tohoto předpokladu odvodíme teoretické vzorce pro výpočet hodnoty MSE a koeficientu autokorelace předpovědních chyb q při použití Holtovy metody pro všechny kombinace přípustné hodnoty vyrovnávacích konstant Holtovy metody a zároveň pro všechny přípustné kombinace hodnot parametrů procesu ARIMA(0, 2, 2) generujícího danou časovou řadu. Uvedeme také ilustrativní grafické zobrazení odvozených vzorců pro konkrétní hodnoty parametrů procesu ARIMA(0, 2, 2).

Dále využijeme odvozených teoretických vzorců z odstavců 3.1 a 3.2 při analýze Brownovy metody (kapitola 4). Zkoumáme zde velikost hodnoty MSE a koeficientu autokorelace předpovědních chyb při použití Brownovy metody s optimální vyrovnávací konstantou alfa na časovou řadu generovanou náhodným procesem ARIMA(0, 2, 2) s libovolnou kombinací hodnot parametrů daného procesu. Výsledky

budou opět graficky prezentovány. V závěru kapitoly formulujeme některé praktické závěry při práci s Holtovou a Brownovou metodou vycházející z výsledků, ke kterým jsme dospěli v kapitolách 3 a 4.

V kapitole 5 jsou teoreticky odvozené vzorce z odstavců 3.1 a 3.2 ověřeny na řadách vygenerovaných procesem $ARIMA(0, 2, 2)$. Dále jsou také vyzkoušeny na reálných časových řadách, aby bylo zjištěno, zda jsou naše výsledky relevantní i v praxi. Je uvedeno srovnání výstupů statistického softwaru a výstupů spočtených za pomoci námi odvozených vzorců.

2 Exponenciální vyrovnávání

Metody známé jako *exponenciální vyrovnávání* byly vyvinuty v 50. letech především k předpovídání budoucích prodejů zboží viz Winters (1960). Myšlenku použít exponenciálního vážení k odhadu úrovně časové řady i jejího trendu a sezónní složky, což následovně umožní předpovídat její budoucí hodnoty, publikoval jako první Američan Charles C. Holt v roce 1957 ve svém memorandu pro Office of Naval Research.

Dodnes je exponenciální vyrovnávání velice populární a v praxi často využívanou metodou předpovídání časových řad s adaptivním přístupem k trendové a sezónní složce. Jeho výhodou je, že využívá všechna známá pozorování řady, přičemž jako nejdůležitější bere v čase nejmladší pozorování. Na popularitě mu také přidává jeho výpočetní a paměťová nenáročnost při použití počítače.

Nejjednodušší variantou exponenciálního vyrovnávání je jednoduché exponenciální vyrovnávání, které se používá pro nesezónní časové řady s lokálně konstantním trendem. Zobecněním této metody dostáváme nejdříve dvojité a dále trojitě exponenciální vyrovnávání pro nesezónní časové řady s lokálně lineárním respektive kvadratickým trendem. Pro nesezónní časové řady s lokálně lineárním trendem byla také navržena Holtova metoda. Holt-Wintersova metoda si navíc poradí i se sezónními časovými řadami. Existuje ještě mnoho dalších metod, v naší práci se však omezíme pouze na nesezónní časové řady s lokálně lineárním trendem, k jejichž předpovídání se používají metoda dvojitěho exponenciálního vyrovnávání (neboli Brownova metoda) a Holtova metoda.

2.1 Brownova metoda

V tomto odstavci popíšeme dvojité exponenciální vyrovnávání nazývané také Brownova metoda. Předpokládejme nyní, že máme nesezónní časovou řadu y , která vypadá takto

$$y_n = Tr_n + E_n, \quad (2.1.1)$$

kde Tr_n je trendová a E_n je residuální složka řady y .

Brownova metoda se používá na časové řady s lokálně lineárním trendem. Předpokládejme tedy dále, že v krátkých úsecích lze trendovou složku pokládat za lineární, tj.

$$Tr_{n-j} = b_0 + b_1(-j). \quad (2.1.2)$$

Odhady parametrů $b_0 \in \mathbb{R}$ a $b_1 \in \mathbb{R}$ v čase n označme $\hat{b}_0(n)$ a $\hat{b}_1(n)$. Odhadněme nyní tyto neznámé parametry metodou DLS (*discounted least squares*) s *diskontním faktorem* $\beta = 1 - \alpha$, kde $\alpha \in (0,1)$ je pevně zvolená *vyrovnávací konstanta*. Běžnou metodu nejmenších čtverců zde modifikujeme tak, že se váhy jednotlivých čtverců směrem do minulosti exponenciálně zmenšují. Minimalizujeme tedy výraz

$$\sum_{j=0}^{\infty} \beta^j (y_{n-j} - b_0 + b_1 j)^2. \quad (2.1.3)$$

Když výraz (2.1.3) parciálně derivujeme podle b_0 a b_1 a derivace položíme rovny nule, dostaneme soustavu normálních rovnic pro b_0 a b_1 ve tvaru

$$\begin{aligned} \sum_{j=0}^{\infty} \beta^j (y_{n-j} - b_0 + b_1 j) &= 0, \\ \sum_{j=0}^{\infty} j \beta^j (y_{n-j} - b_0 + b_1 j) &= 0. \end{aligned} \quad (2.1.4)$$

Použijeme-li součtové vzorce

$$\sum_{j=0}^{\infty} \beta^j = \frac{1}{1-\beta}, \quad \sum_{j=0}^{\infty} j \beta^j = \frac{\beta}{(1-\beta)^2} \quad \text{a} \quad \sum_{j=0}^{\infty} j^2 \beta^j = \frac{\beta(1+\beta)}{(1-\beta)^3}, \quad (2.1.5)$$

přepíšeme po jednoduchých algebraických úpravách soustavu (2.1.4) do tvaru

$$\begin{aligned} b_0 - \frac{\beta}{1-\beta} b_1 &= (1-\beta) \sum_{j=0}^{\infty} y_{n-j} \beta^j, \\ \beta b_0 - \frac{\beta(1+\beta)}{1-\beta} b_1 &= (1-\beta)^2 \sum_{j=0}^{\infty} y_{n-j} j \beta^j. \end{aligned} \quad (2.1.6)$$

Definujme nyní *jednoduchou vyrovnávací statistiku*

$$S_n^{[1]} = (1-\beta) \sum_{j=0}^{\infty} y_{n-j} \beta^j \quad (2.1.7)$$

a podobně *dvojitou vyrovnávací statistiku*

$$S_n^{[2]} = (1-\beta) \sum_{j=0}^{\infty} S_{n-j} \beta^j. \quad (2.1.8)$$

Pro vyrovnávací statistiky platí následující rekurentní formule:

$$S_{n+1}^{[1]} = (1 - \alpha) \cdot S_n^{[1]} + \alpha \cdot y_{n+1}, \quad (2.1.9)$$

$$S_{n+1}^{[2]} = (1 - \alpha) \cdot S_n^{[2]} + \alpha \cdot S_{n+1}^{[1]}. \quad (2.1.10)$$

Pomocí těchto dvou vyrovnávacích statistik lze soustavu (2.1.6) zapsat do tvaru

$$\begin{aligned} b_0 - \frac{\beta}{1 - \beta} b_1 &= S_n^{[1]}, \\ \beta b_0 - \frac{\beta(1 + \beta)}{1 - \beta} b_1 &= S_n^{[2]} - S_n^{[1]}. \end{aligned} \quad (2.1.11)$$

Řešením této soustavy získáme hledané odhady jako

$$\begin{aligned} \hat{b}_0(n) &= 2S_n^{[1]} - S_n^{[2]}, \\ \hat{b}_1(n) &= \frac{\alpha}{1 - \alpha} (S_n^{[1]} - S_n^{[2]}). \end{aligned} \quad (2.1.12)$$

Předpověď hodnoty $y_{n+\tau}$ z času n pro předpovědní horizont $\tau > 0$ potom získáme ve tvaru

$$\hat{y}_{n+\tau}(n) = \hat{b}_0(n) + \tau \hat{b}_1(n) = \left(2 + \frac{\alpha\tau}{1 - \alpha}\right) S_n^{[1]} - \left(1 + \tau \frac{\alpha}{1 - \alpha}\right) S_n^{[2]}. \quad (2.1.13)$$

Pro $\tau = 0$ speciálně dostaneme vyrovnanou hodnotu řady y v čase n jako

$$\hat{y}_n = \hat{b}_0(n) = 2S_n^{[1]} - S_n^{[2]}. \quad (2.1.14)$$

Jako všechny metody založené na exponenciálním vyrovnávání, je i Brownova metoda evidentně rekurentní metodou, viz (2.1.9) a (2.1.10), proto musíme vždy nejprve určit počáteční hodnoty statistik $S_0^{[1]}$ a $S_0^{[2]}$. Ty se obvykle počítají ze vzorců (2.1.11), kdy za $\hat{b}_0(n)$ a $\hat{b}_1(n)$ klademe regresní odhady těchto parametrů, které získáme proložení nějakého počátečního úseku řady přímkou.

Na závěr se ještě zmiňme o volbě vyrovnávací konstanty α . Tutu volbu lze provést kupříkladu tak, že se postupně volí různé hodnoty $\alpha \in (0,1)$ a vybere se ta hodnota, pro kterou metoda poskytuje nejlepší předpovědi například podle MSE (*Mean Square Error*), jiné metody volby konstanty viz např. Cipra (1986).

2.2 Holtova metoda

Druhou metodou, kterou se budeme zabývat, je taktéž velice známá a používaná Holtova metoda. Tato metoda se dá v praxi s úspěchem použít na většinu nesezónních časových řad s lokálně lineárním trendem. Její výhodou je výpočetní nenáročnost i jednoduchost a názornost jejích vzorců.

Předpokládáme, že daná časová řada y má lineární trend, jehož směrnice se ale v čase může postupně měnit. Stav dané časové řady je tedy v okamžiku n určen jednak její úrovní S_n a jednak směrnicí trendu T_n , dále budeme psát namísto směrnice trendu jen trend. Úroveň S_n je zároveň vyrovnanou hodnotou řady y v okamžiku n , tzn.

$$\hat{y}_n = S_n . \quad (2.2.1)$$

Trend T_n vyjadřuje očekávanou změnu úrovně časové řady při jednotkové změně času. Chceme-li tedy pomocí Holtovy metody předpovídat hodnotu časové řady o $\tau > 0$ časových jednotek dopředu, klademe

$$\hat{y}_{n+\tau}(n) = S_n + \tau \cdot T_n . \quad (2.2.2)$$

Dále budeme potřebovat znát způsob, jakým z hodnot S_n , T_n a nově napozorované hodnoty y_{n+1} dostaneme hodnoty S_{n+1} , T_{n+1} . Ze vzorce (2.2.2), dosadíme-li $\tau = 1$, plyne, že naší předpovědí hodnoty y_{n+1} z času n je výraz $S_n + T_n$. Z toho bychom mohli usuzovat, že $S_{n+1} \approx S_n + T_n$. Úroveň i trend řady se ovšem v čase může měnit a signálem toho je hodnota y_{n+1} , tzn. mohli bychom též očekávat $S_{n+1} \approx y_{n+1}$. Rozumné se tedy může zdát volit hodnotu S_{n+1} jako konvexní lineární kombinaci hodnot $S_n + T_n$ a y_{n+1} . Zvolme tedy $\alpha \in (0,1)$ a položme

$$S_{n+1} = (1 - \alpha)(S_n + T_n) + \alpha \cdot y_{n+1} . \quad (2.2.3)$$

Pevná hodnota α se nazývá *vyrovnávací konstanta pro úroveň řady*.

Analogickou úvahou můžeme dojít i ke vzorci pro T_{n+1} . Z předpokladu, že řada má lokálně lineární trend, odhadujeme, že $T_{n+1} \approx T_n$, ale zároveň má smysl očekávat, že $T_{n+1} \approx S_{n+1} - S_n$. Proto opět zvolíme $\gamma \in (0,1)$ a položíme

$$T_{n+1} = (1 - \gamma)T_n + \gamma(S_{n+1} - S_n) . \quad (2.2.4)$$

Hodnota γ je *vyrovnávací konstanta pro směrnici trendu*.

Pro další výpočty bude ještě výhodné přepsat si vzorce pro aktualizaci úrovně a trendu (2.2.3) a (2.2.4) do tzv. chybového tvaru (*error correction*) jako

$$S_{n+1} = S_n + T_n + \alpha \cdot e_{n+1} , \quad (2.2.5)$$

$$T_{n+1} = T_n + \gamma\alpha \cdot e_{n+1} , \quad (2.2.6)$$

kde $e_{n+1} = y_{n+1} - \hat{y}_{n+1}(n)$ je *jednokroková předpovědní chyba z času n* . Vzorce v tomto tvaru jsou opět velice názorné. Hodnota α určuje, jak moc je předpovědní chyba e_{n+1} zahrnuta do nové hodnoty S_{n+1} a podobně hodnota $\gamma\alpha$ určuje, do jaké míry je tato chyba zahrnuta do T_{n+1} .

Holtova metoda je opět rekurentní metodou, viz vzorce (2.2.3) a (2.2.4) či ekvivalentně (2.2.5) a (2.2.6). Volbu počátečních hodnot S_0 a T_0 lze nejjednodušeji provést podle následujících vzorců

$$\begin{aligned} T_0 &= y_2 - y_1, \\ S_0 &= y_1 - T_0 = 2y_1 - y_2. \end{aligned} \quad (2.2.7)$$

Anebo podobně jako v Brownově metodě pomocí regrese.

Nyní již zbývá pouze pevně zvolit vyrovnávací konstanty α a γ , což se dá provést opět podobně jako v Brownově metodě postupným procházením, kdy zkusíme na řadu y použít různé kombinace konstant a z nich vybereme tu, jež je v předpovídání nejúspěšnější.

Dále budeme popisovat vztah mezi Holtovou metodou a procesem ARIMA(0, 2, 2). Nejdříve si proto proces ARIMA(0, 2, 2) popíšeme.

ARIMA je lineární model, složený z následujících tří dílčích částí (přičemž se vždy nemusí vyskytovat všechny tři): AR (*Autoregressive*) tvoří lineární kombinace minulých hodnot, I (*Integrative*) zahrnuje odstranění případných nestacionarit pomocí diferencí a MA (*Moving Average*) se skládá z lineární kombinace minulých chyb. Když řekneme, že řada $\{y_n, n \in \mathbb{Z}\}$ se řídí modelem ARIMA(0, 2, 2), znamená to, že řada druhých diferencí $\{\Delta^2 y_n, n \in \mathbb{Z}\}$ se řídí modelem MA(2), který můžeme zapsat následovně:

$$\Delta^2 y_n = \varepsilon_n + \theta_1 \varepsilon_{n-1} + \theta_2 \varepsilon_{n-2}, \quad (2.2.8)$$

kde $\{\varepsilon_n, n \in \mathbb{Z}\}$ je bílý šum, což znamená, že musí platit $E(\varepsilon_n) = 0$, $var(\varepsilon_n) = \sigma^2$ a platí $cov(\varepsilon_n, \varepsilon_m) = 0$ pro všechna $n \neq m$. θ_1, θ_2 jsou reálné parametry modelu.

Předpokládejme nyní, že jednokrokové předpovědní chyby $\{e_n, n \in \mathbb{Z}\}$ Holtovy metody definované jako $e_n = y_n - \hat{y}_n(n-1)$ tvoří bílý šum s rozptylem $\sigma^2 > 0$. Vyjdeme z následujícího vzorce

$$y_{n+1} = S_n + T_n + e_{n+1}. \quad (2.2.9)$$

Dále již s využitím vzorců (2.2.5) a (2.2.6) pro aktualizaci statistik můžeme počítat

$$\begin{aligned} \Delta y_n &= y_n - y_{n-1} = (S_{n-1} + T_{n-1} + e_n) - (S_{n-2} + T_{n-2} + e_{n-1}) = \\ &= (S_{n-1} - S_{n-2}) + (T_{n-1} - T_{n-2}) + e_n - e_{n-1} = \\ &= T_{n-2} + \alpha \cdot e_{n-1} + \gamma \alpha \cdot e_{n-1} + e_n - e_{n-1} = \\ &= T_{n-2} + (\alpha + \gamma \alpha - 1)e_{n-1} + e_n. \end{aligned} \quad (2.2.10)$$

Dalším diferencováním se dostáváme k výsledku

$$\begin{aligned}\Delta^2 y_n &= \Delta y_n - \Delta y_{n-1} = [T_{n-2} + (\alpha + \gamma\alpha - 1)e_{n-1} + e_n] - \\ &\quad - [T_{n-3} + (\alpha + \gamma\alpha - 1)e_{n-2} + e_{n-1}] = \\ &= e_n + (\alpha + \gamma\alpha - 2) \cdot e_{n-1} + (1 - \alpha) \cdot e_{n-2}.\end{aligned}\quad (2.2.11)$$

Tímto jsme tedy ukázali, že se řada $\{\Delta^2 y_n, n \in \mathbb{Z}\}$ řídí modelem MA(2), tedy také to, že se řada $\{y_n, n \in \mathbb{Z}\}$ řídí modelem ARIMA(0, 2, 2) s parametry

$$\begin{aligned}\theta_1 &= \alpha + \gamma\alpha - 2 \\ \theta_2 &= 1 - \alpha.\end{aligned}\quad (2.2.12)$$

Na závěr odstavce se ještě zmiňme o tzv. modelu VAR a jeho vztahu k Holtově metodě. Model VAR (*Vector autoregression*) se používá k zachycení vývoje a vzájemných závislostí mezi vícerozměrnými časovými řadami. Tento model popisuje vývoj k proměnných po určitou dobu ($n = 1, \dots, T$) jako lineární funkci jejich minulého vývoje. Označme si hodnotu i -té proměnné v čase $n + 1$ jako $y_{i,n+1}$. Je-li každá proměnná $y_{i,n+1}$ lineárně závislá na p předchozích pozorováních svých i ostatních proměnných a na chybovém vektoru, mluvíme o modelu VAR p -tého řádu neboli VAR(p). VAR(p) zapsaný pro k -rozměrný vektor proměnných y_{n+1} tedy vypadá takto

$$y_{n+1} = c + A_1 y_n + A_2 y_{n-1} + \dots + A_p y_{n-p+1} + e_{n+1}, \quad (2.2.13)$$

kde c označuje k -rozměrný vektor konstant, A_i je $k \times k$ -rozměrná matice pro každé $i = 1, \dots, p$ a e_{n+1} je k -rozměrný bílý šum, pro nějž platí $E(e_n) = 0$, $cov(e_n, e_m) = 0$ pro $n \neq m$ a varianční matice $var(e_n)$ je invariantní v čase.

My budeme nadále pracovat pouze s modelem VAR(1) pro dvě proměnné tj. $k = 2$, který můžeme rozepsat soustavu dvou rovnic takto

$$\begin{aligned}y_{1,n+1} &= c_1 + A_{1,1}y_{1,n} + A_{1,2}y_{2,n} + e_{1,n+1}, \\ y_{2,n+1} &= c_2 + A_{2,1}y_{1,n} + A_{2,2}y_{2,n} + e_{2,n+1}.\end{aligned}\quad (2.2.14)$$

Model VAR(1) lze také ekvivalentně vyjádřit v maticovém tvaru, a to následovně

$$\begin{pmatrix} y_{1,n+1} \\ y_{2,n+1} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \begin{pmatrix} y_{1,n} \\ y_{2,n} \end{pmatrix} + \begin{pmatrix} e_{1,n+1} \\ e_{2,n+1} \end{pmatrix}.\quad (2.2.15)$$

Uvedme nyní podmínku (slabé) stacionarity pro proces VAR(1) ve tvaru (2.2.15), tedy podmínku toho, aby daný proces měl konstantní vektor středních hodnot a kovarianční matici invariantní vůči posunům v čase, což speciálně znamená, že má zároveň konstantní varianční matici. Proces (2.2.15) je stacionární, pokud všechna vlastní čísla matice $A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}$ jsou v absolutní hodnotě menší než 1, tedy leží uvnitř jednotkového kruhu v komplexní rovině.

Nyní si prohlédněme vzorce (2.2.5) a (2.2.6) pro aktualizaci úrovně a trendu Holtovy metody v chybovém tvaru, které v čase $n + 1$ vypadají takto

$$S_{n+1} = S_n + T_n + \alpha \cdot e_{n+1}, \quad (2.2.16)$$

$$T_{n+1} = T_n + \gamma\alpha \cdot e_{n+1}. \quad (2.2.17)$$

Zde vidíme, že jak úroveň S_{n+1} , tak trend T_{n+1} závisí lineárně na předchozích hodnotách úrovně a trendu a také na náhodných složkách e_{n+1} přenásobených konstantou. Proto můžeme vyjádřit vývoj úrovně a trendu v Holtově metodě pomocí modelu VAR(1) pro dvě proměnné. Zapsáno maticově příslušný model vypadá takto

$$\begin{pmatrix} S_{n+1} \\ T_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S_n \\ T_n \end{pmatrix} + \begin{pmatrix} \alpha \cdot e_{n+1} \\ \gamma\alpha \cdot e_{n+1} \end{pmatrix}. \quad (2.2.18)$$

Pro korektnost je důležité si uvědomit, že bílý šum zůstane i po přenásobení konstantou bílým šumem a že proces $(\alpha \cdot e_{n+1}, \gamma\alpha \cdot e_{n+1})^T$ je dvourozměrný bílý šum.

Vzorce (2.2.16) a (2.2.17) jsou dále zajímavé i z jiných pohledů. Podívejme se nejprve na vzorec (2.2.17) a povšimněme si, že hodnota trendu v čase $n + 1$ T_{n+1} je součtem předchozí hodnoty trendu T_n a náhodné složky $\gamma\alpha \cdot e_{n+1}$, což vlastně znamená, že vývoj hodnoty trendu tvoří v čase náhodnou procházku.

Na závěr si povšimněme, že pokud od vzorce (2.2.9) odečteme vzorec pro aktualizaci úrovně (2.2.5), dostáváme

$$y_{n+1} - S_{n+1} = S_n + T_n + e_{n+1} - S_n - T_n - \alpha \cdot e_{n+1} = (1 - \alpha) \cdot e_{n+1}, \quad (2.2.19)$$

odkud vidíme, že rozdíl $y_{n+1} - S_{n+1}$ je bílý šum.

2.3 Vztah mezi Brownovou a Holtovou metodou

V tomto odstavci si ukážeme, jaký je vztah mezi Brownovou a Holtovou metodou, a sice, že Brownova metoda je speciálním případem Holtovy metody. Aby nedocházelo k záměně, označíme si vyrovnávací konstantu Brownovy metody jako α a vyrovnávací konstanty Holtovy metody jako α_H a γ_H .

Ukážeme, že pro každou vyrovnávací konstantu Brownovy metody $\alpha \in (0,1)$ existuje dvojice vyrovnávacích konstant Holtovy metody $\alpha_H \in (0,1)$ a $\gamma_H \in (0,1)$ taková, že obě metody jsou ekvivalentní, tzn. poskytují identické předpovědi.

Z předpovědní formule (2.2.2) Holtovy metody, tedy ze vzorce

$$\hat{y}_{n+\tau}(n) = S_n + \tau \cdot T_n, \quad (2.3.1)$$

plyne, že, pokud k předpovídání použijeme Holtovu metodu, bodové předpovědi z času n pro různé předpovědní horizonty $\tau > 0$ tvoří přímkou procházející vyrovnanou hodnotou $\hat{y}_n = S_n$, která má směrnici T_n , tedy směrnici trendu řady y v čase n .

Podobně, upravíme-li předpovědní formule Brownovy metody (2.1.13), dostáváme pro $\tau > 0$

$$\hat{y}_{n+\tau}(t) = 2S_n^{[1]} - S_n^{[2]} + \tau \frac{\alpha}{1-\alpha} (S_n^{[1]} - S_n^{[2]}), \quad (2.3.2)$$

odkud je podobně vidět, že bodové předpovědi Brownovy metody z času n pro předpovědní horizonty $\tau > 0$ tvoří přímkou procházející vyrovnanou hodnotou $\hat{y}_n = 2S_n^{[1]} - S_n^{[2]}$, která má za směrnici hodnotu $\frac{\alpha}{1-\alpha} (S_n^{[1]} - S_n^{[2]})$.

Vidíme, že obě dvě metody vytváří při předpovědích z času n do budoucnosti předpovědní přímkou. Obě metody budou ekvivalentní právě tehdy, budou-li mít stejné předpovědní přímkou v každém čase n . Využijeme nyní toho, že přímkou jsou shodné právě tehdy, když procházejí stejným bodem a mají shodnou směrnici. Budeme předpokládat, že obě přímkou jsou shodné v čase n a ukážeme, že potom budou shodné i v čase $n + 1$ (vzhledem k tomu, že obě metody jsou rekurentní, tak vlastně ani jinou možnost, jak jejich ekvivalenci ukázat, nemáme). Zapsáno vzorcem tedy předpokládáme v bodě n shodu úrovní, viz (2.3.1) a (2.3.2)

$$S_n = 2S_n^{[1]} - S_n^{[2]}, \quad (2.3.3)$$

a zároveň shodu směrnic opět viz (2.3.1) a (2.3.2)

$$T_n = \frac{\alpha}{1-\alpha} (S_n^{[1]} - S_n^{[2]}). \quad (2.3.4)$$

Nejprve dokážeme, že i v čase $n + 1$ mají obě předpovědní přímkou stejnou úroveň, tedy že

$$S_{n+1} = 2S_{n+1}^{[1]} - S_{n+1}^{[2]}. \quad (2.3.5)$$

V rovnici (2.3.5) dosadíme za S_{n+1} , $S_{n+1}^{[1]}$ a $S_{n+1}^{[2]}$ podle vzorců (2.1.9), (2.1.10) a (2.2.3) pro aktualizaci statistik Brownovy a Holtovy metody. Dostáváme

$$\begin{aligned} (1 - \alpha_H)(S_n + T_n) + \alpha_H y_{n+1} &= \\ &= S_n^{[1]}(2 - 3\alpha + \alpha^2) + S_n^{[2]}(\alpha - 1) + y_{n+1}(2\alpha - \alpha^2). \end{aligned} \quad (2.3.6)$$

Sečtením rovnic (2.3.3) a (2.3.4), jejichž platnost předpokládáme, získáme

$$S_n + T_n = 2S_n^{[1]} - S_n^{[2]} + \frac{\alpha}{1-\alpha} (S_n^{[1]} - S_n^{[2]}), \quad (2.3.7)$$

po úpravě

$$S_n + T_n = S_n^{[1]} \left(2 + \frac{\alpha}{1-\alpha}\right) + S_n^{[2]} \left(-1 - \frac{\alpha}{1-\alpha}\right). \quad (2.3.8)$$

Dosadíme-li podle vzorce (2.3.8) do rovnice (2.3.6), dostaneme

$$\begin{aligned} (1 - \alpha_H) \left(2 + \frac{\alpha}{1-\alpha}\right) S_n^{[1]} + (1 - \alpha_H) \left(-1 - \frac{\alpha}{1-\alpha}\right) S_n^{[2]} + \alpha_H y_{t+1} &= \\ &= S_n^{[1]} (2 - 3\alpha + \alpha^2) + S_n^{[2]} (\alpha - 1) + y_{n+1} (2\alpha - \alpha^2). \end{aligned} \quad (2.3.9)$$

Nyní již stačí jen srovnat koeficienty u $S_n^{[1]}$, $S_n^{[2]}$ a y_{n+1} a zjišťujeme, že rovnost v (2.3.9) a tedy i v (2.3.5) platí právě tehdy, když

$$\alpha_H = \alpha(2 - \alpha). \quad (2.3.10)$$

Přitom z předpokladu $\alpha \in (0,1)$ plyne, že i $\alpha_H \in (0,1)$.

Dále potřebujeme dokázat, že se směrnice obou předpovědních přímk v čase $n + 1$ rovnají, tzn.

$$T_{n+1} = \frac{\alpha}{1-\alpha} (S_{n+1}^{[1]} - S_{n+1}^{[2]}). \quad (2.3.11)$$

Opět dosadíme podle vzorců (2.1.9), (2.1.10) a (2.2.3) pro aktualizaci statistik u obou metod za $S_{n+1}^{[1]}$, $S_{n+1}^{[2]}$ a T_{n+1} a dostáváme

$$\begin{aligned} (1 - \gamma_H) T_n + \gamma_H (S_{n+1} - S_n) &= \\ &= \frac{\alpha}{1-\alpha} \left[(1 - \alpha) S_n^{[1]} + \alpha y_{n+1} - (1 - \alpha) S_n^{[2]} - \alpha S_{n+1}^{[1]} \right]. \end{aligned} \quad (2.3.12)$$

Na levé straně rovnice (2.3.12) dosadíme za S_{n+1} podle vzorec (2.2.3). Za T_n dosadíme podle vzorce (2.3.4) vycházející z předpokladu o shodě trendů v čase n a za S_n z předpokladu o shodě úrovní (2.3.3) a nakonec využijeme námi zjištěný vztah (2.3.10). Na pravé straně také rozepíšeme $S_{n+1}^{[1]}$ podle vzorce pro aktualizaci statistik (2.1.9). Po těchto úpravách obdržíme rovnici

$$\begin{aligned} (1 - \gamma_H) \frac{\alpha}{1-\alpha} (S_n^{[1]} - S_n^{[2]}) + \\ + \gamma_H \left[S_n^{[1]} (\alpha^2 - 3\alpha) + S_n^{[2]} \alpha + y_{n+1} (2\alpha - \alpha^2) \right] &= \\ = S_n^{[1]} (\alpha - \alpha^2) + S_n^{[2]} (-\alpha) + y_{n+1} \left(\frac{\alpha^2}{1-\alpha} - \frac{\alpha^3}{1-\alpha} \right), \end{aligned} \quad (2.3.13)$$

v níž můžeme opět srovnat koeficienty u vyrovnávacích statistik $S_n^{[1]}$, $S_n^{[2]}$ a u y_{n+1} , odkud nám vyjde, že rovnost v (2.3.13) a tedy i v (2.3.11) platí tehdy a jen tehdy, když

$$\gamma_H = \frac{\alpha}{2 - \alpha}, \quad (2.3.14)$$

přičemž z předpokladu $\alpha \in (0,1)$ plyne, že i $\gamma_H \in (0,1)$.

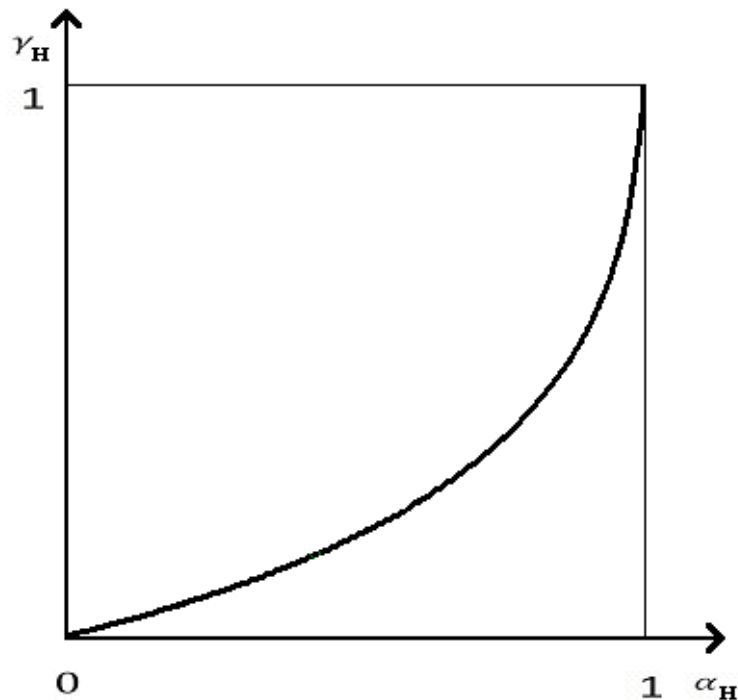
Na závěr shrňme výsledky, které jsme dokázali. Brownova metoda s vyrovnávací konstantou α je ekvivalentní Holtově metodě s vyrovnávacími konstantami α_H a γ_H určenými vzorci

$$\alpha_H = \alpha(2 - \alpha) \text{ a } \gamma_H = \frac{\alpha}{2 - \alpha}. \quad (2.3.15)$$

Je zajímavé si povšimnout, že platí

$$\alpha = \sqrt{\alpha_H \cdot \gamma_H}, \quad (2.3.16)$$

tedy, že vyrovnávací konstanta α Brownovy metody je geometrickým průměrem vyrovnávacích konstant α_H a γ_H ekvivalentní Holtovy metody. Vztah mezi vyrovnávacími konstantami α_H a γ_H si pro ilustraci můžeme znázornit na následujícím obrázku 2.3-1, kde na vodorovné ose máme vyneseny hodnoty α_H a na svislé hodnoty γ_H . Křivka zobrazuje kombinace α_H a γ_H , které vždy odpovídají ekvivalentní Brownově metodě s příslušnou vyrovnávací konstantou α . Pracovně budeme tuto křivku nazývat *brownovská diagonála*.

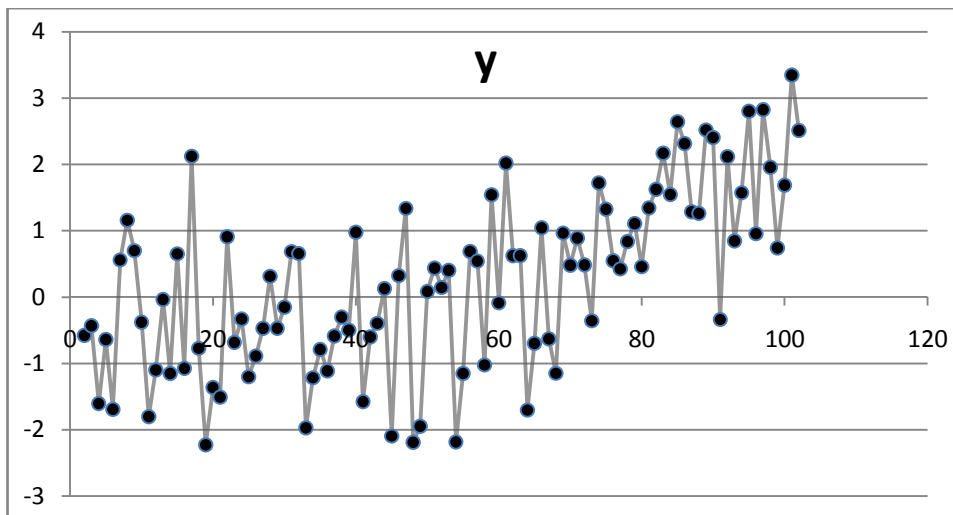


Obrázek 2.3-1 Brownovská diagonála, vztah mezi vyrovnávacími konstantami Brownovy a Holtovy metody

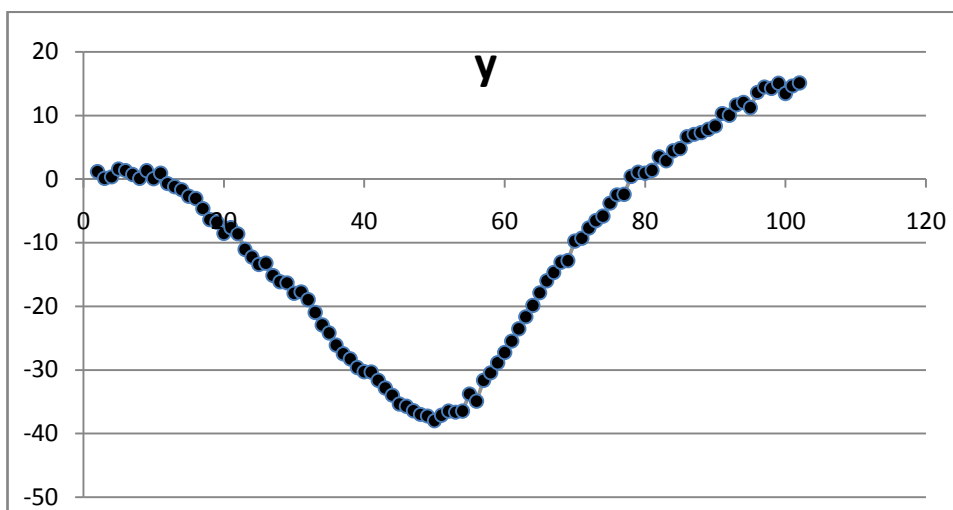
Při použití Holtovy metody můžeme využít pro volbu konstant α_H a γ_H celý čtverec $(0,1) \times (0,1)$, ale při aplikaci Brownovy metody se musíme vlastně omezit pouze na jednorozměrnou křivku z bodu $(0,0)$ do $(1,1)$ parametrizovanou monotónně pomocí $\alpha \in (0,1)$.

Všeobecně bývá v praxi Holtova metoda častěji používána než metoda Brownova, protože se v ní používají dvě spolu nijak nesvázané konstanty, což poskytuje této metodě větší flexibilitu. Nevýhodou může za určitých situací být větší výpočetní náročnost při volbě vyrovnávacích konstant, protože při v odstavci 2.2 popsané volbě je nutné procházet celý čtverec $(0,1) \times (0,1)$ oproti Brownově metodě, kde stačí procházet pouze interval $(0,1)$. Pokud při hledání optimálních vyrovnávacích konstant α_H a γ_H použijeme nějaký numerický optimalizační algoritmus urychlující volbu α_H a γ_H , hrozí nám potom riziko, že najdeme pouze lokální minimum MSE v některém bodě, takže nalezené konstanty α_H a γ_H nebudou optimální a Holtova metoda potom nemusí dávat uspokojivé výsledky.

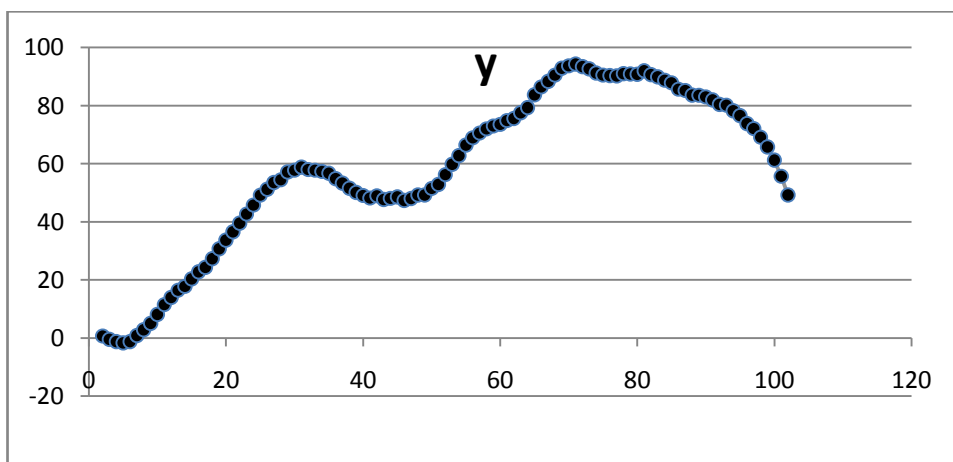
Nyní si ještě ukážeme, jak vlastně vypadají časové řady, pro které je vhodné použít Holtovu metodu s danými vyrovnávacími konstantami. Pokud předpokládáme, že jednokrokové předpovědní chyby e_n tvoří v čase bílý šum, můžeme generovat časové řady pomocí modelu ARIMA(0,2,2), jehož vztah k Holtově metodě jsme ukázali v odstavci 2.2. Pokud příslušné vzorce aplikujeme například v programu MS Excel pro 100 pozorování časové řady, vyjdou nám následující obrázky 2.3-2 až 2.3-6, které ukazují hodnoty časové řady y v čase.



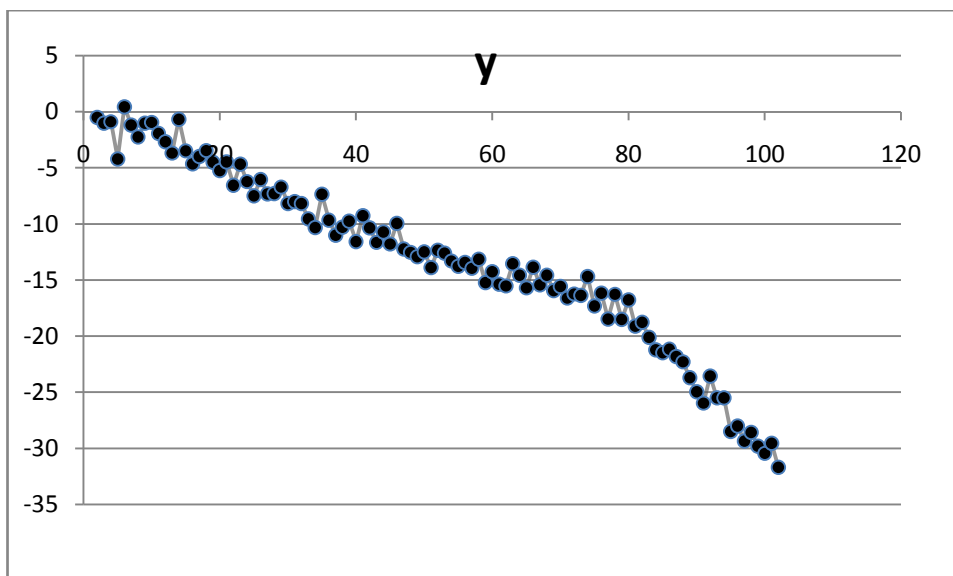
Obrázek 2.3-2 Časová řada generovaná Holtovou metodou s $\alpha = 0,1$; $\gamma = 0,1$



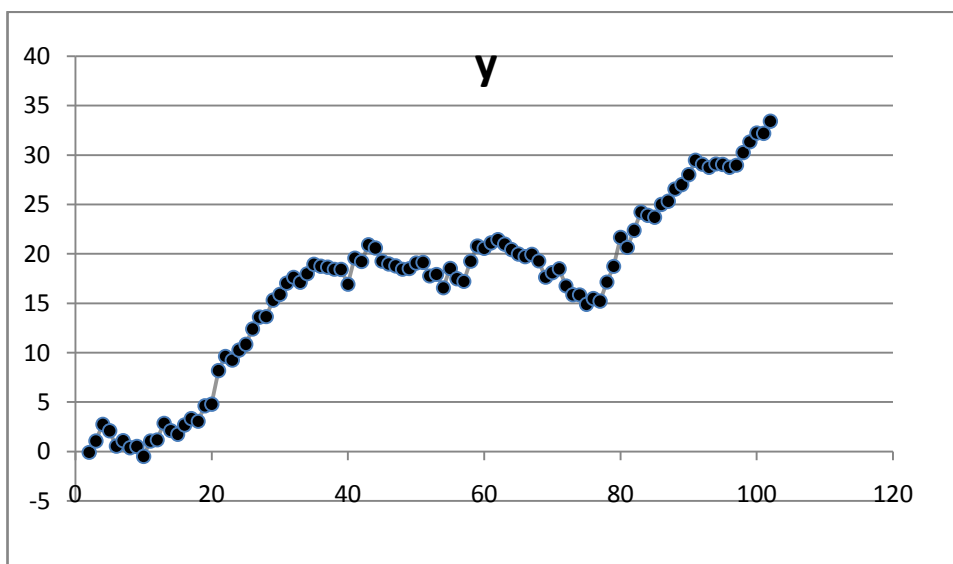
Obrázek 2.3-3 Časová řada generovaná Holtovou metodou s $\alpha = 0,5$; $\gamma = 0,5$



Obrázek 2.3-4 Časová řada generovaná Holtovou metodou s $\alpha = 0,9$; $\gamma = 0,9$



Obrázek 2.3-5 Časová řada generovaná Holtovou metodou s $\alpha = 0,1$; $\gamma = 0,9$



Obrázek 2.3-6 Časová řada generovaná Holtovou metodou s $\alpha = 0,9$; $\gamma = 0,1$

To, že byla časová řada generovaná pomocí rovnic Holtovy metody s vyrovnávacími konstantami α a γ , vlastně znamená, že k vyrovnávání a předpovídání dané časové řady se nejvíce hodí právě Holtova metoda se stejnými vyrovnávacími konstantami α a γ . Na obrázcích 2.3-2 až 2.3-6 si povšimněme, že čím je hodnota konstanty α nižší, tím časová řada dosahuje nižších absolutních hodnot. S vyšší hodnotou konstanty γ potom řada snáze a rychleji mění směrnici trendu.

3 Citlivost Holtovy metody na změny vyrovnávacích konstant

Používáme-li Holtovu metodu k vyrovnávání a předpovídání časových řad, dosti záleží na volbě dvojice vyrovnávacích konstant, jak dobré dostaneme výsledky. V odstavci 2.2, kde jsme představili Holtovu metodu, jsme uvedli, jak se obvykle v praxi provádí hledání optimálních vyrovnávacích konstant. V tomto odstavci se budeme věnovat tomu, jak velkých chyb se dopustíme, když použijeme k předpovídání Holtovu metodu s jinými než optimálními vyrovnávacími konstantami. Jako kritérium velikosti chyby při použití dané neoptimální kombinace vyrovnávacích konstant budeme používat velikost MSE (*Mean Square Error*), což je vlastně rozptyl jednokrokových předpovědních chyb použité metody. Dalším údajem, který bude dobré sledovat, je autokorelační koeficient ρ , což je korelace po sobě jdoucích jednokrokových předpovědních chyb použité metody.

K výpočtu velikosti předpovědní chyby Holtovy metody s neoptimálními vyrovnávacími konstantami bychom mohli použít vhodné reálné časové řady, na které bychom zkoušeli aplikovat Holtovu metodu s různými kombinacemi vyrovnávacích konstant a srovnávali dosažené výsledky. Tento způsob by ovšem předpokládal náročné vyhledání značného množství vhodných dostatečně dlouhých časových řad, abychom dostali relevantní výsledky. Proto raději budeme časové řady generovat pomocí náhodného procesu ARIMA(0, 2, 2), což nám navíc umožní odvodit hodnoty MSE a ρ teoreticky a pro všechny kombinace generujících a použitých konstant.

3.1 Výpočet MSE

Nyní využijeme vztahu mezi Holtovou metodou a náhodným procesem ARIMA(0,2,2), který jsme odvodili v odstavci 2.2. Pokud na proces ARIMA(0, 2, 2) použijeme Holtovu metodu s vyrovnávacími konstantami α , γ , které odpovídají parametrům procesu ARIMA(0, 2, 2) θ_1 a θ_2 ze vzorců (2.2.12), získáme předpovědi s minimální MSE - viz Chatfield (2002). Naším cílem bude spočítat, jak se změní MSE, když na daný proces použijeme Holtovu metodu s jinými konstantami.

Nejdříve si ukážeme způsob, jakým budeme generovat časové řady. Abychom mohli využít vztahu Holtovy metody s procesem ARIMA(0, 2, 2), předpokládejme, že jednotlivé chyby e_n tvoří bílý šum s rozptylem $\sigma^2 > 0$. K vytvoření časové řady nyní můžeme využít následující generující schéma. Toto schéma je sestaveno z rovnic Holtovy metody s vyrovnávacími konstantami α_1 , γ_1 , viz vzorce (2.2.5), (2.2.6) a (2.2.9)

$$S_{n+1} = S_n + T_n + \alpha_1 e_{n+1}, \quad (3.1.1)$$

$$T_{n+1} = T_n + \gamma_1 \alpha_1 e_{n+1}, \quad (3.1.2)$$

$$y_{n+1} = S_n + T_n + e_{n+1}. \quad (3.1.3)$$

Toto schéma nám zaručuje, že jím vytvořená časová řada bude vhodná pro použití Holtovy metody k jejímu vyrovnávání a předpovídání. Přitom také platí, že nejlepších výsledků (nejnižší MSE) lze dosáhnout při použití Holtovy metody se stejnými vyrovnávacími konstantami α_1, γ_1 jako při generování dané řady. Je dobré si uvědomit, že naše generující schéma odpovídá procesu ARIMA(0, 2, 2) parametrizovanému vyrovnávacími konstantami α_1, γ_1 prostřednictvím vzorců (2.2.12).

Dále si označíme rovnice Holtovy metody použité na vygenerovanou časovou řadu y . Budeme aplikovat Holtovu metodu s vyrovnávacími konstantami α_2, γ_2 , které nemusí být nutně shodné s α_1, γ_1 , jejíž rovnice vypadají takto (opět vycházíme z vzorců (2.2.5), (2.2.6) a (2.2.9))

$$\tilde{S}_{n+1} = \tilde{S}_n + \tilde{T}_n + \alpha_2 \varepsilon_{n+1} \quad (3.1.4)$$

$$\tilde{T}_{n+1} = \tilde{T}_n + \gamma_2 \alpha_2 \varepsilon_{n+1} \quad (3.1.5)$$

$$y_{n+1} = \tilde{S}_n + \tilde{T}_n + \varepsilon_{n+1}. \quad (3.1.6)$$

Pomocí této metody budeme tedy časovou řadu y vyrovnávat a předpovídat. Jsou-li hodnoty vyrovnávacích konstant generujícího schématu α_1 a γ_1 rozdílné od vyrovnávacích konstant α_2 a γ_2 aplikované metody, znamená to zároveň, že i spočtené úrovně a trendy řady v jednotlivých časech jsou rozdílné, jelikož vyrovnávací konstanty vždy určují, jak moc je zahrnuto každé nové pozorování y_{n+1} do hodnoty úrovně a trendu v dalším bodě, což je vidět ze vzorců (2.2.3) a (2.2.4). Ze vzorců v chybovém tvaru (2.2.5), (2.2.6) lze podobně nahlédnout, že i velikost chyby v různých časech je v obou případech odlišná. Z těchto důvodů jsme si v aplikované Holtově metodě s vyrovnávacími konstantami α_2 a γ_2 ve vzorcích (3.1.4), (3.1.5) a (3.1.6) označili hodnotu úrovně v bodě n jako \tilde{S}_n , trendu jako \tilde{T}_n a velikost chyby jako ε_n , tedy rozdílně od obdobných veličin použitých v generujícím schématu.

Podle vzorců (3.1.6) a (3.1.3) si nejprve vyjádříme velikost chyby použité metody ε_{n+1} jako

$$\varepsilon_{n+1} = y_{n+1} - (\tilde{S}_n + \tilde{T}_n) = e_{n+1} + (S_n - \tilde{S}_n) + (T_n - \tilde{T}_n). \quad (3.1.7)$$

Ze vzorce (3.1.7) vidíme, že velikost jednokrokové předpovědní chyby použité metody ε_{n+1} odpovídá součtu chyby e_{n+1} a chyby vzniklé použitím metody

s neoptimálními hodnotami vyrovnávacích konstant α_2, γ_2 ve formě rozdílů $S_n - \tilde{S}_n$ a $T_n - \tilde{T}_n$.

Všimněme si nyní vlastností rozdílů $S_{n+1} - \tilde{S}_{n+1}$ a $T_{n+1} - \tilde{T}_{n+1}$. Dosadíme za S_{n+1} , \tilde{S}_{n+1} , T_{n+1} a \tilde{T}_{n+1} podle vzorců (3.1.1), (3.1.2), (3.1.4) a (3.1.5), dostáváme

$$S_{n+1} - \tilde{S}_{n+1} = (1 - \alpha_2)(S_n - \tilde{S}_n) + (1 - \alpha_2)(T_n - \tilde{T}_n) + (\alpha_1 - \alpha_2)e_{n+1}, \quad (3.1.8)$$

$$T_{n+1} - \tilde{T}_{n+1} = (1 - \gamma_2\alpha_2)(T_n - \tilde{T}_n) - \gamma_2\alpha_2(S_n - \tilde{S}_n) + (\gamma_1\alpha_1 - \gamma_2\alpha_2)e_{n+1}. \quad (3.1.9)$$

Nyní srovnáme poslední dva vzorce (3.1.8) a (3.1.9) s modelem VAR(1) popsaným vzorcem (2.2.14). Ve vzorcích (3.1.8) a (3.1.9) jsou oba rozdíly $S_{n+1} - \tilde{S}_{n+1}$ a $T_{n+1} - \tilde{T}_{n+1}$ vyjádřeny pomocí lineární kombinace svých minulých hodnot a chyb e_{n+1} , které jsou bílým šumem, což vlastně znamená, že se proces $(S_{n+1} - \tilde{S}_{n+1}, T_{n+1} - \tilde{T}_{n+1})^T$ řídí modelem VAR(1). Aby byl vztah s modelem VAR(1) jasně patrný, napíšeme vzorce (3.1.8) a (3.1.9) maticově a poté srovnáme s (2.2.15)

$$\begin{pmatrix} S_{n+1} - \tilde{S}_{n+1} \\ T_{n+1} - \tilde{T}_{n+1} \end{pmatrix} = \begin{pmatrix} 1 - \alpha_2 & 1 - \alpha_2 \\ -\gamma_2\alpha_2 & \gamma_1\alpha_1 - \gamma_2\alpha_2 \end{pmatrix} \begin{pmatrix} S_n - \tilde{S}_n \\ T_n - \tilde{T}_n \end{pmatrix} + \begin{pmatrix} (\alpha_1 - \alpha_2)e_{n+1} \\ (\gamma_1\alpha_1 - \gamma_2\alpha_2)e_{n+1} \end{pmatrix}, \quad (3.1.10)$$

kde, jak víme, $(\alpha_1 - \alpha_2)e_{n+1}$ i $(\gamma_1\alpha_1 - \gamma_2\alpha_2)e_{n+1}$ je bílý šum.

Podmínka stacionarity procesu $(S_{n+1} - \tilde{S}_{n+1}, T_{n+1} - \tilde{T}_{n+1})^T$ v našem případě vypadá tak, že pro vlastní čísla matice $A = \begin{pmatrix} 1 - \alpha_2 & 1 - \alpha_2 \\ -\gamma_2\alpha_2 & \gamma_1\alpha_1 - \gamma_2\alpha_2 \end{pmatrix}$, které si označíme jako λ_1 a λ_2 , musí platit $|\lambda_1| < 1$ a $|\lambda_2| < 1$. Jako kořeny charakteristické rovnice $\det(A - \lambda E) = 0$ spočteme obě vlastní čísla, která vyjdou následovně

$$\lambda_{1,2} = \frac{2 - \gamma_2\alpha_2 - \alpha_2 \pm \sqrt{(-2\gamma_2\alpha_2 + \alpha_2)^2 - 4(1 - \alpha_2)}}{2}. \quad (3.1.11)$$

Algebraickými úpravami lze ověřit, že obě podmínky $|\lambda_1| < 1$ a $|\lambda_2| < 1$ jsou splněny pro všechna $\alpha_2, \gamma_2 \in (0,1)$, čímž je zároveň ověřena stacionarita procesu $(S_{n+1} - \tilde{S}_{n+1}, T_{n+1} - \tilde{T}_{n+1})^T$.

Označme nyní

$$v_S = \text{var}(S_n - \tilde{S}_n), \quad (3.1.12)$$

$$v_T = \text{var}(T_n - \tilde{T}_n), \quad (3.1.13)$$

$$w = \text{cov}(S_n - \tilde{S}_n, T_n - \tilde{T}_n). \quad (3.1.14)$$

Stacionarita procesu $(S_n - \tilde{S}_n, T_n - \tilde{T}_n)^T$, zaručuje, že kovarianční matice rozdílů $cov(S_n - \tilde{S}_n, T_n - \tilde{T}_n) = \begin{pmatrix} v_S & w \\ w & v_T \end{pmatrix}$ je invariantní vůči posunům v čase. Speciálně tedy platí

$$v_S = var(S_n - \tilde{S}_n) = var(S_{n+1} - \tilde{S}_{n+1}), \quad (3.1.15)$$

$$v_T = var(T_n - \tilde{T}_n) = var(T_{n+1} - \tilde{T}_{n+1}), \quad (3.1.16)$$

$$w = cov(T_n - \tilde{T}_n, S_n - \tilde{S}_n) = cov(T_{n+1} - \tilde{T}_{n+1}, S_{n+1} - \tilde{S}_{n+1}). \quad (3.1.17)$$

Z předpokladu, který jsem o chybách učinili, víme, že chyby e_n a e_m a jsou nekorelované pro $m \neq n$, z toho vyplývá, že i $S_n - \tilde{S}_n$ a e_{n+1} jsou nekorelované, protože rozdíl $S_n - \tilde{S}_n$ závisí pouze na posloupnosti chyb do času n , tzn. na e_n, e_{n-1}, \dots , viz (3.1.8). Odtud $cov(S_n - \tilde{S}_n, e_{n+1}) = 0$ a analogicky také $cov(T_n - \tilde{T}_n, e_{n+1}) = 0$. S využitím tohoto poznatku a dosazením za $S_{n+1} - \tilde{S}_{n+1}$ a $T_{n+1} - \tilde{T}_{n+1}$ podle vzorců (3.1.8) a (3.1.9) vyjádříme postupně v_S, v_T a w následovně

$$\begin{aligned} v_S &= var(S_{n+1} - \tilde{S}_{n+1}) = (1 - \alpha_2)^2 v_S + 2(1 - \alpha_2)^2 w + (1 - \alpha_2)^2 v_T + \\ &\quad + (\alpha_1 - \alpha_2)^2 \sigma^2, \\ v_T &= var(T_{n+1} - \tilde{T}_{n+1}) = (1 - \gamma_2 \alpha_2)^2 v_T - 2(1 - \gamma_2 \alpha_2) \gamma_2 \alpha_2 w + \\ &\quad + \gamma_2^2 \alpha_2^2 v_S + (\gamma_1 \alpha_1 - \gamma_2 \alpha_2)^2 \sigma^2, \\ w &= cov((T_{n+1} - \tilde{T}_{n+1}), (S_{n+1} - \tilde{S}_{n+1})) = (1 - \alpha_2)(1 - 2\gamma_2 \alpha_2) w - \\ &\quad - (1 - \alpha_2) \gamma_2 \alpha_2 v_S + (1 - \alpha_2)(1 - \gamma_2 \alpha_2) v_T + \\ &\quad + (\alpha_1 - \alpha_2)(\gamma_1 \alpha_1 - \gamma_2 \alpha_2) \sigma^2. \end{aligned} \quad (3.1.18)$$

Tím jsme vlastně získali soustavu tří rovnic o třech neznámých v_S, v_T a w s parametry $\alpha_1, \gamma_1, \alpha_2, \gamma_2, \sigma^2$, která po úpravě vypadá takto

$$\begin{aligned} [(1 - \alpha_2)^2 - 1]v_S + (1 - \alpha_2)^2 v_T + 2(1 - \alpha_2)^2 w &= -(\alpha_1 - \alpha_2)^2 \sigma^2, \\ \gamma_2^2 \alpha_2^2 v_S + [(1 - \gamma_2 \alpha_2)^2 - 1]v_T - 2(1 - \gamma_2 \alpha_2) \gamma_2 \alpha_2 w &= -(\gamma_1 \alpha_1 - \gamma_2 \alpha_2)^2 \sigma^2, \\ -(1 - \alpha_2) \gamma_2 \alpha_2 v_S + (1 - \alpha_2)(1 - \gamma_2 \alpha_2) v_T + [(1 - \alpha_2)(1 - 2\gamma_2 \alpha_2) - 1]w &= \\ &= -(\alpha_1 - \alpha_2)(\gamma_1 \alpha_1 - \gamma_2 \alpha_2) \sigma^2. \end{aligned} \quad (3.1.19)$$

Lze si povšimnout, že soustava rovnic (3.1.19) je regulární pro všechny přípustné hodnoty parametrů $\alpha_1, \gamma_1, \alpha_2, \gamma_2$ a σ^2 a má tedy právě jedno řešení. Pravá strana soustavy je nenulová právě tehdy, když platí $\alpha_1 \neq \alpha_2$ nebo $\gamma_1 \neq \gamma_2$, v tomto případě má tedy soustava právě jedno, a to netriviální řešení. Pravá strana je nulová právě tehdy, když platí $\alpha_1 = \alpha_2, \gamma_1 = \gamma_2$ a soustava má právě jedno, a to triviální řešení. Dále si povšimněme, že parametr σ^2 leží pouze na pravé straně soustavy a

násobí jí, což znamená, že hodnota parametru σ^2 pouze úměrně zvětšuje hodnoty neznámých v_S , v_T a w . Bez újmy na obecnosti proto lze volit například $\sigma^2 = 1$.

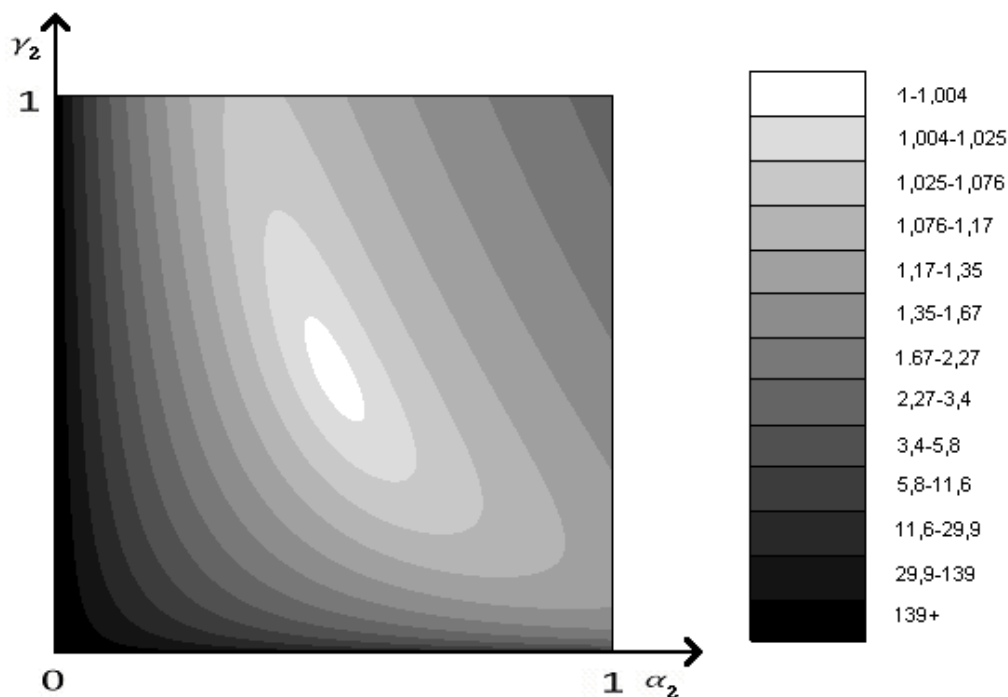
Vyjadřovat přímo řešení této soustavy rovnic pomocí parametrů by bylo příliš složité a nepřehledné, ale bez problémů můžeme tyto rovnice řešit pro konkrétní hodnoty parametrů (například pomocí Cramerova pravidla). Předpokládejme tedy dále, že jsme již vyřešili soustavu (3.1.19) a známe hodnoty v_S , v_T a w .

Naším cílem je spočítat MSE metody použité k vyrovnání a předpovídání časové řady y . Jak víme MSE je rozptyl chyb použité metody, tedy $var(\varepsilon_n)$. Nyní si uvědomme, že posloupnost chyb ε_n je stacionární, což je vidět ze vzorce (3.1.7), protože součet tří stacionárních posloupností je také stacionární. Využijeme-li tedy stacionaritu chyb ε_n , dostáváme $var(\varepsilon_n) = var(\varepsilon_{n+1})$. Dále s využitím vztahu (3.1.7) a s využitím nekorelovanosti e_{n+1} s $S_n - \tilde{S}_n$ i $T_n - \tilde{T}_n$ počítejme

$$\begin{aligned} MSE = var(\varepsilon_{n+1}) &= var\left(e_{n+1} + (S_n - \tilde{S}_n) + (T_n - \tilde{T}_n)\right) = \\ &= \sigma^2 + v_S + v_T + 2w. \end{aligned} \quad (3.1.20)$$

Z tohoto vzorce můžeme vyčíst, že velikost MSE je úměrná rozptylu jednokrokových předpovědních chyb σ^2 (jelikož již víme, že i hodnoty proměnných v_S , v_T a w jsou úměrné σ^2) a navíc je přičtena chyba způsobená volbou metody s neoptimální kombinací vyrovnávacích konstant α_2, γ_2 . Zároveň si můžeme všimnout, že pokud k předpovídání časové řady použijeme Holtovu metodu se stejnými vyrovnávacími konstantami jaké byli použity v generujícím schématu, tedy $\alpha_2 = \alpha_1, \gamma_2 = \gamma_1$, potom nastávají i rovnosti $S_n = \tilde{S}_n, T_n = \tilde{T}_n$ a ze vzorce (3.1.20) vidíme, že $MSE = \sigma^2$.

Nyní již tedy umíme vypočítat MSE pro libovolnou kombinaci parametrů $\alpha_1, \gamma_1, \alpha_2, \gamma_2 \in (0,1), \sigma^2 > 0$. Zvolme pevně $\sigma^2 = 1$. Nalezli jsme tedy vlastně funkci $MSE(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$. Abychom si funkci $MSE(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$ mohli lépe představit, nakreslíme si ji na obrázku 3.1-1, kde pevně volíme $\alpha_1 = 0,5, \gamma_1 = 0,5$, na vodorovné ose leží hodnoty α_2 a na svislé γ_2 . Obrázek nám nyní ukazuje, jak se mění MSE, když na časovou řadu generovanou schématem s $\alpha_1 = 0,5, \gamma_1 = 0,5$ použijeme Holtovu metodu s vyrovnávacími konstantami α_2, γ_2 . Každá vrstevnice při přechodu ze světlejší do tmavší barvy představuje zvýšení hodnoty MSE, když začínáme v bodě $(0,5; 0,5)$, kde je dosaženo hodnoty MSE 1. V oblasti nejtmaší barvy potom dosahuje MSE hodnoty 139 a vyšší, což již značí naprosté selhání dané metody při vyrovnávání a předpovídání dané časové řady. K obrázku je přiložena stupnice, která vysvětluje použité odstíny barev, jako intervaly hodnot MSE.

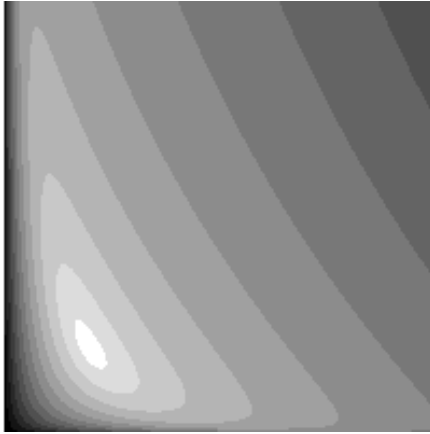


Obrázek 3.1-1 $MSE(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$ pro $\alpha_1 = 0,5$; $\gamma_1 = 0,5$

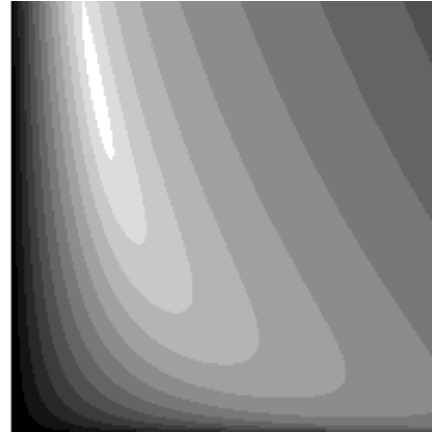
Na obrázku 3.1-1 vidíme, že přibližně platí intuitivní představa, že čím dále jsou konstanty použité metody od metody generující, tím jsou dosažené předpovědi horší. Zároveň se ale ukazuje, že pokud jsou vyrovnávací konstanty α_2 a γ_2 oproti generujícím podhodnoceny, vede to k větším chybám v předpovědích, než pokud jsou nadhodnoceny. Vidíme dokonce, že pokud se buď hodnota α_2 nebo γ_2 limitně blíží 0, MSE roste do nekonečna. Naopak při limitním přiblížení k bodu (1,1), zůstává hodnota MSE konečná, tzn. funkce $MSE(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$ je zde shora omezena.

Dále si můžeme povšimnout, že funkce není v blízkosti bodu optima (0,5; 0,5) příliš strmá (viz vysvětlivky jednotlivých barev vpravo od obrázku), což znamená, že když se při volbě vyrovnávacích konstant použité metody o malou hodnotu odchýlíme od optimálních hodnot, nemělo by se to příliš projevit ve výsledcích této metody. Podlouhlý tvar vrstevnic zase značí, že větších chyb se dopustíme, pokud konstanty použité metody budou obě větší nebo obě menší než konstanty generující. Naopak menší chyby se dopustíme, pokud jednu z konstant zmenšíme a druhou zvětšíme, tedy zde dochází k jakési vzájemné kompenzaci.

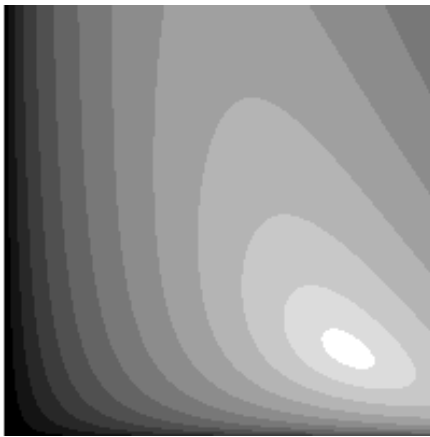
Pro ilustraci si ještě na následujících obrázcích 3.1-2 až 3.1-5 ukažme, jak se situace změní, zvolíme-li jinak hodnoty generujících konstant α_1, γ_1 . Pro všechny čtyři obrázky platí stejný popis jako pro 3.1-1.



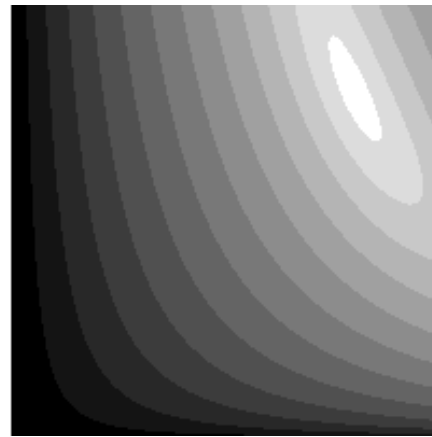
Obrázek 3.1-2 MSE pro $\alpha_1 = 0, 2$; $\gamma_1 = 0, 2$



Obrázek 3.1-3 MSE pro $\alpha_1 = 0, 2$; $\gamma_1 = 0, 8$



Obrázek 3.1-4 MSE pro $\alpha_1 = 0, 8$; $\gamma_1 = 0, 2$



Obrázek 3.1-5 MSE pro $\alpha_1 = 0, 8$; $\gamma_1 = 0, 8$

3.2 Výpočet ρ

Zkusíme nyní ještě získat podobnou funkci $\varrho(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$, která nám ukáže, jak se mění autokorelační koeficient jednorokových předpovědních chyb ϱ při vyrovnávání a předpovídání řady y , když se změní jednotlivé konstanty v generujícím schématu a použité Holtově metodě. Autokorelační koeficient ϱ je korelační koeficient mezi sousedními chybami použité metody, tedy ε_n a ε_{n+1} . Ze stacionarity chyb ovšem dostáváme, že je to také korelační koeficient mezi ε_{n+1} a ε_{n+2} . Pro ε_{n+1} platí vzorec (3.1.7) a ε_{n+2} vyjádříme s využitím (3.1.7) a vzcrců (3.1.8) a (3.1.9) jako

$$\begin{aligned}
 \varepsilon_{n+2} &= e_{n+2} + (S_{n+1} - \tilde{S}_{n+1}) + (T_{n+1} - \tilde{T}_{n+1}) = \\
 &= e_{n+2} + (1 - \alpha_2 - \gamma_2\alpha_2)(S_n - \tilde{S}_n) + \\
 &\quad + (2 - \alpha_2 - \gamma_2\alpha_2)(T_n - \tilde{T}_n) + \\
 &\quad + (\gamma_1\alpha_1 - \gamma_2\alpha_2 + \alpha_1 - \alpha_2)e_{n+1}.
 \end{aligned} \tag{3.2.1}$$

Nyní spočítejme kovarianci chyb s využitím nekorelovanosti e_{n+2} a e_{n+1} s $S_n - \tilde{S}_n$ i $T_n - \tilde{T}_n$ a také nekorelovanosti e_{n+2} s e_{n+1} navzájem, takto

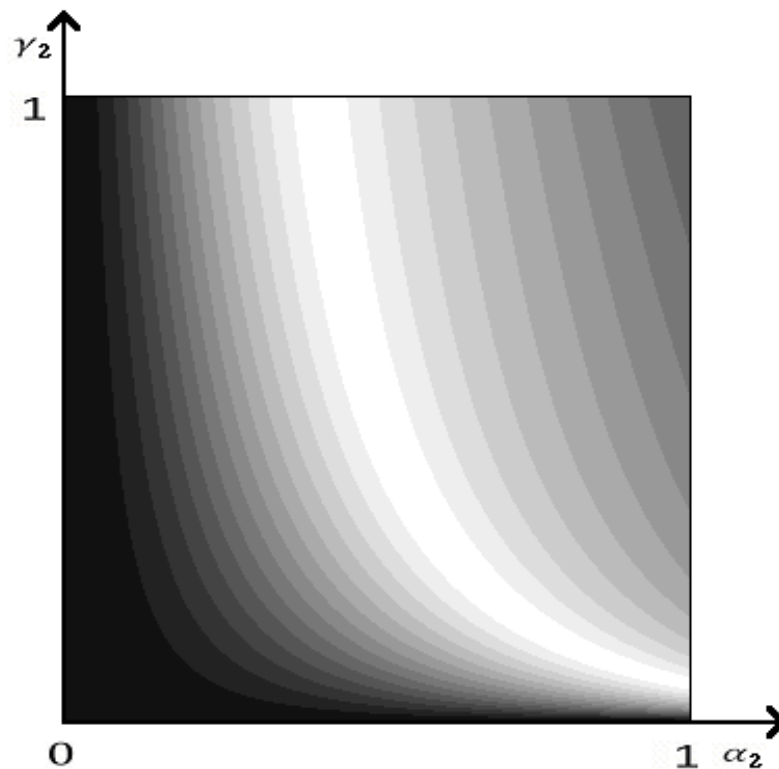
$$\begin{aligned} cov(\varepsilon_{n+1}, \varepsilon_{n+2}) &= (\alpha_1 - \alpha_2 + \gamma_1 \alpha_1 - \gamma_2 \alpha_2) \sigma^2 + (1 - \alpha_2 - \gamma_2 \alpha_2) v_S + \\ &+ (2 - \alpha_2 - \gamma_2 \alpha_2) v_T + (3 - 2\alpha_2 - 2\gamma_2 \alpha_2) w. \end{aligned} \quad (3.2.2)$$

Uvědomme si, že $var(\varepsilon_{n+1})$ jsme již počítali jako MSE, to samé platí díky stacionaritě i pro $var(\varepsilon_{n+2})$. Nyní si tedy můžeme hledaný autokorelační koeficient ρ vyjádřit jako

$$\rho = corr(\varepsilon_{n+1}, \varepsilon_{n+2}) = \frac{cov(\varepsilon_{n+1}, \varepsilon_{n+2})}{\sqrt{var(\varepsilon_{n+1})var(\varepsilon_{n+2})}} = \frac{cov(\varepsilon_{n+1}, \varepsilon_{n+2})}{MSE}, \quad (3.2.3)$$

kde $cov(\varepsilon_{n+1}, \varepsilon_{n+2})$ máme již vyjádřenou v (3.2.2) a MSE v (3.1.20).

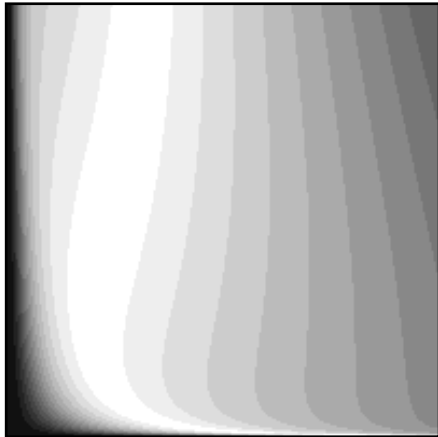
Tímto jsme tedy našli $\rho(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$, což znamená způsob, jak spočítat autokorelační koeficient ρ pro libovolnou kombinaci parametrů $\alpha_1, \gamma_1, \alpha_2, \gamma_2 \in (0,1)$. Stejně jako v případě $MSE(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$ si pro ilustraci tuto funkci znázorníme na obrázku 3.2-1 pomocí vrstevnic. Zvolme znovu generující konstanty $\alpha_1 = 0,5, \gamma_1 = 0,5$ a na osy vynesme hodnoty α_2, γ_2 . Vrstevnice nyní představují hodnotu autokorelačního koeficientu ρ pro dané kombinace konstant v absolutní hodnotě. Přechod ze světlejší barvy do tmavší vyjadřuje přírůstek absolutní hodnoty ρ o 0,067. Nejsvětlejší oblast tedy označuje body, pro které je $|\rho| \leq 0,067$ (například včetně bodu $\alpha_2 = 0,5; \gamma_2 = 0,5$), nalevo od této oblasti je $\rho > 0$ a směrem k bodu (0,0) se jeho hodnota zvyšuje až k jedné a naopak napravo je $\rho < 0$ a směrem k bodu (1,1) se jeho hodnota zmenšuje.



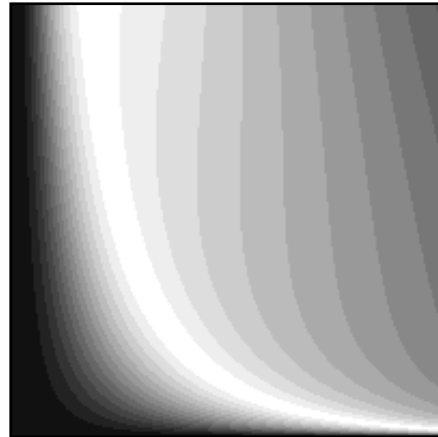
Obrázek 3.2-1 $\rho(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$ pro $\alpha_1 = 0,5; \gamma_1 = 0,5$

Je zajímavé si povšimnout, že oblast nejsvětější barvy, na které leží body, pro něž přibližně vychází $\rho = 0$, tvoří křivku připomínající hyperbolu. Podobně jako v případě $MSE(\alpha_1, \gamma_1, \alpha_2, \gamma_2)$, pokud se α_2 nebo γ_2 blíží 0, koeficient autokorelace ρ se blíží 1, kdežto v blízkosti bodu (1,1) je ρ omezen hodnotou větší než -1 .

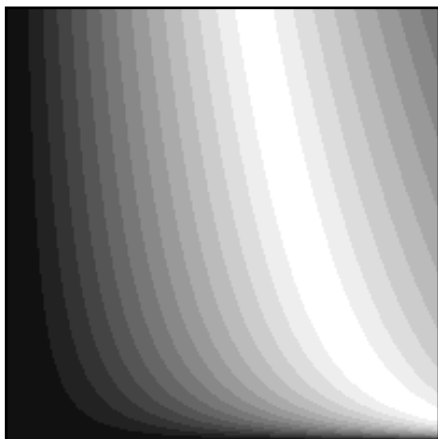
Nyní si pro ilustraci na následujících obrázcích ukažme, jak se situace změní, zvolíme-li jinak hodnoty α_1, γ_1 . Opět pro všechny čtyři následující obrázky 3.2-2 až 3.2-5 platí popis stejný jako pro obrázek 3.2-1.



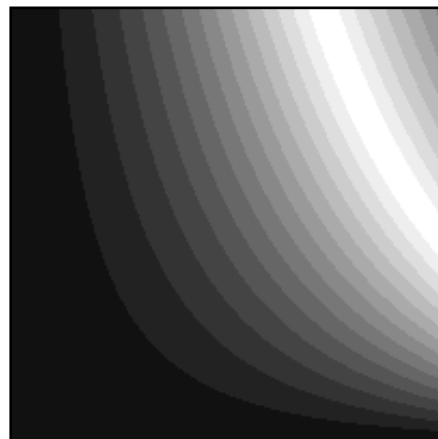
Obrázek 3.2-2 ρ pro $\alpha_1 = 0, 2$; $\gamma_1 = 0, 2$



Obrázek 3.2-3 ρ pro $\alpha_1 = 0, 2$; $\gamma_1 = 0, 8$



Obrázek 3.2-4 ρ pro $\alpha_1 = 0, 8$; $\gamma_1 = 0, 2$



Obrázek 3.2-5 ρ pro $\alpha_1 = 0, 8$; $\gamma_1 = 0, 8$

4 Aplikace na Brownovu metodu

V tomto odstavci se budeme zabývat především otázkou, jak velké chyby se dopustíme, pokud místo Holtovy metody použijeme Brownovu. To by nám potom mohlo napovědět, zda je použití Brownovy metody opravdu tak nevýhodné oproti metodě Holtově, nebo přeci jenom jeho používání může mít za určitých situací smysl, jak jsme již nastínili v odstavci 2.2.

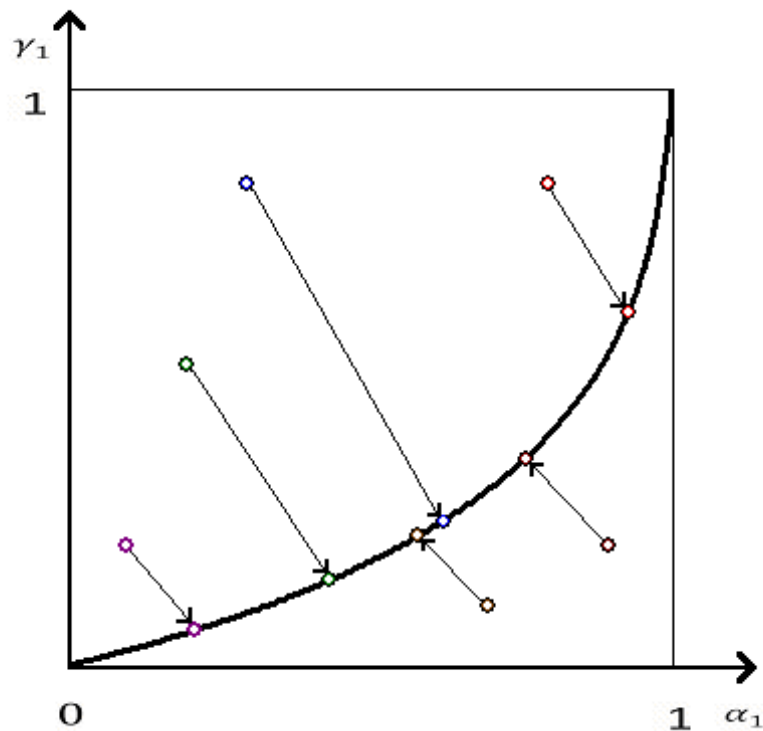
Představme si, že máme řadu vytvořenou opět rovnicemi generujícího schématu (3.1.1), (3.1.2) a (3.1.3) s vyrovnávacími konstantami $\alpha_1, \gamma_1 \in (0,1)$. Pokud k vyrovnání této řady použijeme Holtovu metodu s vyrovnávacími konstantami $\alpha_2, \gamma_2 \in (0,1)$, známe již z odstavce 3.1 výsledek a víme, jak kvalitní předpověď dostaneme. Nyní chceme znát odpověď na otázku, jak kvalitní bude předpověď, když na tuto řadu použijeme Brownovu metodu s vyrovnávací konstantou $\alpha \in (0,1)$. To je však pro nás jednoduchý úkol, uvědomíme-li si, že již z odstavce 2.3 známe vztahy mezi konstantami Holtovy a Brownovy metody, tedy vzorce (2.3.15)

$$\alpha_H = \alpha(2 - \alpha) \quad \text{a} \quad \gamma_H = \frac{\alpha}{2 - \alpha}. \quad (4.1.1)$$

Pomocí těchto vzorců lze Brownovu metodu s libovolnou vyrovnávací konstantou $\alpha \in (0,1)$ považovat za ekvivalentní Holtovu metodu s vyrovnávacími konstantami α_H a γ_H , pro něž již umíme spočítat příslušné hodnoty MSE a ϱ pomocí vzorců (3.1.20) a (3.2.3). Máme-li nyní časovou řadu vytvořenou pomocí generujícího schématu s vyrovnávacími konstantami α_1, γ_1 a chceme nalézt Brownovu metodu s optimální vyrovnávací konstantou α , provedeme to procházením přípustných hodnot $\alpha \in (0,1)$. Pro každou tuto vybranou hodnotu α si spočteme pomocí vzorce (3.1.20) hodnotu MSE a na závěr vybereme α s nejnižší hodnotou MSE. Vyjádřeno vzorcem hledáme

$$\underset{\alpha \in (0,1)}{\operatorname{argmin}} \operatorname{MSE}(\alpha_1, \gamma_1, \alpha_H(\alpha), \gamma_H(\alpha)). \quad (4.1.2)$$

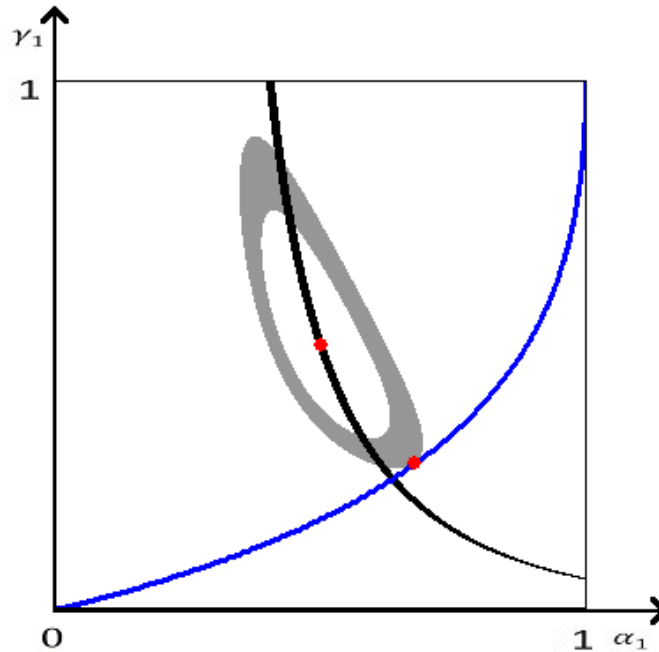
Pro lepší představu si na následujícím obrázku 4.1-1 znázorníme, kde leží optimální α Brownovy metody pro různé kombinace generujících konstant α_1, γ_1 , které jsou vyneseny na osách. Na obrázku 4.1-1 je znázorněna Brownovská diagonála z obrázku 2.3-1, jak jsme si tuto křivku v odstavci 2.3 nazvali. Mimo diagonálu leží body odpovídající kombinaci generujících konstant α_1, γ_1 , ze kterých vede vždy šipka přiřazující jim bod na diagonále odpovídající optimální hodnotě α Brownovy metody.



Obrázek 4.1-1 Optimální hodnota α Brownovy metody pro různé kombinace α_1, γ_1

Všimněme si, že ne vždy je bodu mimo diagonálu přiřazen jemu nejbližší bod (ve smyslu eukleidovské metriky) na diagonále. Vrstevnice MSE, jak je vidíme na obrázcích 3.1-1 až 3.1-5 netvoří totiž kružnice, a proto se mohou diagonály dotknout v jiném než nejbližším bodě.

Zvolme nyní kombinaci generujících konstant, například $\alpha_1 = 0,5$, $\gamma_1 = 0,5$. Pomocí výše zmíněného postupu spočteme optimální hodnotu $\alpha = 0.432$ odpovídající bodu na brownovské diagonále, Brownova metoda s touto vyrovnávací konstantou dosahuje $MSE = 1.0414$ a $\rho = -0.09003$. Na následujícím obrázku 4.1-2 si zobrazme opět Brownovskou diagonálu, bod mimo diagonálu reprezentující generující kombinaci konstant $\alpha_1 = 0,5$, $\gamma_1 = 0,5$, bod na diagonále odpovídající optimálnímu α , šedou plochu, která označuje množinu bodů, pro které hodnota MSE leží v intervalu $(1.03, 1.0414)$, tedy vlastně vrstevnici podobnou těm na obrázku 3.1-1 a nakonec ještě vidíme černě oblast bodů, ve kterých hodnota autokorelačního koeficientu ρ leží mezi $(-0.0125, 0.0125)$, opět tedy vidíme jednu vrstevnici podobnou těm z obrázku 3.2-1.

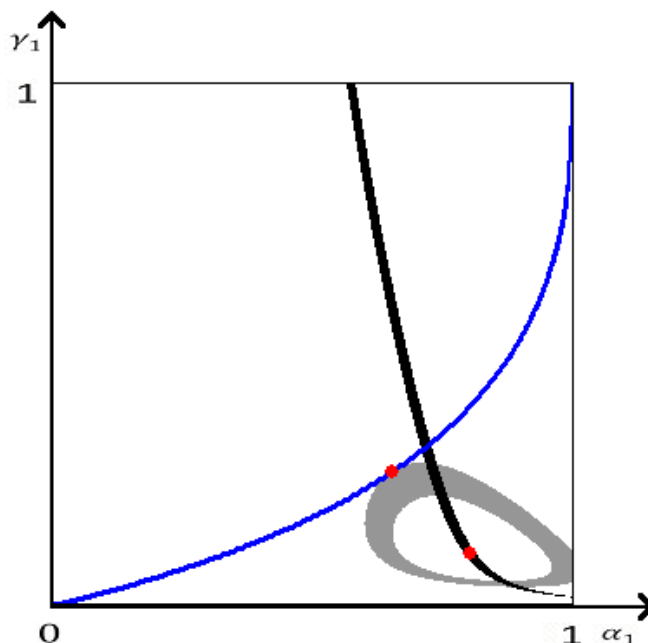


Obrázek 4.1-2 Hledání optimálního α Brownovy metody pro $\alpha_1 = 0,5, \gamma_1 = 0,5$

Všimněme si, že bod mimo diagonálu leží v černé množině bodů, je tomu ze stejného důvodu jako na obrázku 3.2-1. Když na řadu generovanou schématem s konstantami $\alpha_1 = 0,5, \gamma_1 = 0,5$ použijeme Holtovu metodu se stejnými konstantami, vyjde autokorelační koeficient ρ vždy nulový. Bod na diagonále leží na okraji šedého pásu, v místě dotyku s brownovskou diagonálou. Je to jediný bod diagonály, ve kterém je hodnota MSE takto nízká.

Dále si povšimněme polohy bodu na diagonále vzhledem k vrstevnici přibližně nulové hodnoty autokorelačního koeficientu ρ . Tento bod leží nalevo od této vrstevnice, a proto by měla být v bodě na diagonále, stejně jako to platí pro body na obrázku 3.2-1, hodnota ρ záporná. Opravdu, jak jsme již dříve spočetli, je v tomto bodě autokorelační koeficient $\rho = -0.09003 < 0$.

Pro ukázkou zopakujme postup ještě pro jiný bod. Zvolme nyní bod pod diagonálou, například bod $\alpha_1 = 0,8, \gamma_1 = 0,1$, který představuje dvojici generujících konstant a pro něj sestavme analogicky obrázek 4.1-3.



Obrázek 4.1-3 Hledání optimálního α Brownovy metody pro $\alpha_1 = 0,8, \gamma_1 = 0,1$

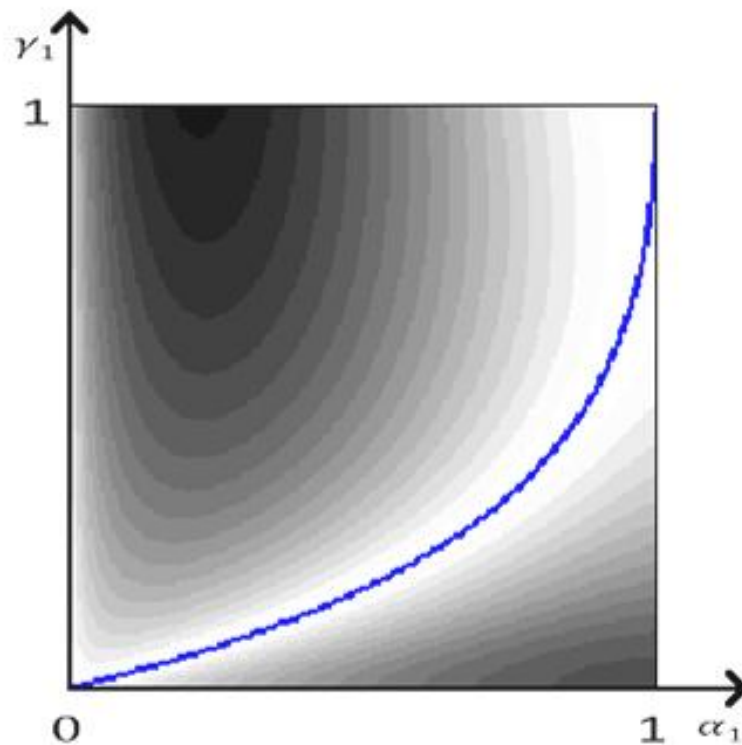
Na obrázku vidíme vrstevnice MSE v intervalu $(1.025, 1.04854)$ a ϱ v intervalu $(-0.0125, 0.0125)$. Tentokrát však leží bod na diagonále nalevo od nulové vrstevnice ϱ , proto by v tomto bodě měl být autokorelační koeficient kladný, stejně jako na obrázku 3.2-1. Ověření výpočtem nám pro $\alpha_1 = 0,8, \gamma_1 = 0,1$ dává optimální vyrovnávací konstantu použité Brownovy metody $\alpha = 0.408$ a v příslušném bodě na diagonále také $MSE = 1.04854$ a $\varrho = 0.094824 > 0$.

Je zajímavé si povšimnout, že na obrázku 4.1-2 ležel bod mimo diagonálu nad brownovskou diagonálou a jemu příslušný bod na diagonále měl záporný autokorelační koeficient ϱ , oproti tomu na obrázku 4.1-3 zakreslený bod mimo diagonálu ležel pod diagonálou a jemu příslušný autokorelační koeficient ϱ byl kladný. Mohlo by nás tedy napadnout, jestli je to pouze náhoda, nebo je to pravidlem, že body nad brownovskou diagonálou mají kladný koeficient ϱ a ty pod ní ho mají záporný.

Podívejme se tedy na další obrázek 4.1-4. Na vodorovné ose je opět vynesena vyrovnávací konstanta α_1 a na svislé γ_1 , pomocí generujícího schématu s těmito vyrovnávacími konstantami tedy budeme generovat řadu, na niž použijeme Brownovu metodu s optimální vyrovnávací konstantou α . Opět pomocí vrstevnic budeme v jednotlivých bodech (α_1, γ_1) vynášet hodnotu autokorelačního koeficientu Brownovy metody s optimální vyrovnávací konstantou α .

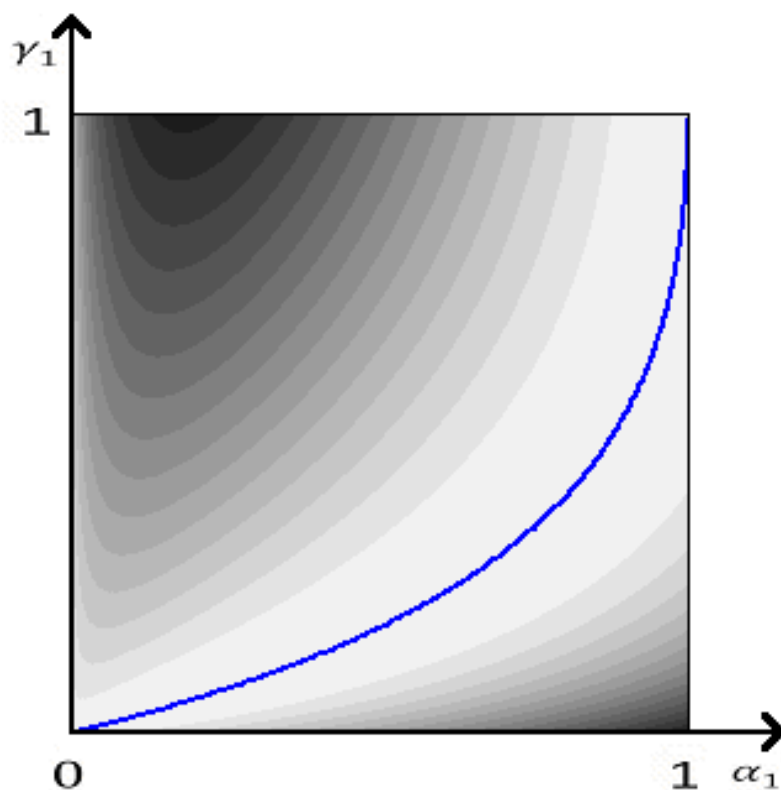
Nejsvětlejší oblast znázorňuje oblast kolem brownovské diagonály, na níž je hodnota koeficientu ϱ blízká nule, protože v bodech diagonály generující konstanty α_1, γ_1 přesně odpovídají konstantě α použité Brownovy metody, tj. pro $\alpha, \alpha_1, \gamma_1$

platí rovnost ze vzorců (4.1.1). Jinými slovy, pokud konstanty α_1, γ_1 leží na brownovské diagonále, potom použitá Brownova metoda s vyrovnávací konstantou α bude ekvivalentní Holtově metodě s vyrovnávacími konstantami α_1, γ_1 , a tudíž v těchto bodech bude mít Brownova metoda nulový autokorelační koeficient ρ a minimální hodnotu MSE. Dále můžeme na obrázku vidět, že směrem do bodu (0,1) koeficient ρ nabývá záporných hodnot a postupně snižuje svojí hodnotu. Směrem od diagonály do bodu (1,0) nabývá koeficient ρ kladných hodnot a postupně zvyšuje svojí hodnotu. Pro úplnost dodejme, že přechod z vrstevnice světlejší barvy do tmavší znamená zvýšení absolutní hodnoty ρ o 0.009.



Obrázek 4.1-4 Autokorelační koeficient ρ při použití Brownovy metody

Na závěr si ještě na obrázku 4.1-5 vykresleme opět pomocí vrstevnic, jak se mění hodnoty MSE při použití Brownovy metody s optimální konstantou α na řady generované konstantami α_1, γ_1 . Přechod ze světlejší vrstevnice do tmavší znamená zvýšení hodnoty MSE o 0.012. Podél diagonály je hodnota $MSE = 1$ (předpokládáme-li stále $\sigma^2 = 1$) a směrem od diagonály do rohů se jeho hodnota zvyšuje.



Obrázek 4.1-5 MSE při použití Brownovy metody

Lze si povšimnout, že nejvyšší hodnoty MSE (a také nejvyšší absolutní hodnoty ϱ), se nedosahuje v těsné blízkosti bodu $(0,1)$, jak by se při letmém pohledu mohlo zdát, ale teprve v místě napravo od něj. Je to způsobeno tím, že když se hodnota vyrovnávací konstanty α_1 Holtovy metody blíží 0, tak potom ztrácí hodnota γ_1 na významu, protože se vyskytuje pouze ve vzorci (3.1.2) v součinu s téměř nulovou hodnotou α_1 .

Všimněme si, že poslední dva obrázky 4.1-4 a 4.1-5 spolu poměrně dobře korespondují. Na diagonále je hodnota MSE minimální a $\varrho = 0$, směrem od diagonály se zvětšuje jak hodnota MSE, tak absolutní hodnota koeficientu ϱ . To jestli jsme pod a nebo nad diagonálou nám určuje znaménko koeficientu ϱ .

Pokusme se nyní shrnout některé možné praktické závěry plynoucí z výsledků čtvrté kapitoly. Představme si nejprve, že z nějakých důvodů nemůžeme nebo nechceme používat Holtovu metodu a k dispozici máme pouze metodu Brownovu. Jaká doporučení nám dají naše zjištění z kapitoly 4? Předně srovnáním obrázků 4.1-4 a 4.1-5 ve shodě s článkem Cipra, Hanzák (2008) docházíme k závěru, že pokud nám při použití Brownovy metody vyjde hodnota koeficientu ϱ blízká nule, dosáhli jsme zároveň hodnoty MSE, kterou by ani použití Holtovy metody nemohlo výrazně zlepšit.

Pokud nám vyjde koeficient ρ různý od nuly, jeho znaménko nám může napovědět, jaké by měly být konstanty Holtovy metody, abychom dosáhli při jejím použití optimálních výsledků. Pokud je ρ záporný, znamená to, že jsme v situaci jako na obrázku 4.1-2, optimální vyrovnávací konstanty Holtovy metody se nachází v bodě nad diagonálou a při použití Holtovy metody bychom měli zmenšit konstantu α_H a zvětšit γ_H (oproti těm odpovídajícím vyrovnávací konstantě Brownovy metody ve smyslu vzorců (2.3.15)). Pokud je naopak ρ kladný, jsme v situaci jako na obrázku 4.1-3 a měli bychom konstantu zvětšit α_H a zmenšit γ_H . Shrňme-li naše zjištění, znaménko nám určuje směr, kterým bychom měli měnit konstanty a absolutní hodnota vzdálenost o jakou bychom se měli posunout.

Naopak pokud bychom znali výstup (MSE_H) Holtovy metody s optimálními koeficienty, mohli bychom z nich usoudit na výsledky Brownovy metody, aniž bychom znali data, na která byla Holtova metoda použita. Ze znalosti optimálních koeficientů Holtovy metody totiž, jak už jsme to dělali na obrázku 4.1-6, umíme zjistit optimální hodnotu vyrovnávací konstanty Brownovy metody α za pomoci vzorce (3.1.20), pro tuto hodnotu již není problém ze vzorců (3.1.20) a (3.2.3) spočítat hodnoty MSE a ρ . Jediný rozdíl by byl, že bychom již nepředpokládali $\sigma^2 = 1$, ale položili bychom $\sigma^2 = MSE_H$ (viz komentář k vzorci (3.1.20)).

5 Numerické aplikace

V tomto odstavci si teoretické výsledky, ke kterým jsme došli v předchozích kapitolách, ověříme nejprve na časových řadách generovaných procesem ARIMA(0, 2, 2) a následně vyzkoušíme, zda závěry platí i pro vhodné reálné časové řady.

5.1 Ověření vzorců pomocí simulace

Abychom ověřili správnost v kapitole 3 odvozených vzorců, použili jsme program MS Excel, ve kterém jsme vytvořili časovou řadu o 50 000 pozorování pomocí generujícího schématu s α_1, γ_1 , tedy pomocí vzorců (3.1.1), (3.1.2) a (3.1.3) (tzn. generovanou procesem ARIMA(0, 2, 2)). Na vytvořenou řadu jsme potom použili Holtovu metodu danou vzorci (3.1.4), (3.1.5) a (3.1.6) s vyrovnávacími konstantami α_2, γ_2 , poté jsme spočetli jednokrokové předpovědní chyby této metody ε_n a jejich rozptyl $var(\varepsilon_n)$, což je hodnota MSE pro použitou metodu. Spočtené hodnoty MSE jsme pak pro dané kombinace konstant zapsali do tabulky 5.1-1 do sloupce označeného simulované MSE. Podobně jsme poté spočetli simulované hodnoty ρ jako $corr(\varepsilon_n, \varepsilon_{n+1})$.

Dále jsme využili našich teoreticky odvozených vzorců (3.1.20) a (3.2.3) a podle nich spočetli hodnoty MSE a ρ pro dané kombinace konstant. Tyto hodnoty jsme také zapsali do tabulky 5.1-1 do sloupců označených teoreticky spočtené MSE a ρ .

α_1	γ_1	α_2	γ_2	Simulované MSE	Teoreticky spočtené MSE	Simulované ρ	Teoreticky spočtené ρ
0,7	0,1	0,7	0,1	0,9966	1	0,0053	0
0,9	0,1	0,8	0,2	1,0208	1,0239	0,0446	0,0471
0,4	0,5	0,3	0,6	1,0305	1,0288	0,1324	0,1348
0,8	0,8	0,9	0,9	1,0687	1,0745	-0,2511	-0,2558
0,1	0,9	0,2	0,8	1,1454	1,1295	-0,0567	-0,0544
0,3	0,4	0,9	0,7	2,2568	2,2641	-0,5175	-0,5191
0,6	0,6	0,2	0,3	5,445	5,605	0,8884	0,8913
0,8	0,7	0,2	0,4	10,1143	10,2045	0,932	0,9325
0,8	0,9	0,1	0,2	128,34	129,33	0,9885	0,9887

Tabulka 5.1-1 Srovnání výsledků teoreticky odvozených vzorců a simulace pomocí ARIMA(0,2,2)

Srovnáme-li tučně zvýrazněné hodnoty MSE a ρ získané pomocí obou postupů, zjistíme, že rozdíly mezi příslušnými hodnotami jsou nepatrné. Tímto se nám tedy podařilo ověřit platnost výše zmíněných vzorců (3.1.20) a (3.2.3) za předpokladu generování řady procesem ARIMA(0, 2, 2).

5.2 Vyzkoušení vzorců na reálných časových řadách

Nyní zbývá ještě ověřit, zda teoreticky odvozené vzorce fungují i pro vhodné (nesezónní s lokálně lineárním trendem) reálné časové řady. Zde budeme srovnávat hodnoty spočtené podle vzorců (3.1.20) a (3.2.3) s výstupy statistického softwaru. Budeme používat statistický software DMITS, který je součástí diplomové práce Hanzák (2007). Použité časové řady y_1 a y_2 jsme převzali ze stránek Time Series Data Library www.robjhyndman.com/TSDL.

Jako první uvažujme časovou řadu y_1 roční spotřeby kávy v USA v letech 1910-1970 o 61 pozorováních. Použijeme-li na ni ve statistickém softwaru Holtovu metodu s optimalizovanou volbou konstant, zjistíme, že optimální vyrovnávací konstanty této metody jsou $\alpha_{opt} = 0,571973$, $\gamma_{opt} = 0,038043$, dále z výstupů softwaru vyčteme, že $MSE = 1,21785$ a koeficient autokorelace $\rho = 0,011326$.

Malá hodnota koeficientu autokorelace nám umožňuje přibližně považovat tento koeficient za nulový (p-hodnota standardního testu nulovosti vyšla 0,93), což nám zároveň umožňuje předpokládat, že časová řada y_1 byla generována procesem ARIMA(0, 2, 2) s parametry odpovídajícími nalezeným optimálním vyrovnávacím konstantám $\alpha_{opt} = 0,5720$, $\gamma_{opt} = 0,03804$ Holtovy metody. Hodnota MSE potom odpovídá pouze velikosti rozptylu předpovědních chyb e_n v dané řadě, tedy $\sigma^2 = 1,21785$.

Zkusme nyní využít našich výsledků z kapitoly 4 a spočítejme, jak by teoreticky mělo dopadnout použití Brownovy metody na řadu y_1 . Optimální vyrovnávací konstantu Brownovy metody volíme minimalizací hodnoty MSE, k jehož výpočtu používáme vzorec (3.1.20). Tak nám vyjde $\alpha = 0.2704$, dále hodnota $MSE = 1.2757$ a $\rho = 0.0902$, a to podle vzorce (3.2.3).

Nyní na časovou řadu y_1 použijeme opět prostřednictvím statistického softwaru Brownovu metodu s optimalizovanou volbou vyrovnávací konstanty α . Tímto postupem vyjde optimální hodnota $\alpha = 0,242841$, $MSE = 1,27378$ a $\rho = 0,149959$. Pro lepší přehlednost uvedeme srovnání výsledků teoreticky odvozených vzorců a výstupů statistického programu v tabulce 5.2-1. Stejně srovnání provedeme ve zmíněné tabulce zároveň pro řadu y_2 . Řada y_2 o 30 pozorováních potom obsahuje roční počty stávek v USA mezi lety 1951 a 1980 a řada y_3 o 59 pozorováních zachycuje roční vývoj HDP Brazílie v letech 1900-1958 (zdroj: Duke University, www.stat.duke.edu). Pro ilustraci uvádíme v tabulce 5.2-1 vždy u Brownovy metody jednak vyrovnávací konstantu α , ale také vyrovnávací konstanty ekvivalentní Holtovy metody α_H, γ_H spočtené pomocí vzorců (2.3.15).

	statistický software					teoreticky odvozené vzorce		
	Holtova metoda		Brownova metoda			Brownova metoda		
	α_{opt} γ_{opt}	MSE ϱ	α	α_H γ_H	MSE ϱ	α	α_H γ_H	MSE ϱ
y_1	0,5720 0,0380	1,2179 0,0113	0,2428	0,4267 0,1382	1,2738 0,1500	0,2704	0,4316 0,1403	1,2757 0,0902
y_2	0,3862 0,5814	299579 0,0334	0,3876	0,6250 0,2404	316493 -0,1032	0,3904	0,6284 0,2425	323353 -0,1199
y_3	0,8235 0,4648	16,2076 -0,0203	0,6087	0,8649 0,4375	16,2154 -0,0302	0,6076	0,8460 0,4364	16,2173 -0,0088

Tabulka 5.2-1 Použití teoretických vzorců na reálné časové řady

Srovnáme-li tučně vyznačené údaje z tabulky, zjistíme, že se opravdu hodnoty získané použitím našich teoretických vzorců a údaje ze statistického softwaru zásadně neliší. Malé odlišnosti jsou způsobeny jednak tím, že jsme předpokládali nulovou hodnotu koeficientu ϱ při použití Holtovy metody, což v ani jedné z reálných časových řad beze zbytku splněno nebylo. Dále se pozorované řady mohly poněkud lišit od řad generovaných procesem ARIMA(0, 2, 2) a v neposlední řadě byly všechny tři řady poměrně krátké, takže i kdyby opravdu byly generovány procesem ARIMA(0, 2, 2), mohou se nalezené hodnoty vyrovnávacích konstant, spočtená MSE a ϱ lišit od jejich teoretických hodnot, jejichž jsou vlastně bodovými odhady. Celkově ovšem výsledky zapsané v tabulce 5.2-1 spíše potvrzují použitelnost našich výsledků i v praxi.

Podívejme se ještě na druhý závěr z konce kapitoly 4. Předpokládejme, že známe jenom výsledky softwarově realizované Brownovy metody. Je-li koeficient ϱ záporný (řada y_2), měla by být optimální vyrovnávací konstanta Holtovy metody α_{opt} menší než α_H a γ_{opt} větší než γ_H . V tabulce 5.2-1 vidíme, že toto doporučení se ukázalo správným a analogicky platí i pro řadu y_2 . Pro řadu y_3 zase platí další závěr z konce kapitoly 4, že pokud Brownova metoda dosáhne téměř nulového koeficientu ϱ , tak to znamená, že se nacházíme v těsné blízkosti optima a použití Holtovy metody nám již hodnotu MSE příliš nesníží.

6 Závěr

Zadáním práce bylo vyhledat v literatuře metody typu exponenciálního vyrovnávání a popsat je. Dále se pak zaměřit na některý z problémů praktického použití těchto metod v praxi. Součástí práce měly být příklady použití na reálné či simulované časové řady.

Vzniklá práce se zabývá výhradně nesezónními časovými řadami s lokálně lineárním trendem. K vyrovnávání a předpovídání takovýchto řad slouží z metod typu exponenciálního vyrovnávání především Holtova metoda a dvojitě exponenciální vyrovnávání neboli Brownova metoda. Obě metody byly v odstavcích 2.1 a 2.2 popsány, dále byl v odstavci 2.2 ukázán vztah mezi Holtovou metodou a procesem ARIMA(0, 2, 2) i vícerozměrným procesem VAR(1). Dále byl odvozen vztah mezi oběma metodami (odstavec 2.3), konkrétně byly odvozeny vzorce, podle nichž se dají vyrovnávací konstanty Holtovy metody vyjádřit pomocí vyrovnávací konstanty Brownovy metody.

Abychom mohli v odstavcích 3.1 a 3.2 teoreticky odvodit vzorce pro MSE a autokorelační koeficient předpovědních chyb q Holtovy metody, předpokládali jsme použití metody na časovou řadu generovanou procesem ARIMA(0, 2, 2), v jehož vyjádření jsme použili vztah mezi Holtovou metodou a procesem ARIMA(0, 2, 2) z kapitoly 3.2. Pro větší názornost jsme pro konkrétní koeficienty procesu ARIMA(0, 2, 2) znázornili oba vzorce graficky.

V kapitole 5 jsme mohli využít vztahu mezi Holtovou metodou a Brownovou metodou pro odvození vztahů pro hodnoty MSE a autokorelačního koeficientu předpovědních chyb q použité Brownovy metody. Všechny výsledky jsme opět ilustrovali grafickým zobrazením a formulovali některé praktické závěry, které podle autora přináší lepší vhled do otázky volby mezi Holtovou a Brownovou metodou.

Na závěr jsme v kapitole 6 simulačně ověřili teoreticky odvozené vzorce a vyzkoušeli použitelnost vzorců na reálných časových řadách.

Literatura

Chatfield, C. (2002): Time-Series Forecasting. Chapman & Hall/CRC.

Cipra, T. (1986): Analýza časových řad s aplikacemi v ekonomii, Praha , SNTL, 1986

Cipra, T., Hanzák, T. (2008): Exponential smoothing for irregular time series, Kybernetika, Volume 44, str. 385-399

Hanzák, T. (2007): Dekompoziční metody pro časové řady s nepravidelně pozorovanými hodnotami, diplomová práce MFF UK

Winters, P. R. (1960): Forecasting sales by exponentially weighted moving averages, Management Science 6, str. 324-342