

# Stochastické modely v ekonomii

---

## 1.1 Základní přístupy k analýze časových řad

V této části se omezíme jen na stručnou charakteristiku čtyř nejčastěji používaných přístupů k analýze časových řad. Uvažujme časovou řadu  $Y_1, Y_2, \dots, Y_n$ .

### A. Dekompozice časové řady.

Princip tohoto přístupu je velmi jednoduchý. Časová řada se rozkládá na čtyři základní složky, jimiž jsou trend ( $Tr$ ), sezónní složka ( $Sz$ ), cyklická složka ( $C$ ) a náhodná (reziduální) složka ( $\varepsilon$ ). To znamená, že časovou řadu chápeme jako trend, na který se „nabalují“ periodické složky (sezónní a cyklická složka) a náhodná složka (nejčastěji bílý šum).

Dekompozice (rozklad) časové řady je dvojího typu:

a) **aditivní** ve tvaru

$$Y_t = Tr_t + Sz_t + C_t + \varepsilon_t,$$

kdy všechny složky se měří ve stejných jednotkách jako  $Y_t$ ,

b) **multiplikativní** ve tvaru

$$Y_t = Tr_t Sz_t C_t \varepsilon_t,$$

kdy pouze trendová složka je měřena ve stejných jednotkách jako  $Y_t$  a ostatní složky jsou bezrozměrné veličiny.

### B. Boxova-Jenkinsova metodologie

U tohoto přístupu se zpravidla předpokládá, že časová řada je slabě stacionární. Základním prvkem při konstrukci modelu je reziduální složka. Uvedeme dva typické příklady modelu:

- model klouzavých součtů 1. řádu ve tvaru

$$Y_t = \varepsilon_t + k\varepsilon_{t-1},$$

kde  $k$  je nějaká reálná konstanta a  $\varepsilon_t$  reprezentuje tzv. bílý šum;

- autoregresní model 1. řádu definovaný předpisem

$$Y_t = \varepsilon_t + kY_{t-1}.$$

Boxova-Jenkinsova metodologie umožňuje modelovat i časové řady s výrazným trendovým a/nebo sezónním charakterem, přičemž trendová i sezónní složka mohou být (na rozdíl od dekompozice) modelovány stochasticky.

Boxovy-Jenkinsovy modely jsou mnohem flexibilnější než modely dekompoziční, což znamená, že se lépe adaptují na změny v průběhu časové řady.

Základním matematickým nástrojem pro analýzu časové řady jsou v tomto případě **metody korelační analýzy**, které umožňují zkoumat závislosti mezi jednotlivými pozorováními dané časové řady.

### C. Lineární kauzální (faktorové) modely

Takové modely, běžné v ekonometrii, jsou zpravidla konstruovány tak, že se hodnoty sledované časové řady vysvětlují pomocí jiných, tzv. **faktorových časových řad**. Uvedeme jednoduchý ekonometrický model, převzatý z monografie [8], ve tvaru

$$C_t = \alpha + \beta C_{t-1} + \gamma X_t + \delta P_t + \varepsilon_t.$$

V uvažovaném modelu se výdaje  $C_t$  obyvatelstva na nákup spotřebního zboží v roce  $t$  vysvětlují pomocí těchto výdajů  $C_{t-1}$  v roce bezprostředně předcházejícím a navíc pomocí peněžních příjmů  $X_t$  obyvatelstva a cenového indexu  $P_t$  spotřebního zboží v roce  $t$ .

V této oblasti se používají především metody regresní analýzy.

### D. Spektrální analýza časových řad

Zkoumaná časová řada se považuje za nekonečnou lineární kombinaci sinusových a kosinusových funkcí s různými amplitudami a frekvencemi. Pomocí speciálních statistických nástrojů (např. periodogram) je možno získat představu o intenzitě zastoupení jednotlivých frekvencí v časové řadě. Ve spektrální analýze časových řad se využívá především **Fourierovy analýzy**.

## 1.2 Základní charakteristiky časových řad

Základními **teoretickými charakteristikami** časové řady jsou:

**střední hodnota**  $\mu_t = \mathbf{E}(Y_t)$ ,

**rozptyl (variance)**  $\sigma_t^2 = \mathbf{var}(Y_t) = \mathbf{E}(Y_t - \mu)^2$ ,

**autokovarianční funkce řádu  $k$  ( $k = 0, \pm 1, \dots$ )**

$$\gamma_k = \text{cov}(Y_t, Y_{t+k}) = \mathbf{E}(Y_t - \mu_t)(Y_{t+k} - \mu_{t+k})$$

**autokorelační funkce řádu  $k$  ( $k = 0, \pm 1, \dots$ )**

$$\rho_k = \frac{\text{cov}(Y_t, Y_{t+k})}{\sigma_t \sigma_{t+k}}$$

Uvedené vztahy se zjednoduší, omezíme-li se na tzv. slabě stacionární řady, tj. na řady s konstantní střední hodnotou, konstantním rozptylem a autokovarianční funkcí, která závisí pouze na hodnotě  $k$ . Pak pro uvedené teoretické charakteristiky můžeme psát:

- a) **střední hodnota**  $\mu = \mathbf{E}(Y_t)$  pro všechna  $t$ ,
- b) **rozptyl (variance)**  $\sigma^2 = \text{var}(Y_t) = \mathbf{E}(Y_t - \mu)^2$  pro všechna  $t$ ,
- c) **autokovarianční funkce řádu  $k$  ( $k = 0, \pm 1, \dots$ )**

$$\gamma_k = \text{cov}(Y_t, Y_{t+k}) = \mathbf{E}[(Y_t - \mu)(Y_{t+k} - \mu)],$$

- d) **autokorelační funkce řádu  $k$  ( $k = 0, \pm 1, \dots$ )**

$$\rho_k = \frac{\text{cov}(Y_t, Y_{t+k})}{\sigma_t \sigma_{t+k}}$$

Autokovarianční i autokorelační funkce jsou funkce sudé, tj.  $\gamma_k = \gamma_{-k}$ ,  $\rho_k = \rho_{-k}$ , proto se určují pouze pro  $k \geq 0$ .

Odhady teoretických charakteristik (empirické charakteristiky) jsou:

- a) **aritmetický průměr**  $\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t,$
- b) **odhad rozptylu (variance)**  $S_Y^2 = \frac{1}{n} \sum_{t=1}^n (Y_t - \bar{Y})^2 = \frac{1}{n} \sum_{t=1}^n Y_t^2 - \bar{Y}^2,$
- c) **odhad autokovarianční funkce** ( $k = 0, 1, \dots$ )  
$$c_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y}) = \frac{1}{n-k} \sum_{t=1}^{n-k} Y_t Y_{t+k} - \bar{Y}^2,$$
- d) **odhad autokorelační funkce** ( $k = 0, 1, \dots$ )  
$$r_k = \frac{c_k}{S_Y^2} = \frac{c_k}{c_0}.$$

Výpočet odhadů (empirických charakteristik) má praktický význam pro  $n > 50, k \leq n/4$ .

Příklad:

```
data=read.csv("C:/Users/Tomas/Disk Google/vyuka/Stochastické Modely/Kurzy.csv",  
header = TRUE, sep = ",")
```

```
head(data)
```

```
CZK=data$Czech.koruna  
CZK=na.omit(CZK)
```

```
mean(CZK)  
sd(CZK)  
plot(CZK)  
acf(CZK)  
acf(CZK, type="covariance")
```

```
pacf(CZK)
```

```
ldCZK=diff(log(CZK), lag=1)
```

```
plot(ldCZK, type="l")  
mean(ldCZK)  
acf(ldCZK)
```

```
pacf(ldCZK)
```

```
hist(ldCZK, freq=FALSE)
```

```
curve(dnorm(x, mean=mean(ldCZK), sd=sd(ldCZK)), col="red", lwd=2, add=TRUE,  
yaxt="n")
```

## 2.1 Stacionarita časové řady

V Boxově – Jenkinsově metodologii lze modelovat pouze stacionární časové řady, resp. takové řady, které mohou být převedeny na stacionární. V teorii se rozlišuje dvojí stacionarita striktní a slabá. **Striktní stacionarita** předpokládá, že chování příslušného náhodného procesu, tj. jeho rozdělení, je invariantní vůči časovým posunům. Naproti tomu **slabá stacionarita** je méně omezující; požaduje se, aby příslušný náhodný proces měl konstantní střední hodnotu, konstantní rozptyl a pro kovariance platilo

$$\text{cov}(Y_t, Y_s) = \text{cov}(Y_{t+h}, Y_{s+h}),$$

a to pro libovolné  $h$ . Uvedený vztah představuje požadavek, aby závislost mezi dvěma libovolnými pozorováními závisela jen na jejich časové vzdálenosti a nikoli na jejich časovém umístění v řadě. V dalším výkladu budeme pod pojmem „stacionarita“ chápat slabou stacionaritu.

## 2.2 Autokorelační funkce (ACF)

Teoretická autokorelační funkce řádu  $k$  ( $\rho_k$ ), stejně jako její odhad  $r_k$ , byly již definovány v odstavci 1. Její hodnoty se označují stručně jako autokorelace řádu  $k$ . Připomínáme, že autokorelační funkce je funkce sudá, tj.  $\rho_{-k} = \rho_k$ , a proto se při práci s ní omezuje pouze na hodnoty řádu  $k \geq 0$ . Přitom vždy platí

$$\rho_0 = 1 \text{ a } |\rho_k| \leq 1 \text{ pro } k > 0.$$

Pro danou časovou řadu však teoretické autokorelace  $\rho_k$  neznáme, a proto je musíme odhadnout pomocí vypočtených hodnot  $r_k$ . Při rozhodování o nulové hypotéze  $\rho_k = 0$  porovnáváme vypočtenou hodnotu  $|r_k|$  s dvojnásobkem směrodatné odchylky  $r_k$ , tedy s hodnotou  $2\sigma(r_k)$ . Tato konkrétní volba hodnoty 2 pro multiplikativní konstantu odpovídá zhruba hladině významnosti testu rovné 0,05. Pro stanovení zmíněné směrodatné odchylky se používá následující aproximace.

**Bartlettova aproximace** [9]. Je-li  $\rho_k = 0$  pro  $k > k_0$ , pak platí

$$\sigma(r_k) \approx \sqrt{\frac{1}{n} \left( 1 + 2 \sum_{j=1}^{k_0} r_j^2 \right)}, \quad k > k_0,$$

přičemž  $n$  je délka analyzované časové řady.

### 2.3 Parciální autokorelační funkce (PACF)

Korelace mezi dvěma náhodnými veličinami bývá často způsobena tím, že obě náhodné veličiny jsou korelovány s jinými veličinami. Parciální autokorelace udávají korelaci mezi  $Y_t$  a  $Y_{t+k}$  očištěnou o vliv veličin ležících mezi nimi.

Teoretická **parciální autokorelační funkce řádu  $k$**  (**parciální autokorelace řádu  $k$** )  $\rho_{kk}$  je definována jako parciální korelační koeficient  $Y_t$  a  $Y_{t+k}$  při pevných hodnotách  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$ .

Odvození vztahů pro parciální autokorelace je poměrně jednoduché. Vyjdeme z toho, že parciální autokorelace  $\rho_{kk}$  představují parciální regresní koeficient v autoregresi  $k$ -tého řádu.

$$Y_t = \rho_{k1} Y_{t-1} + \rho_{k2} Y_{t-2} + \dots + \rho_{kk} Y_{t-k} + e_t.$$

Odhady parciálních autokorelací se počítají pomocí následujících rekurentních vztahů.

$$r_{11} = r_1,$$

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \text{ pro všechna } k > 1,$$

kde  $r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,j}$  pro  $j = 1, 2, \dots, k-1$ .

Také u parciální autokorelační funkce  $\rho_{kk}$  je pro výběr vhodného modelu důležité zjistit, zda má či nemá identifikační bod, a pokud jej má, určit jeho hodnotu  $k_0$ . Přitom se postupuje analogicky jako v případě autokorelační funkce  $\rho_k$ : vypočtené hodnoty  $|r_{kk}|$  se porovnávají s dvojnásobkem směrodatné odchylky odhadu  $r_{kk}$ , tj. s hodnotou  $2\sigma(r_{kk})$ . Ke stanovení zmíněné směrodatné odchylky slouží následující aproximace.

**Quenouilleova aproximace** (viz např. [9]). Je-li  $\rho_{kk} = 0$  pro  $k > k_0$ , potom platí

$$\sigma(r_{kk}) \approx \sqrt{\frac{1}{n}}, \quad k > k_0.$$



```
Příklad:  
auco=acf(ldczk) #sigma(r_k)^2 ~ 1/n  
auco1=acf(ldczk, ci.type = "ma") #Bartlett approximation  
plot(auco1)  
pacf(ldczk) #Quenouilleova aproximace sigma(r_k)^2 ~ 1/n
```

### 3. LINEÁRNÍ MODELY STACIONÁRNÍCH ČASOVÝCH ŘAD

#### 3.1 Pojem lineárního procesu

Lineární proces je definován jako nekonečná řada

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \quad (1)$$

kde jsou  $\psi_i$  nějaké parametry a  $\varepsilon_t$  vzájemně nekorelované složky reprezentují tzv. bílý šum s nulovou střední hodnotou, konstantním rozptylem  $\sigma^2$  a autokovarianční funkcí invariantní vůči posunutí v čase  $t$ .

Poznámka. Pro obecný stacionární lineární proces je třeba vztah (7.1) přepsat ve tvaru

$$Y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$$

kde  $\mu$  značí jeho střední hodnotu. V dalším výkladu budeme používat jednodušší reprezentaci lineárního procesu.

Stacionární lineární proces je tedy vyjádřen jako lineární kombinace řady nekorelovaných stejně rozdělených náhodných veličin. Taková reprezentace se někdy označuje jako **Woldova reprezentace**.

Při zápisu lineárního procesu se obvykle používá tzv. operátor zpětného posunutí (operátor zpoždění)  $B$ , pro nějž obecně platí

$$B^j Y_t = Y_{t-j} \text{ a samozřejmě také } B^j \varepsilon_t = \varepsilon_{t-j}$$

S použitím tohoto operátoru lze vztah (1) zapsat ve tvaru

$$y_t = \Psi(B) \varepsilon_t,$$

přičemž

$$\Psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots = 1 + \sum_{j=1}^{\infty} \psi_j B^j.$$

Lineární proces může existovat pouze tehdy, když nekonečná řada náhodných veličin na pravé straně vztahu (7.1) konverguje (předněji konverguje podle kvadratického středu). **Postačující podmínka pro existenci lineárního procesu** má tedy tvar: Lineární proces existuje tehdy, jestliže  $\Psi(B)$  konverguje pro  $|B| \leq 1$ , přitom  $B$  se považuje za obyčejnou číselnou proměnnou. Tato podmínka současně zaručuje i **stacionaritu lineárního procesu** a nulovost jeho střední hodnoty, tj.  $E(Y_t) = 0$ .

Z praktického hlediska je důležité, aby se současná hodnota  $Y_t$  lineárního procesu dala vyjádřit pomocí hodnot předcházejících  $Y_{t-1}, Y_{t-2}, \dots$  a současné hodnoty bílého šumu, tj. ve tvaru

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \dots + \varepsilon_t, \quad (7.2)$$

Použijeme-li operátor zpětného posunutí, můžeme vztah (7.2) přepsat ve tvaru

$$\pi(B)Y_t = \varepsilon_t,$$

kde

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots = 1 - \sum_{j=1}^{\infty} \pi_j B^j.$$

Lineární procesy, které lze takto vyjádřit, se nazývají **invertibilní**. **Postačující podmínku pro invertibilitu lineárního procesu** můžeme formulovat takto: lineární proces je invertibilní, jestliže  $\pi(B)$  konverguje pro  $|B| \leq 1$ .

Praktický význam v Boxově – Jenkinsově metodologii mají speciální lineární procesy s konečným (co nejmenším počtem) počtem nenulových parametrů. Přitom se parametry volí tak, aby výsledné modely vyhovovaly podmínkám pro stacionaritu a invertibilitu.

V původní Boxově – Jenkinsově metodologii se pracuje s třemi základními typy lineárních procesů:

- procesy klouzavých součtů –  $MA(q)$ ,
- autoregresní procesy –  $AR(p)$ ,
- smíšené procesy –  $ARMA(p, q)$ ,

proměnné  $p$  a  $q$  přitom udávají počet parametrů.

## 7.2 Proces klouzavých součtů

Proces klouzavých součtů řádu  $q$  (označení **MA( $q$ )**) je popsán modelem

$$Y_t = \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \dots + \vartheta_q \varepsilon_{t-q} = \mathcal{G}(B) \varepsilon_t,$$

kde

$$\mathcal{G}(B) = 1 + \sum_{j=1}^q \vartheta_j B^j$$

představuje **operátor klouzavých součtů**,  $\vartheta_1, \vartheta_2, \dots, \vartheta_q$  parametry procesu (reálná čísla) a  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$  složky bílého šumu, pro něž platí:

- $\mathbf{E}(\varepsilon_t) = 0$  pro všechna  $t$ ,
- $\text{var}(\varepsilon_t) = \mathbf{E}(\varepsilon_t^2) = \sigma^2 > 0$ ,
- $\text{cov}(\varepsilon_t, \varepsilon_s) = \mathbf{E}(\varepsilon_t \varepsilon_s) = 0$  pro  $t \neq s$ .

1. Proces MA( $q$ ) je stacionární pro libovolné hodnoty parametrů.
2. Střední hodnota procesu MA( $q$ ) je nulová, tj.  $\mathbf{E}Y_t = 0$ .
3. Pro varianci (rozptyl) procesu MA( $q$ ) platí

$$\text{var}(Y_t) = \sigma^2 \left( 1 + \sum_{j=1}^q \vartheta_j^2 \right).$$

4. Autokorelační funkce  $\rho_k$  procesu MA( $q$ ) má tvar

$$\rho_k = \frac{\mathcal{G}_k + \mathcal{G}_1 \mathcal{G}_{k+1} + \dots + \mathcal{G}_{q-k} \mathcal{G}_q}{1 + \sum_{j=1}^q \mathcal{G}_j^2}.$$

Tato funkce má identifikační bod  $k_0 = q$ .

5. Parciální autokorelační funkce  $\rho_{kk}$  procesu MA( $q$ ) nemá identifikační bod, je omezena geometricky klesající posloupností nebo sinusoidou s geometricky klesající amplitudou.
6. Proces MA( $q$ ) je invertibilní, jestliže všechny nulové body polynomu  $\mathcal{G}(B)$  leží vně jednotkového kruhu v komplexní rovině.
7. Při praktických aplikacích se nejčastěji využívají procesy MA(1) a MA(2).

```
#####MA(1)#####  
library(forecast)  
r=arima.sim(model = list(ma = c(0.8897)), sd = sqrt(0.1796),n=500)  
plot(r)  
acf(r)  
pacf(r)  
fit=Arima(r,order=c(0,0,1))  
plot(residuals(fit))  
pacf(residuals(fit))  
acf(residuals(fit))  
  
forecast(fit,h=20)  
plot(forecast(fit,h=20))  
summary(fit)  
  
auto.arima(r,ic="bic")
```

```
#####Test
2*(1-pnorm(abs(fit$coef[1]),mean=0, sd = sqrt(fit$var.coef[1,1])))

#tseries
library(tseries)
fit1=arma(r,order=c(0,1))
summary(fit1)
plot(fit1$residuals)
```



### 7.3 Autoregresní proces

**Autoregresní proces řádu  $p$**  (označení **AR( $p$ )**) je popsán modelem

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t \quad (7.3)$$

neboli (s použitím operátoru zpětného posunutí)

$$\varphi(B)Y_t = \varepsilon_t,$$

kde

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

je tzv. **autoregresní operátor**.

- 1) Autoregresní proces AR( $p$ ) je stacionární, jestliže všechny nulové body polynomu  $\varphi(B)$  leží vně jednotkového kruhu v komplexní rovině.
- 2) Střední hodnota autoregresního procesu AR( $p$ ) je nulová, tj.  $\mathbf{E}(Y_t) = 0$ .
- 3) Pro rozptyl (varianci) autoregresního procesu AR( $p$ ) platí

$$\text{var}(Y_t) = \frac{\sigma^2}{1 - \varphi_1 \rho_1 - \varphi_2 \rho_2 - \dots - \varphi_p \rho_p}.$$

- 4) Autokorelační funkce  $\rho_k$  autoregresního procesu AR( $p$ ) nemá identifikační bod a splňuje soustavu diferenčních rovnic

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \quad \text{pro } k > 0.$$

- 5) Parciální autokorelační funkce  $\rho_{kk}$  procesu AR( $p$ ) má identifikační bod  $k_0 = p$ , což znamená, že platí  $\rho_{kk} = 0$  pro  $k > p$ .

6) Proces  $AR(p)$  je invertibilní pro libovolné hodnoty parametrů.

7) V praxi se nejčastěji uplatňují procesy  $AR(1)$  a  $AR(2)$ .

```
#####AR(1)#####  
library(forecast)  
r=arima.sim(model = list(ar = c(0.8897)), sd = sqrt(0.1796),n=100)  
plot(r)  
acf(r)  
pacf(r)  
fit=Arima(r,order=c(1,0,0))  
plot(residuals(fit))  
pacf(residuals(fit))  
forecast(fit,h=20)  
plot(forecast(fit,h=20))  
summary(fit)  
  
auto.arima(r,ic="bic")  
  
#####Test  
2*(1-pnorm(abs(fit$coef[1]),mean=0, sd = sqrt(fit$var.coef[1,1])))  
  
#tseries  
library(tseries)  
fit1=arma(r,order=c(1,0))  
summary(fit1)  
plot(fit1$residuals)
```

## 7.4 Smíšený proces

Smíšený proces řádu  $p$  a  $q$  s označením  $\text{ARMA}(p, q)$  se definuje modelem

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \vartheta_2 \varepsilon_{t-2} + \dots + \vartheta_q \varepsilon_{t-q}$$

nebo ekvivalentně s použitím operátoru zpětného posunutí

$$\varphi(B)Y_t = \vartheta(B)\varepsilon_t,$$

kde  $\varphi(B)$  je autoregresní operátor a  $\vartheta(B)$  operátor klouzavých součtů.

1. Smíšený proces  $\text{ARMA}(p, q)$  je stacionární, když je stacionární proces  $\text{AR}(p)$ .
2. Střední hodnota stacionárního smíšeného procesu  $\text{ARMA}(p, q)$  je nulová.
3. Autokorelační funkce  $\rho_k$  smíšeného procesu  $\text{ARMA}(p, q)$  vyhovuje vztah; m pro  $\text{AR}(p)$  pro  $k > q$ . Nemá identifikační bod a představuje (po prvních  $q - p$  hodnotách) lineární kombinaci klesajících geometrických posloupností a sinusoid různých frekvencí s geometricky klesajícími amplitudami.
4. Parciální autokorelační funkce  $\rho_{kk}$  smíšeného procesu  $\text{ARMA}(p, q)$  nemá rovněž identifikační bod a je omezena (po prvních  $p - q$  hodnotách) geometricky klesající posloupností nebo sinusoidou s geometricky klesající amplitudou.
5. Smíšený proces  $\text{ARMA}(p, q)$  je invertibilní, jestliže je invertibilní proces  $\text{MA}(q)$ .
6. V praxi se nejčastěji setkáváme s procesem  $\text{ARMA}(1, 1)$ .

```
#####gold white noise after transformation
data=read.csv("C:/Users/Tomas/Disk Google/vyuka/Stochasticke
Modely/KurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
```

```
plot(data$X1gold.USD,type="l")
plot(log(na.omit(data$X1gold.USD)))
ldgold=diff(log(na.omit(data$X1gold.USD)),1)
plot(ldgold, type="l")
acf(ldgold)
pacf(ldgold)
auto.arima(ldgold,ic="bic")
```

```
#####Euro-Brazilian.real MA(2)
data=read.csv("C:/Users/Tomas/Disk Google/vyuka/Stochasticke
Modely/KurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
library(forecast)
```

```
plot(data$X1EUR.Brazilian.real,type="l")
plot(log(na.omit(data$X1EUR.Brazilian.real)))
ldEu_Br=diff(log(na.omit(data$X1EUR.Brazilian.real)),1)
plot(ldEu_Br, type="l")
acf(ldEu_Br)
pacf(ldEu_Br)
auto.arima(ldEu_Br,ic="bic")
```

```
fit=Arima(ldEu_Br,order=c(0,0,2))
plot(residuals(fit))
acf(residuals(fit))
pacf(residuals(fit))
```

```
forecast(fit,h=20)
plot(forecast(fit,h=20))
```

```
summary(fit)
```

```
#####Test  
2*(1-pnorm(abs(fit$coef[1]),mean=0, sd = sqrt(fit$var.coef[1,1])))  
2*(1-pnorm(abs(fit$coef[2]),mean=0, sd = sqrt(fit$var.coef[2,2])))  
#tseries  
library(tseries)  
fit1=arma(1dEu_Br,order=c(0,2))  
summary(fit1)  
plot(fit1$residuals)
```

## 4. LINEÁRNÍ MODELY NESTACIONÁRNÍCH ČASOVÝCH ŘAD

V předcházející kapitole jsme se zabývali výhradně stacionárními procesy. V ekonomické praxi se však často setkáváme s časovými řadami tvořenými nestacionárními procesy, které jsou charakterizovány přítomností výrazného trendu.

### 4.1 Proces náhodné procházky

Proces náhodné procházky (proces RW) se popisuje modelem

$$Y_t = Y_{t-1} + \varepsilon_t.$$

Jde zřejmě o speciální případ modelu AR(1). Pomocí operátoru zpětného posunutí jej můžeme zapsat jako

$$(1 - B)Y_t = \varepsilon_t.$$

Vzhledem k tomu, že platí

$$Y_t = (Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = (Y_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1} + \varepsilon_t =$$

lze model (8.1) přepsat do tvaru

$$Y_t = Y_0 + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \dots$$

kde  $Y_0$  představuje hodnotu odpovídajícího procesu v čase  $t = 0$ . Z toho ovšem vyplývá, že proces náhodné procházky je tvořen nekonečným součtem náhodných veličin, jež jsou složkami bílého šumu. První diference tohoto procesu ( $Y_t - Y_{t-1}$ ) zřejmě představuje proces bílého šumu. Proces s takovou vlastností se nazývá integrovaný proces 1. řádu a označuje symbolicky I(1).

Má-li proces RW v čase  $t = 0$  hodnotu  $Y_0$ , pak jeho střední hodnota  $\mathbf{E}(Y_t) = Y_0 = \mu$  je konstantní, nezávislá na čase.

Pro rozptyl procesu RW platí

$$\text{var}(Y_t) = \gamma_0 = \mathbf{E}\left[(\varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1)(\varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1)\right] = t\sigma^2, \quad (8.2)$$

což znamená, že je lineární funkcí času a pro  $t \rightarrow \infty$  roste neomezeně.

Autokovarianční funkce řádu  $k$  je dána vztahem (viz např. [10])

$$\begin{aligned} \gamma_k &= \mathbf{E}\left[(Y_t - Y_0)(Y_{t+k} - Y_0)\right] = \mathbf{E}\left[(\varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1)(\varepsilon_{t-k} + \varepsilon_{t-k-1} + \dots + \varepsilon_1)\right] \\ &= (t-k)\sigma^2, \end{aligned}$$

takže závisí nejen na posunutí  $k$ , ale i na čase  $t$  a s rostoucím časem neomezeně roste.

Ze vztahu (8.2) vyplývá, že rozptyl v čase  $(t-k)$  je roven  $\text{var}(Y_{t-k}) = (t-k)\sigma^2$ , pak autokorelační funkce

$$\rho_k = (t-k) / \sqrt{t(t-k)} = \sqrt{(t-k)/t}$$

také závisí na čase a pro  $t \rightarrow \infty$  při dané hodnotě zpoždění ( $k$ ) konverguje k jedné.

Z uvedených vlastností procesu RW vyplývá, že náhodná procházka je nestacionární proces, přičemž zdrojem nestacionarity je stochastický trend.

## 4.2 Smíšené integrované procesy

Smíšené integrované procesy (procesy ARIMA) jsou určeny pro popis časových řad s náhodnými změnami trendu (úrovně a sklonu). Výchozí časová řada nemusí být stacionární, je však nutné, aby byla převoditelná na stacionární. Tento převod se uskutečňuje diferencováním výchozí časové řady.

**Smíšený integrovaný proces ARIMA( $p, d, q$ )** se popisuje modelem

$$\varphi(B)W_t = \vartheta(B)\varepsilon_t,$$

v němž

$$W_t = \Delta^d Y_t$$

$W_t$  reprezentuje časovou řadu zkonstruovanou diferenciací výchozí řady

$Y_t$ ,  $d$  je **řád diferencování** a  $\Delta$  **diferenční operátor** definovaný jako

$$\Delta^d Y_t = (1-B)^d Y_t.$$

Model ARIMA( $p, d, q$ ) je možno také zapsat ve tvaru

$$\varphi(B)(1-B)^d Y_t = \vartheta(B)\varepsilon_t. \quad (8.3)$$



Proč nazýváme proces integrovaný?

Pro jednoduchost budeme předpokládat, že  $d=1$ , a tedy  $W_t = \Delta Y_t$ .

Opakovaným použitím vztahu  $W_t = \Delta Y_t$  dostaneme

$$Y_t = W_t + Y_{t-1} = W_t + W_{t-1} + Y_{t-2} = W_t + W_{t-1} + \dots + W_{t-k} + Y_{t-k-1}.$$

Můžeme si představit, že vliv členu  $Y_{t-k-1}$  s rostoucí hodnotou  $k$  postupně slábne, takže původní proces  $Y_t$  dostaneme z procesu  $W_t$  postupným sčítáním (integrovaním). Proto se proces  $ARIMA(p,1,q)$  nazývá integrovaným procesem (přesněji integrovaným procesem 1. řádu), tj.  $ARIMA(p,1,q) \sim I(1)$ . Obecně platí  $ARIMA(p,d,q) \sim I(d)$ .

Jak stanovit hodnotu  $d$ ?

1. Vizuální posouzení stacionarity výchozí řady  $Y_t$  a diferencovaných řad.
2. Studium odhadů autokorelační funkce  $r_k$  pro řady  $Y_t$  a  $\Delta^d Y_t$ ,  $t = 1, 2, \dots$ .  
Jestliže hodnoty  $r_k$  klesají pomalu (přibližně lineárně), pak je nutno provést další diferenciaci.
3. Studium odhadů rozptylu pro řady  $Y_t$  a  $\Delta^d Y_t$ ,  $t = 1, 2, \dots$ . Za optimální se volí taková hodnota parametru  $d$ , která poskytuje nejmenší odhad rozptylu.

```
library(forecast)
setwd("H:/StochMod")
data=read.csv("KurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
plot(data$X1gold.USD,type="l")
plot(log(na.omit(data$X1gold.USD)),type="l")
ldgold=diff(log(na.omit(data$X1gold.USD)),1)
plot(ldgold, type="l")
acf(ldgold)
pacf(ldgold)
auto.arima(ldgold,ic="bic")

lgold=log(na.omit(data$X1gold.USD))
acf(lgold)
acf(diff(lgold,differences = 1))
plot(diff(lgold,differences = 1),type="l")
plot(diff(lgold,differences = 2),type="l")

var(lgold)
var(diff(lgold,differences = 1))
var(diff(lgold,differences = 2))

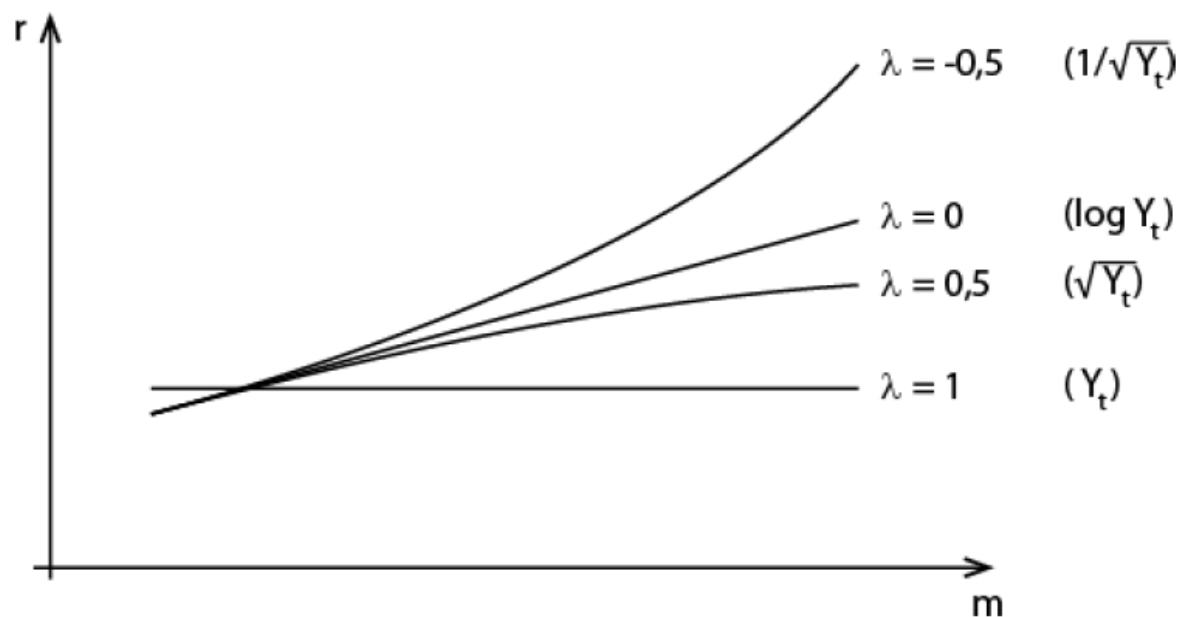
auto.arima(lgold)

acf(ldgold)
auto.arima(ldgold)
```

### Boxova Jenkinsonova transformace pro linearizaci časových řad.

$$Y_t^{(\lambda)} = \begin{cases} Y_t^\lambda & \text{pro } \lambda \neq 0, \\ \log Y_t & \text{pro } \lambda = 0. \end{cases}$$

Optimální hodnota parametru  $\lambda$  se určuje graficky (viz např.[8]). Daná časová řada se rozdělí na krátké úseky (4 až 12 pozorování) a v každém z těchto úseků se určí aritmetický průměr hodnot pozorování  $m$  a rozdíl mezi maximální a minimální hodnotou  $r$ . Sestrojí se graf závislosti  $r$  na  $m$  (viz obr. 8.1).



```
data=read.csv("C:/Users/Tomas/Disk Google/vyuka/Stochasticke
Modely/KurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
library(forecast)
gold=na.omit(data$X1gold.USD)
plot(gold,type="l")
#k - vzdalenost
k=100
m=NULL
r=NULL
for(i in 1:(length(gold)-k)){
  m=c(m,mean(gold[i:(i+k-1)]))
  r=c(r,max(gold[i:(i+k-1)])-min(gold[i:(i+k-1)]))
}
plot(m,r,ylim = c(0,300))
#d=1
ldgold=diff(log(na.omit(data$X1gold.USD)),1)
plot(ldgold, type="l")
acf(ldgold)
pacf(ldgold)
auto.arima(ldgold,ic="bic")
```

### 4.3 Modely volatility

```
#####Simulace
e=rnorm(100,0,1)
Y=array(0,100)
for(i in 1:100){
  Y[i]=sum(e[1:i])
}
plot(Y,type="l",ylim = c(-20,20))

for(j in 1:10){e=rnorm(100,0,1);
Y=array(0,100);
for(i in 1:100){
  Y[i]=sum(e[1:i])
};
}lines(x=1:100,y=Y)
}
```

```
#####Volatilita
plot(lsgold,type="l")
k=200
v=NULL
for(i in 1:(length(gold)-k)){
  v=c(v,sd(gold[i:(i+k-1)]))
}
plot(1:length(v),v,type = "l")
```

Standardně používané modely AR nebo ARMA poskytují špatné výsledky při analýze finančních časových řad. Takové řady vykazují různé nelinearity, zejména silnou závislost okamžitého rozptylu řady na její historii. Ukázalo se, že vhodným nástrojem pro odstranění těchto problémů jsou modely volatility. Při jejich popisu se preferují takové charakteristiky, jakými jsou podmíněná střední hodnota a podmíněný rozptyl.

V případě jednoduchého autoregresního modelu AR(1), popsaného vztahem  $Y_t = \phi_1 Y_{t-1} + \varepsilon_t$ , platí pro podmíněné charakteristiky následující vztahy:

$$\mathbf{E}(Y_t | Y_{t-1}) = \varphi_1 Y_{t-1} \text{ a } \text{var}(Y_t | Y_{t-1}) = \sigma^2.$$

Z uvedeného vyplývá, že podmíněná střední hodnoty náhodné veličiny  $Y_t$  se v čase mění, ale její podmíněný rozptyl zůstává konstantní.

Označme veškerou informaci známou do času  $t-1$  symbolem  $\Omega_{t-1}$ . Tato „minulá“ informace je generována všemi minulými hodnotami časové řady  $\{Y_{t-1}, Y_{t-2}, \dots\}$ , všemi minulými hodnotami  $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  a vhodnými funkcemi těchto hodnot. V modelech volatility se pracuje s podmíněnými charakteristikami ve tvaru vhodných nelineárních funkcí informace  $\Omega_{t-1}$ :

$$\begin{aligned} \mu_t &= \mathbf{E}(Y_t | \Omega_{t-1}) = g(\Omega_{t-1}), \\ h_t &= \text{var}(Y_t | \Omega_{t-1}) = h(\Omega_{t-1}), \quad h(\cdot) > 0. \end{aligned}$$

Podmíněný rozptyl se označuje jako **volatilita časové řady** a funkce  $h(\cdot)$  jako **skedastická funkce**. V případě modelu AR(1) je  $h(\cdot)$  funkce **homoskedastická**, při proměnlivém podmíněném rozptylu funkce **heteroskedastická**.

Modely volatility se nejčastěji zapisují ve tvaru

$$Y_t = \mu_t + e_t = \mu_t + \varepsilon_t \sqrt{h_t} = g(\Omega_{t-1}) + \varepsilon_t \sqrt{h(\Omega_{t-1})},$$

Kde  $e_t$  jsou nekorelované identicky rozdělené náhodné veličiny s podmíněnými charakteristikami  $\mathbf{E}(e_t | \Omega_{t-1}) = 0$  a  $\text{var}(e_t | \Omega_{t-1}) = h_t$  a  $\varepsilon_t$  nezávislé identicky rozdělené veličiny s nulovou střední hodnotou a jednotkovým rozptylem. Uvažované modely jsou určeny dvěma rovnicemi (8.4), z nichž první se obvykle nazývá rovnice střední hodnoty a druhá rovnice volatility.

### 4.3.1 Modely ARCH

Prvním vhodným nástrojem pro modelování volatility byl model ARCH (Autoregressive Conditionally Heteroscedastic Model), který vytvořil a aplikoval Engle v roce 1982.

Obecný model ARCH( $m$ ) můžeme zapsat ve tvaru

$$Y_t = \mu_t + \varepsilon_t \sqrt{h_t}, \quad h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_m e_{t-m}^2.$$

$$\varepsilon_t \sqrt{h_t} = \mathbf{e}_t \quad \text{a} \quad \varepsilon_t \text{ je bílý šum. Obvykle } N(0,1) \text{ rozdělení.}$$

Koeficienty  $\alpha_0, \alpha_1, \dots, \alpha_m$  mohou nabývat jen takových hodnot, aby platilo  $h_t > 0$ . Postačující podmínka, zaručující kladnost volatility, má tvar  $\alpha_0 > 0, \alpha_1 \geq 0, \dots, \alpha_m \geq 0$ .

ARCH(1)

1. Nepodmíněná střední hodnota veličin  $e_t$  je nulová, tj.  $\mathbf{E}(e_t) = 0$ .

2. Pro nepodmíněný rozptyl veličin  $e_t$  platí

$$\text{var}(e_t) = \frac{\alpha_0}{1 - \alpha_1} \text{ za podmínky } 0 < \alpha_1 < 1. \text{ To znamená, že je konstantní}$$

v čase a proces  $\{e_t\}$  je nepodmíněně homoskedastický.

3. Pro nepodmíněný koeficient špičatosti veličin  $e_t$  platí  $\frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2} \geq 3$  za

podmínky  $0 \leq \alpha_1 < \sqrt{1/3}$ , což znamená, že může být větší než

u normálního rozdělení.

```
library(rugarch)
data=read.csv("C:/Users/Tomas/Disk Google/vyuka/Stochastické Modely/
KurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
library(forecast)
Eur=na.omit(data$X1EUR.USD)
```

```

plot(Eur,type="l")
ldEur=diff(log(na.omit(Eur)),1)
plot(ldEur, type="l")
acf(ldEur)
pacf(ldEur)
auto.arima(ldEur,ic="bic")
library(zoo)
plot(rollapply(ldEur, width = 150, FUN = sd))

#ARCH(1)
spec = ugarchspec(mean.model = list(armaOrder = c(0, 0)),variance.model =
list(model = "SGARCH", garchOrder = c(1,0)))
spec
fit = ugarchfit(data=ldEur, spec = spec)
summary(fit)
fit
ugarchforecast(fit,n.ahead = 10)
plot(residuals(fit))
plot(rollapply(residuals(fit), width = 150, FUN = sd, na.pad = TRUE))
plot(sigma(fit))
Standardisedresiduals=residuals(fit)/sigma(fit)
plot(rollapply(Standardisedresiduals, width = 150, FUN = sd))

```

### 4.3.1 Modely GARCH

Nevýhodou modelů ARCH je skutečnost, že často vyžadují ke správnému popsání volatility vysoký řád  $m$  a s tím souvisí nutnost odhadovat velký počet parametrů. Tuto nevýhodu odstraňují modely GARCH (Generalized ARCH). V těchto modelech může volatilita (tj. podmíněný rozptyl)



záviset navíc na minulých hodnotách podmíněného rozptylu.

Obecný model GARCH( $m, s$ ) má tvar

$$Y_t = \mu_t + \varepsilon_t \sqrt{h_t}, \quad h_t = \alpha_0 + \sum_{i=1}^m \alpha_i e_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j},$$

kde  $\varepsilon_t$  jsou nezávislé identicky rozdělené veličiny s nulovou střední hodnotou, jednotkovým rozptylem (často normálním rozdělením) a  $\alpha_i, \beta_j$  ( $i = 0, 1, \dots, m, j = 1, 2, \dots, s$ ) jsou parametry modelu, které splňují

podmínky:  $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^{\max\{m,s\}} (\alpha_i + \beta_i) < 1$ .

Nejjednodušším zástupcem modelů GARCH je model GARCH(1,1), který je možno zapsat ve tvaru

$$Y_t = \mu_t + \varepsilon_t \sqrt{h_t}, \quad h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 h_{t-1}, \quad \alpha_0 > 0, \alpha_1 > 0, \beta_1 \geq 0.$$

Základní vlastnosti **modelu GARCH(1,1)** jsou podobné vlastnostem modelu ARCH(1):

Model GARCH(1,1) je slabě stacionární, jestliže  $\alpha_1 + \beta_1 < 1$ .

2. Pro nepodmíněný rozptyl veličin  $e_t$  platí

$$\text{var}(e_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \text{ za podmínky } \alpha_1 + \beta_1 < 1.$$

Rozptyl je tedy konstantní v čase a proces  $\{e_t\}$  je nepodmíněně homoskedastický.

3. Pro nepodmíněný koeficient špičatosti veličin  $e_t$  platí

$$\frac{3(1 - (\alpha_1 + \beta_1)^2)}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} \geq 3 \text{ za podmínky } 1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0.$$

```
library(rugarch)
data=read.csv("C:/Users/Tomas/Disk Google/vyuka/Stochasticke Modely/
KurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
library(forecast)
Eur=na.omit(data$X1EUR.USD)
plot(Eur,type="l")
ldEur=diff(log(na.omit(Eur)),1)
plot(ldEur, type="l")
acf(ldEur)
pacf(ldEur)
auto.arima(ldEur,ic="bic")
library(zoo)
plot(rollapply(ldEur, width = 150, FUN = sd, na.pad = TRUE))

#GARCH(1,1)
```

```

spec = ugarchspec(mean.model = list(armaOrder = c(0, 0)),variance.model =
list(model = "SGARCH", garchOrder = c(1,1)))
spec
fit = ugarchfit(data=ldEur, spec = spec)
summary(fit)
fit
ugarchforecast(fit,n.ahead = 10)
sim = ugarchsim(fit,n.sim=500, n.start=1, m.sim=1, startMethod="sample")
plot(sim)
plot(residuals(fit))

plot(rollapply(residuals(fit), width = 150, FUN = sd))
plot(sigma(fit))
Standardisedresiduals=residuals(fit)/sigma(fit)
plot(rollapply(Standardisedresiduals, width = 150, FUN = sd))

```

```
#####ARMA + GARCH
```

```

library(rugarch)
data=read.csv("C:/Users/Tomas/Disk Google/vyuka/Stochasticke Modely/
KurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
library(forecast)
library(zoo)
plot(data$X1EUR.Brazilian.real,type="l")
plot(log(na.omit(data$X1EUR.Brazilian.real)))
ldEu_Br=diff(log(na.omit(data$X1EUR.Brazilian.real)),1)
plot(ldEu_Br, type="l")
acf(ldEu_Br)
pacf(ldEu_Br)
auto.arima(ldEu_Br,ic="bic")

fit=Arima(ldEu_Br,order=c(0,0,2))
plot(residuals(fit))
acf(residuals(fit))
pacf(residuals(fit))

```

```

plot(rollapply(residuals(fit), width = 150, FUN = sd))

#####ARMA(0,2) + GARCH(1,1)
spec = ugarchspec(mean.model = list(armaOrder = c(0, 2)),variance.model =
list(model = "SGARCH", garchOrder = c(1,1)))
spec
fit = ugarchfit(data=ldEu_Br, spec = spec)
summary(fit)
fit
ugarchforecast(fit,n.ahead = 10)
sim = ugarchsim(fit,n.sim=500, n.start=1, m.sim=1, startMethod="sample")
plot(sim)
plot(residuals(fit))

plot(rollapply(residuals(fit), width = 150, FUN = sd))
plot(sigma(fit))
Standardisedresiduals=residuals(fit)/sigma(fit)
plot(rollapply(Standardisedresiduals, width = 150, FUN = sd))

```

## 5 Hodnota v riziku

Hodnota v riziku je hodnota číselné řady, kterou řada nepřekročí s pravděpodobností  $\alpha$ .

Nejčastěji používané jsou hodnoty v riziku s pravděpodobností 0,05 a 0,95.

Hodnota v riziku mi určuje míru rizika.

```
#####ARMA + GARCH
library(rugarch)
setwd("H:/StochMod")
setwd("C:/Users/Tomas/Disk Google/vyuka/Stochasticke Modely")
data=read.csv("kurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
library(forecast)
library(zoo)
plot(data$X1EUR.Brazilian.real,type="l")
plot(log(na.omit(data$X1EUR.Brazilian.real)))
ldEu_Br=diff(log(na.omit(data$X1EUR.Brazilian.real)),1)
plot(ldEu_Br, type="l")
#####ARMA(0,2) + GARCH(1,1)
spec = ugarchspec(mean.model = list(armaOrder = c(0, 2)),variance.model =
                    list(model = "sGARCH", garchOrder = c(1,1)))

spec
fit = ugarchfit(data=ldEu_Br, spec = spec)
fit

#####VaR
pred=ugarchforecast(fit,n.ahead = 10)
pred
#Predikce je:
mu= 2.706e-04
sig=0.01200
curve(dnorm(x, mean=mu, sd=sig),
      col="red", lwd=2, add=FALSE, yaxt="n",xlim = c(-0.2,0.2))

VaR95=qnorm(0.95, mean=mu, sd = sig)
VaR05=qnorm(0.05, mean=mu, sd = sig)
#zpatky pretransformovana hodnota v riziku
```

```

SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]
exp(log(SoucHod)+VaR95)
exp(log(SoucHod)+VaR05)
#Normalita
Standardisedresiduals=residuals(fit)/sigma(fit)
plot(rollapply(Standardisedresiduals, width = 150, FUN = sd))
hist(Standardisedresiduals,freq=FALSE,breaks = 100)
curve(dnorm(x, mean=mu, sd=1),
      col="red", lwd=2, add=TRUE, yaxt="n")
s=as.vector(Standardisedresiduals[1:3000])
shapiro.test(s)
qqnorm(Standardisedresiduals)
qqline(Standardisedresiduals)
#indikuje heavy tails
#parametricky

qqplot(rt(300, df = 5,ncp=mu),as.vector(Standardisedresiduals) )

pt5=function(q){pt(q,df=5,ncp=mu)}
hist(Standardisedresiduals,freq=FALSE,breaks = 100)
curve(dt(x, df=5,ncp=mu),
      col="red", lwd=2, add=TRUE, yaxt="n")
curve(dnorm(x, mean=mu, sd=1),
      col="blue", lwd=2, add=TRUE, yaxt="n")

ks.test(s, "pt5")

VaR95=qt(0.95, df=5, ncp=mu)*sig
VaR05=qt(0.05, df=5, ncp=mu)*sig
#zpatky pretransformovana hodnota v riziku
SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]
exp(log(SoucHod)+VaR95)
exp(log(SoucHod)+VaR05)

#neparametricky

```

```

VaR95=quantile(Standardisedresiduals, 0.95)*sig
VaR05=quantile(Standardisedresiduals, 0.05)*sig
#zpatky pretransformovana hodnota v riziku
SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]
exp(log(SoucHod)+VaR95)
exp(log(SoucHod)+VaR05)

```

Místo normality je možné použít jiná rozdělení, např. t rozdělení, které je heavy tailed, nebo neparametrické rozdělení.

Aplikace

<https://var.ef.jcu.cz/>

umožňuje predikovat riziko ve formě Hodnoty v riziku pro libovolné měny s aktuálními stavu kurzů. Navíc pro dlouhodobé transakce.

### Expected shortfall – Podmíněná hodnota v riziku

Očekávaná hodnota ztráty portfolia X pro nejhorších  $\alpha$  případů. Expected shortfall je míra nejistoty, riziku a je méně sensitivní na okrajové hodnoty rozdělení X než hodnota v riziku.

$$ES_{\alpha}(X) = -\frac{1}{\alpha} \int_0^{\alpha} VaR_{\gamma}(X) d\gamma$$

Pro malé  $\alpha$ , měří nejhorší ztráty

Pro velké  $\alpha$ , měří ztráty, přičemž ignoruje nejpozitivnější, nepravděpodobné scénáře.

```

#####Expected Shortfall
gamma=seq(0,0.05, by=0.001)
gamma=gamma[-1]
VaRgamma=qnorm(gamma, mean=mu, sd = sig)
plot(VaRgamma)
ES=-mean(VaRgamma)
#zpatky pretransformovana hodnota v riziku
SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]
exp(log(SoucHod)-ES)

```

```

gamma=seq(0,0.95, by=0.001)
gamma=gamma[-1]
VaRgamma=qnorm(gamma, mean=mu, sd = sig)
plot(VaRgamma)
ES=-mean(VaRgamma)
#zpatky pretransformovana hodnota v riziku
SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]
#konzervativni odhad stredni hodnoty
exp(log(SoucHod)-ES)

#heavy tails expected shortfall parametricky
gamma=seq(0,0.05, by=0.001)
gamma=gamma[-1]
VaRgamma=qt(gamma, df=5, ncp=mu)*sig
plot(VaRgamma)
ES=-mean(VaRgamma)
#zpatky pretransformovana hodnota v riziku
SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]
exp(log(SoucHod)-ES)
gamma=seq(0,0.95, by=0.001)
gamma=gamma[-1]
VaRgamma=qt(gamma, df=14, ncp=mu)*sig
plot(VaRgamma)
ES=-mean(VaRgamma)
#zpatky pretransformovana hodnota v riziku
SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]
#konzervativni odhad stredni hodnoty
exp(log(SoucHod)-ES)

#heavy tails expected shortfall neparametricky
gamma=seq(0,0.05, by=0.001)
gamma=gamma[-1]
VaRgamma=quantile(Standardisedresiduals, gamma)*sig
plot(VaRgamma)
ES=-mean(VaRgamma)
#zpatky pretransformovana hodnota v riziku
SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]

```



```
exp(log(SoucHod)-ES)
gamma=seq(0,0.95, by=0.001)
gamma=gamma[-1]
VaRgamma=quantile(Standardisedresiduals, gamma)*sig
plot(VaRgamma)
ES=-mean(VaRgamma)
#zpatky pretransformovana hodnota v riziku
SoucHod=data$X1EUR.Brazilian.real[length(data$X1EUR.Brazilian.real)]
#konzervativni odhad stredni hodnoty
exp(log(SoucHod)-ES)
```

## 6 Regresní modely s autokorelační strukturou

Testování nezávislosti residuí:

### Durbin Watson Test

**H<sub>0</sub> (null hypothesis):** There is no correlation among the residuals at lag 1.

**H<sub>A</sub> (alternative hypothesis):** The residuals are autocorrelated.

Je založen na statistice

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2},$$

```
setwd("C:/Users/Tomas/Disk Google/vyuka/Stochasticke Modely")
library(readxl)
Covid <- read_excel("Covid.xlsx")
View(Covid)
library(forecast)
plot(Covid$Unknowndeaths~Covid$Coviddeaths)
ccc=lm(Covid$Unknowndeaths~Covid$Coviddeaths)
summary(ccc)
```

Předpoklad: residua jsou iid – nezávislá, stejně rozdělená, a navíc normální

```
#Testing the autocorrelation structure
library(car)
durbinwatsonTest(ccc)
```

## Ljung-Box test

The Ljung–Box test may be defined as:

**H<sub>0</sub>:** The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).

**H<sub>a</sub>:** The data are not independently distributed; they exhibit serial correlation.

The test statistic is:

$$Q = n(n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n - k}$$

where  $n$  is the sample size,  $\hat{\rho}_k$   $\{\displaystyle \{\hat{\rho}\}_{k}\}$  is the sample autocorrelation at lag  $k$ , and  $h$  is the number of lags being tested.

```
#Ljung - Box test
Box.test(residuals(ccc), lag = 3, type = c("Box-Pierce", "Ljung-Box"), fitdf = 2)
#lag the statistic will be based on lag autocorrelation coefficients.
#fitdf number of degrees of freedom to be subtracted if x is a series of
residuals.
# lag > fitdf.
#For ARMA(p,q) residuals fitdf = p+q.
```

## ARMA methodology

```
plot(residuals(ccc))
acf(residuals(ccc))
pacf(residuals(ccc))
auto.arima(residuals(ccc))
fit=Arima(residuals(ccc),order=c(1,0,0))
summary(fit)
2*(1-pnorm(abs(fit$coef[1]),mean=0, sd = sqrt(fit$var.coef[1,1])))
```

Rezidua zjevně obsahují autocorelaci, takže nemůžeme věřit výsledkům testu v lineárním modelu. Předpoklad nezávislosti je nejdůležitější ze všech předpokladů lineárního modelu. Důležitější než homoscedasticita a mnohem důležitější než normalita.

Zkušenosti statistiků říkají, že je důležité autocorelaci nějak modelovat, ale není zásadní mít úplně perfektní model.

Lineární model:  $Y = X\beta + \varepsilon$ ,  $\varepsilon \sim N(0, V)$

Nezávislá data:  $V = \sigma^2 I$ ,  $I$  je identická matice

Změníme  $V$ , tak aby obsahovalo korelační strukturu a použijeme zobecněné lineární modely, kde může být libovolná struktura.

Modely autokorelace:

1) Compound symmetry

$$\text{Cor}(\varepsilon) = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}$$

```
library(nlme)
```

```
#Compound symmetry
```

```
ccc1=glm(Unknowndeaths~Coviddeaths,data=Covid,correlation =corCompSymm(form = ~  
Number))
```

```
summary(ccc1)
```

2) AR(1)

$$\text{Cor}(\varepsilon) = \begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & & \rho^{n-2} \\ \vdots & & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix}$$

```
#AR(1) autocorrelation structure
```

```
ccc2=glm(Unknowndeaths~Coviddeaths,data=Covid,correlation =corAR1(form = ~
Number))
summary(ccc2)
```

3) ARMA(p,q)

```
##ARMA autocorrelation structure
ccc3=glm(Unknowndeaths~Coviddeaths,data=Covid,correlation =corARMA(p=2,q=0,form =
~ Number))
summary(ccc3)
```

Měla pandemie různý vliv na počet neobjasněných úmrtí na jaře a na podzim?

Různý intercept:

Neobjasněná úmrtí =  $\beta_1$ \*Covidová úmrtí +  $\beta_2$ \*I(Podzim) +  $\varepsilon$

```
c4=glm(Unknowndeaths~Coviddeaths+Autumn,data=Covid,correlation =corAR1(form = ~
Number))
summary(c4)
```

Na podzim máme méně neobjasněných úmrtí než na jaře, ovšem výsledek je naprosto nesignifikantní

Různý trend: přidáme interakční efekt

Neobjasněná úmrtí =  $\beta_1$ \*Covidová úmrtí +  $\beta_2$ \*I(Podzim) +  $\beta_3$ \* I(Podzim)\* Covidová úmrtí +  $\varepsilon$

```
c5=gls(Unknowndeaths~Coviddeaths+Autumn+Coviddeaths:Autumn,data=Covid,correlation
=corAR1(form = ~ Number))
summary(c5)
```

Na podzim je trend závislosti neobjasněných úmrtí na covidových úmrtí menší než na jaře, ovšem výsledek je naprosto nesignifikantní

## 7 LINEÁRNÍ MODELY SEZÓNŇNÍCH ČASOVÝCH ŘAD

### 7.1 Sezónní smíšené integrované procesy SARIMA

Sezónní smíšené integrované procesy SARIMA slouží k popisu časových řad, jejichž trend i sezónní složka mají stochastický charakter. Nyní si ukážeme, jak se postupuje při konstrukci příslušného modelu.

Přitom budeme předpokládat, že máme k dispozici řadu  $Y_t$  měsíčních pozorování, která vykazuje výrazné sezónní vlivy s počtem sezón  $L=12$  v průběhu roku.

1. Nejprve se zkonstruuje model ARIMA pro lednová měření ve tvaru

$$\Phi(B^{12}) \Delta_{12}^D Y_t = \Theta(B^{12}) \eta_t, \quad (9.1)$$

kde

$$\Phi(B^{12}) = 1 - \Phi_1 B^{12} - \Phi_2 B^{24} - \dots - \Phi_P B^{12P}$$

je **sezónní autoregresní operátor řádu  $P$**

$$\Theta(B^{12}) = 1 + \Theta_1 B^{12} + \Theta_2 B^{24} + \dots + \Theta_Q B^{12Q}$$

**sezónní operátor klouzavých součtů řádu  $Q$**  a

$$\Delta_{12} = 1 - B^{12}$$

**sezónní diferenční operátor**, pro nějž platí  $\Delta_{12}^D Y_t = (1 - B^{12})^D Y_t$ .

Model (8.2) se považuje za ARMA model pro lednová měření.

2. Provede se konstrukce ARMA modelů pro měření realizovaná v měsících únor až prosinec. Předpokládá se, tyto modely jsou přibližně stejné jako model (9.1).
3. Náhodné složky  $\eta_t$  by měly být vzájemně korelované, protože zřejmě existuje nějaká souvislost mezi lednovými a únorovými měřeními, mezi únorovými a březnovými měřeními atd. Předpokládá se tedy, že časovou řadu  $\eta_t$  lze popsat ARIMA modelem ve tvaru

$$\varphi(B) \Delta^d \eta_t = \vartheta(B) \varepsilon_t, \quad (9.2)$$

kde  $\varphi(B)$ ,  $\vartheta(B)$  jsou operátory zavedené v odstavci 8.1 a  $\varepsilon_t$  označuje bílý šum.

4. Spojením modelů (9.1) a (9.2) dostaneme výsledný SARIMA model ve tvaru

$$\varphi(B) \Phi(B^{12}) \Delta^d \Delta_{12}^D Y_t = \vartheta(B) \Theta(B^{12}) \varepsilon_t. \quad (9.3)$$



Takto zkonstruovaný **multiplikativní sezónní smíšený integrovaný model SARIMA** se zapisuje ve tvaru

$$\text{SARIMA}(p, d, q) \times (P, D, Q)_{12}.$$

V obecném případě má takový model tvar

$$\text{SARIMA}(p, d, q) \times (P, D, Q)_L,$$

kde  $L$  označuje počet sezón v daném období (počet měsíců nebo kvartálů v roce, ale také počet dnů v týdnu apod.).

Aditivní varianta sezónního smíšeného integrovaného modelu se používá jen zřídka.

## 7.2 Sezónní procesy klouzavých součtů SMA

**Sezónní proces klouzavých součtů** řádu  $Q$ , tj.  $SMA(Q)$ , můžeme obecně popsat modelem

$$Y_t = \varepsilon_t + \Theta_1 \varepsilon_{t-L} + \Theta_2 \varepsilon_{t-2L} + \dots + \Theta_Q \varepsilon_{t-QL}$$

nebo pomocí operátoru  $B$  jako

$$Y_t = (1 + \Theta_1 B^L + \Theta_2 B^{2L} + \dots + \Theta_Q B^{QL}) \varepsilon_t = \Theta_Q (B^L)^Q \varepsilon_t,$$

kde  $L$  značí počet sezón v daném období. Je zřejmé, že model  $SMA(Q)$  je speciálním případem modelu SARIMA, a to  $SARIMA(0,0,0) \times (0,0,Q)_L$ . Uvažovaný model je stacionární pro libovolné hodnoty parametrů a invertibilní v případě, že všechny kořeny rovnice  $\Theta_Q(B)^L = 0$  leží vně jednotkového kruhu v komplexní rovině.

### 7.3 Sezónní autoregresní modely SAR

**Sezónní autoregresní model** řádu  $P$ , tj.  $SAR(P)$ , má v obecném případě tvar

$$Y_t = \Phi_1 Y_{t-L} + \Phi_2 Y_{t-2L} + \dots + \Phi_P Y_{t-PL} + \varepsilon_t.$$

Pomocí operátoru  $B$  jej můžeme zapsat jako

$$\Phi_P(B^L) = (1 - \Phi_1 B^L - \Phi_2 B^{2L} - \dots - \Phi_P B^{PL}) = \varepsilon_t.$$

Uvedený model je také speciálním případem modelu SARIMA, a to  $SARIMA(0,0,0) \times (P,0,0)_L$ . Je invertibilní při libovolných hodnotách parametrů a stacionární, pokud všechny kořeny rovnice  $\Phi_P(B^L) = 0$  leží vně jednotkového kruhu v komplexní rovině.

#### 7.4 Sezónní smíšené modely SARMA

Uvažujeme **smíšený sezónní model SARMA**  $(P, Q)$ , který můžeme považovat za speciální případ modelu s označením  $\text{SARIMA}(0, 0, 0) \times (P, 0, Q)_L$ . Tento model lze zapsat ve tvaru

$$Y_t = \Phi_1 Y_{t-L} + \Phi_2 Y_{t-2L} + \dots + \Phi_P Y_{t-PL} + \varepsilon_t + \Theta_1 \varepsilon_{t-L} + \Theta_2 \varepsilon_{t-2L} + \dots + \Theta_Q \varepsilon_{t-QL}.$$

S použitím operátoru B jej můžeme vyjádřit jako

$$(1 - \Phi_1 B^L - \Phi_2 B^{2L} - \dots - \Phi_P B^{PL}) Y_t = (1 + \Theta_1 B^L + \Theta_2 B^{2L} + \dots + \Theta_Q B^{QL}) \varepsilon_t$$

neboli  $\Phi_P(B^L) Y_t = \Theta_Q(B^L) \varepsilon_t$ . Z vlastností modelů SMA a SAR vyplývá, že uvažované smíšené modely SARMA jsou stacionární, jestliže všechny kořeny rovnice  $\Phi_P(B^L) = 0$  leží vně jednotkového kruhu, a invertibilní, pokud všechny kořeny rovnice  $\Theta_Q(B)^L = 0$  leží také vně jednotkového kruhu v komplexní rovině.

```

#####SARIMA
library(astsa)
# Monthly totals of international airline passengers
AirPassengers
plot(AirPassengers)
plot(log(AirPassengers))
plot(diff(log(AirPassengers)))
acf(diff(log(AirPassengers)))
pacf(diff(log(AirPassengers)))
acf(diff(log(AirPassengers)),12)
pacf(diff(log(AirPassengers)),12)

#lednova mereni
leden=(0:11)*12+1
la=log(AirPassengers)
acf(la[leden])
pacf(la[leden])
#Ar(1) model
arima(la[leden],order=c(1,0,0))
auto.arima(la[leden])
#zvolime 0,1,0 pro lednovy model
acf(diff(la[leden]))
pacf(diff(la[leden]))
plot(diff(la,12))
#zde je nejaky trend
plot(diff(diff(la,12),1))

las=diff(diff(la,12),1)
acf(las)
pacf(las)
#predpokladame 0,0,1 model
sarima(log(AirPassengers),p=0,d=1,q=1,P=0,D=1,Q=0,S=12)
#Zjevne mi zustava autocorelace po 12 mesici
#viz Ljung-Box test - H0: autocorrelation rho_k=0, k is the lag
#takze musim modelovat autocorelaci v sezonim model P,D,Q
res=sarima(log(AirPassengers),p=0,d=1,q=1,P=0,D=1,Q=0,S=12)
acf(res$fit$residuals)

```

```
pacf(res$fit$residuals)

sarima(log(AirPassengers), p=0, d=1, q=1, P=0, D=1, Q=1, S=12)
#Trosicku lepsi BIC ma
sarima(log(AirPassengers), p=0, d=1, q=1, P=1, D=1, Q=0, S=12)

# fun with diagnostics
sarima(log(AirPassengers), 0, 1, 1, 1, 1, 0, 12, gg=TRUE, col=4)

# fun with the graphic
sarima.for(log(AirPassengers), 12, 0, 1, 1, 1, 1, 0, 12, gg=TRUE, col=4, main='arf')

out=sarima.for(log(AirPassengers), 12, 0, 1, 1, 1, 1, 0, 12, gg=TRUE, col=4, main='arf')
out$pred
out$se

??salmon
```

### 9.3 Výhody a nevýhody Boxova-Jenkinsova přístupu

Boxova-Jenkinsova metodologie má ve srovnání s metodami založenými na dekompozici následující výhody.

- Modely se rychle adaptují na změny v průběhu časové řady, proto je Boxova-Jenkinsova metodologie úspěšná i v těch případech, kde dekompozice selhává.
- Boxův-Jenkinsův přístup vykazuje nejlepší výsledky při analýze ekonomických časových řad.
- Boxův-Jenkinsův přístup je systematický, proto může být plně automatizován.

Nevýhody Boxova-Jenkinsova přístupu spatřujeme v tom, že:

- je vhodný jen pro časové řady o délce nejméně 50 pozorování,
- jeho praktické aplikace jsou mnohem náročnější než aplikace dekompozičních metod,
- výsledné modely, zejména modely s větším počtem parametrů, se obtížně interpretují.

## 9 VÍCEROZMĚRNÉ ČASOVÉ ŘADY A JEJICH CHARAKTERISTIKY

### 9.1 Pojem vícerozměrné časové řady

Vícerozměrné časové řady chápeme jako zobecnění jednorozměrných časových řad. Pod pojmem ***m*-rozměrné časové řady** rozumíme posloupnost chronologicky uspořádaných výsledků pozorování *m* různých náhodných veličin, tedy *m*-rozměrného náhodného vektoru  $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, \dots, Y_{mt})$ . Vyjdeme-li z teorie náhodných procesů, můžeme říci, že *m*-rozměrná časová řada představuje konkrétní realizaci nějakého *m*-rozměrného náhodného procesu.

***m*-rozměrnou časovou řadu** je možno definovat jako posloupnost náhodných vektorů ve tvaru

$\{Y(s, t) = s \in S, t \in T, S \times T \rightarrow \mathbb{R}\}$ , kde *T* je časová množina nabývající hodnot  $\{0, 1, 2, \dots\}$  a množina *S* reprezentuje *m*-prvkový výběrový prostor odpovídající *m* různým složkám časové řady. Veličina *m* se nazývá rozměr časové řady. V dalším výkladu budeme vícerozměrné časové řady značit zkráceně  $\{\mathbf{Y}_t\} = \{Y_{1t}, Y_{2t}, \dots, Y_{mt}\}$ .

**Slabá stacionarita** vícerozměrného časové řady je definována analogicky jako stacionarita jednorozměrné časové řady. Říkáme, že *m*-rozměrná časová řada je slabě stacionární, jestliže jsou splněny následující podmínky:



- a)  $\mathbf{E}(\mathbf{Y}_t) = \boldsymbol{\mu} < \infty$  pro všechny hodnoty  $t$ , kde  $\boldsymbol{\mu}$  je  $m$ -rozměrný sloupcový vektor středních hodnot jednotlivých časových řad.
- b)  $\mathbf{E}\left[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})^T\right] = \boldsymbol{\Sigma}$  pro všechny hodnoty  $t$ , kde  $\boldsymbol{\Sigma}$  je kovarianční matice typu  $m \times m$ , na jejíž diagonále jsou rozptyly jednotlivých časových řad a mimo diagonálu jejich kovariance.
- c)  $\mathbf{E}\left[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t+k} - \boldsymbol{\mu})^T\right] = \boldsymbol{\Gamma}_k$  pro všechny hodnoty  $t$ , kde  $\boldsymbol{\Gamma}_k$  je autokovarianční maticová funkce řádu  $k$  (typu  $m \times m$ ), na jejíž diagonále jsou autokovariance řádu  $k$  jednotlivých časových řad v čase  $t$  pro danou hodnotu  $k$  a mimo diagonálu kovariance těchto řad, přičemž jedna je uvažována v čase  $t$  a druhá v čase  $t+k$ .

Zřejmě platí:  $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma}$ .

Za předpokladu slabé stacionarity jsou teoretické a empirické charakteristiky  $m$ -rozměrné časové řady podobné jednorozměrným, s tím rozdílem, že střední hodnota je vektorem, rozptyl odpovídá  $m \times m$  rozměrné kovarianční matici a autokovariance odpovídá  $m \times m$  autokovarianční matici. Také autokorelace bude odpovídat  $m \times m$  autokorelační matici.

## 12 LINEÁRNÍ MODELY VÍCEROZMĚRNÝCH ČASOVÝCH ŘAD

### 12.1 Vícerozměrný lineární proces

Stacionární vícerozměrný lineární proces  $\{\mathbf{Y}_t\}$  s nulovou střední hodnotou může být vyjádřen jako lineární kombinace řady nekorelovaných identicky rozdělených náhodných vektorů. Takový  **$m$ -rozměrný lineární proces** je definován jako nekonečná řada

$$\mathbf{Y}_t = \mathbf{I}_1 \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\psi}_2 \boldsymbol{\varepsilon}_{t-2} \dots,$$

v níž  $\mathbf{I}_1$  je jednotková matice typu  $m \times m$ ,  $\boldsymbol{\psi}_k$ ,  $k=1,2, \dots$ , matice parametrů typu  $m \times m$

a  $\{\boldsymbol{\varepsilon}_t\}$  představuje  **$m$ -rozměrný proces bílého šumu** s nulovým vektorem středních hodnot a autokovarianční maticovou funkcí.

Pro autokovarianční matici bílého šumu platí:

$$\boldsymbol{\Gamma}_k = \mathbf{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-k}) = \begin{cases} \boldsymbol{\Sigma}_\varepsilon & \text{pro } k = 0, \\ \mathbf{0} & \text{pro } k \neq 0. \end{cases}$$

Pomocí operátoru zpětného posunutí (operátoru zpoždění)  $B$  lze vztah (12.1) zapsat jako

$$\mathbf{Y}_t = \boldsymbol{\Psi}(B)\boldsymbol{\varepsilon}_t,$$

přičemž

$$\boldsymbol{\Psi}(B) = \mathbf{I}_1 + \sum_{k=1}^{\infty} \boldsymbol{\Psi}_k B^k.$$

Vícerozměrný lineární proces je slabě **stacionární**, jestliže platí

$$\sum_{j=k}^{\infty} \psi_{ij,k}^2 < \infty,$$

kde  $\psi_{ij,k}$  je  $(i, j)$ -tý prvek matice  $\boldsymbol{\Psi}_k$ .

## 12.1 Vícerozměrné smíšené procesy

**Model** vícerozměrného smíšeného procesu řádu  $p$  a  $q$  (VARMA( $p,q$ )) má tvar

$$\mathbf{Y}_t = \varphi_1 \mathbf{Y}_{t-1} + \varphi_2 \mathbf{Y}_{t-2} + \dots + \varphi_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t + \vartheta_1 \boldsymbol{\varepsilon}_{t-1} + \vartheta_2 \boldsymbol{\varepsilon}_{t-2} + \dots + \vartheta_q \boldsymbol{\varepsilon}_{t-q}$$

nebo (s použitím operátoru B)

$$\varphi_p(B) \mathbf{Y}_t = \vartheta_q(B) \boldsymbol{\varepsilon}_t,$$

příčemž operátory

$$\varphi_p(B) = (\mathbf{I}_1 - \boldsymbol{\varphi}_1 B - \boldsymbol{\varphi}_2 B^2 - \dots - \boldsymbol{\varphi}_p B^p)$$

$$\text{a } \vartheta_q(B) = (\mathbf{I}_1 + \boldsymbol{\vartheta}_1 B + \boldsymbol{\vartheta}_2 B^2 \dots + \boldsymbol{\vartheta}_q B^q)$$

Jsou polynomiální matice. Proces VARMA( $p,q$ ) je stacionární, jestliže všechny kořeny rovnice  $|\mathbf{I}_1 - \boldsymbol{\varphi}_1 B - \boldsymbol{\varphi}_2 B^2 - \dots - \boldsymbol{\varphi}_p B^p| = 0$  leží vně jednotkového kruhu v komplexní rovině, a invertibilní, pokud se všechny kořeny rovnice  $|\mathbf{I}_1 + \boldsymbol{\vartheta}_1 B + \boldsymbol{\vartheta}_2 B^2 \dots + \boldsymbol{\vartheta}_q B^q| = 0$  nacházejí vně jednotkového kruhu v komplexní rovině.

## 12.2 Kauzální vztahy

Jedním ze základních problémů, jímž se zabývá analýza vícerozměrných časových řad, je studium kauzálních (příčinných) vztahů mezi jednotlivými časovými řadami. Kauzalita je pojem

filozofický. Aristoteles chápe kauzalitu jako vnitřní spojení mezi příčinou a účinkem nějaké síly. Empirik D. Hume se domnívá, že když je jeden jev následován jevem druhým a výskyt druhého jevu po jevu prvním je pravidelný a mnohokrát ověřený, pak jev první je příčinou a jev druhý důsledkem. Filozofické pojetí kauzality není pro statistiku vhodné. V současnosti se při ověřování kauzální souvislosti mezi jednotlivými časovými řadami používají dva základní přístupy:

1. Grangerovo pojetí kauzality,
2. analýza odezvy na impuls (analýza impuls – reakce).

### 12.3 Grangerovo pojetí kauzality

Grangerova koncepce kauzality není v souladu s filozofickým pojetím kauzality. O **kauzalitě podle Grangera** se mluví v případě, že existuje korelovanost mezi současnou hodnotou jedné časové řady a minulými (zpožděnými) hodnotami jiných časových řad.

Pro testování kauzality jsou nejvhodnější modely VAR. Struktura modelů VAR umožňuje převést testování kauzality na vyšetřování nulovosti bloků určitých parametrů modelu VAR.

Nejprve objasníme běžně používanou terminologii v případě, že analyzujeme nějakou  $m$ -rozměrnou časovou řadu, tj. sledujeme paralelně časové změny náhodných veličin  $Y_1, Y_2, \dots, Y_m$ .

- Jestliže zpožděné hodnoty náhodné veličiny  $Y_i$  v modelu VAR významně ovlivňují jinou náhodnou veličinu  $Y_j$ , pak veličina  $Y_i$  kauzálně působí podle Grangera na veličinu  $Y_j$ .
- Jestliže veličina  $Y_i$  kauzálně působí podle Grangera na veličinu  $Y_j$ , ale veličin  $Y_j$  kauzálně nepůsobí podle Grangera na veličinu  $Y_i$ , pak existuje jednosměrná závislost  $Y_j$  na  $Y_i$ .
- Jestliže veličina  $Y_i$  kauzálně působí podle Grangera na veličinu  $Y_j$  a také veličina  $Y_j$  kauzálně působí podle Grangera na veličinu  $Y_i$ , pak mezi veličinami  $Y_i$  a  $Y_j$  existuje zpětná vazba.
- Jestliže veličina  $Y_j$  kauzálně nepůsobí podle Grangera na veličinu a také veličina  $Y_j$  kauzálně nepůsobí podle Grangera na veličinu  $Y_i$ , pak veličiny  $Y_i$  a  $Y_j$  jsou podle Grangera nezávislé.

Použití zavedené terminologie ukážeme na příkladu převzatém z monografie [9]. Budeme uvažovat dvourozměrný model VAR(1). Z obecného (maticového) vztahu (12.2) dostaneme pro model VAR(1) dvě rovnice

$$\begin{aligned} Y_{1t} &= \varphi_{11}Y_{1,t-1} + \varphi_{12}Y_{2,t-1} + \varepsilon_{1t}, \\ Y_{2t} &= \varphi_{21}Y_{1,t-1} + \varphi_{22}Y_{2,t-1} + \varepsilon_{2t}. \end{aligned}$$

Pak stačí otestovat nulové hypotézy o nulovosti parametrů uvažovaného modelu a na základě výsledků těchto testů rozhodnout, zda existují kauzální vztahy mezi náhodnými veličinami  $Y_1$  a  $Y_2$ . Pak existují následující možnosti.

- Je-li  $\varphi_{12} \neq 0$ , pak veličina  $Y_2$  kauzálně působí podle Grangera na veličinu  $Y_1$ .
- Je-li  $\varphi_{21} \neq 0$ , pak veličina  $Y_1$  kauzálně působí podle Grangera na veličinu  $Y_2$ .
- Jestliže platí  $\varphi_{12} \neq 0$  a  $\varphi_{21} = 0$ , pak existuje jednosměrná závislost veličiny  $Y_1$  na veličině  $Y_2$ .

- Jestliže platí  $\varphi_{12} = 0$  a  $\varphi_{21} \neq 0$ , pak existuje jednosměrná závislost veličiny  $Y_2$  na veličině  $Y_1$ .
- Jestliže platí  $\varphi_{12} = 0$  a  $\varphi_{21} = 0$ , pak jsou veličiny  $Y_1$  a  $Y_2$  podle Grangera nezávislé.

Vyšetřování kauzality podle Grangera se samozřejmě netýká jen kauzální závislosti mezi dvěma jednorozměrnými časovými řadami.

```
#####VARIMA
```

```
install.packages("devtools")
library(devtools)
install_github("d-/MTS")
library(MTS)
setwd("H:/StochMod")
data=read.csv("KurzyPlus.csv", header = TRUE, sep = ";", dec=",")
head(data)
data=na.omit(data)
library(forecast)
Euro=diff(log(na.omit(data$X1EUR.USD)))
Gold=diff(log(na.omit(data$X1gold.USD)))
cor(Euro,Gold)
cor(Euro[-1],Gold[-length(Gold)]) #zavislost Eura na predchozim goldu
```



```

d=cbind(Euro,Gold)
ccm(d)

#####simulovany priklad
phi=matrix(c(0,-0.6,0.6,0),2,2); theta=matrix(c(0,0,0,0),2,2)
phi
# prvni radek zavislost prvni promenne phi11,phi12
# druhy radek zavislost druhe promenne phi21,phi22
theta
sigma=diag(2)
m1=VARMAsim(300,arlags=c(1),malags=c(1),phi=phi,theta=theta,sigma=sigma)
zt=m1$series
zt
cor(zt)
cor(cbind(zt[1:299,1],zt[2:300,2])) #phi21 phi 2rada v zavislosti na 1rade
phi
cor(cbind(zt[2:300,1],zt[1:299,2])) #phi12 phi 1rada v zavislosti na 2rade
ccm(zt) #cross correlation
Eccm(zt) #Extended cross correlation with Ljung box p-values
m2=VARMA(zt,p=1,q=1,include.mean=FALSE)

#####Euro Gold

```

```
Eccm(d)
m3=VARMA(d,p=1,q=0,include.mean = FALSE)
#significantni je phi21 a phi22
#Euro pusobi jednosmerne kauzalne podle Grangera na Zlato
#Zlato ma navíc AR(1)
```

```
VARMAPred(m3,h=1) #prediction na horizont 1
ccm(m3$residuals)
m4=VARMA(d,p=1,q=1,include.mean = FALSE)
```

## 12.4 Odezva na impuls

V této části ukážeme, jak se vysvětluje **odezva** vybrané vysvětlované proměnné jedné časové řady **na impuls** v některé rovnici modelu VAR. V m-rozměrném modelu VAR můžeme sledovat (v čase měřeném od okamžiku impulsu) celkem  $m^2$  odezev, pro každou z m vysvětlovaných proměnných vždy m odezev na impulsy v jednotlivých rovnicích. Za předpokladu, že model VAR je stacionární (s nulovým vektorem středních hodnot  $\mu$ ), vliv impulsů ve všech  $m^2$  případech odezní, i když s různou rychlostí.

```
library(MTS)
```

```
phi=matrix(c(0.6,0,0.2,0.3),2,2); theta=matrix(c(0,0,0,0),2,2)
```

```
# první radek závislost na první prom
```

```
# druhý radek závislost na druhé proměnné
```

```
phi
```

```
sigma=diag(2)
```

```
m1=VARMAsim(300,arlags=c(1),malags=c(1),phi=phi,theta=theta,sigma=sigma)
```

```
zt=m1$series
```

```
cor(zt)
```

```
cor(cbind(zt[1:299,1],zt[2:300,2])) #phi 2rada v závislosti na 1rade
```

```
cor(cbind(zt[2:300,1],zt[1:299,2])) #phi 1rada v závislosti na 2rade
```

```
ccm(zt) #cross correlation
```

```
Eccm(zt) #Extended cross correlation with Ljung box p-values
```

```
m2=VARMA(zt,p=1,q=1,include.mean=FALSE)
```

**Příklad 12.1.** Uvažujme dvourozměrný modelů VAR(1) s nulovou střední hodnotou

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} 0,6 & 0,2 \\ 0 & 0,3 \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

a jeho odezvu na jednotkový impuls v čase  $t = 0$  v první rovnici, tj. pro  $\varepsilon_{1,0} = 0$ . V tomto případě (při nulových ostatních hodnotách bílého šumu) budou odezvy nabývat postupně následujících hodnot:

$$\begin{pmatrix} Y_{1,0} \\ Y_{2,0} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,0} \\ \varepsilon_{2,0} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} Y_{1,1} \\ Y_{2,1} \end{pmatrix} = \begin{pmatrix} 0,6 & 0,2 \\ 0 & 0,3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,6 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} Y_{1,2} \\ Y_{2,2} \end{pmatrix} = \begin{pmatrix} 0,6 & 0,2 \\ 0 & 0,3 \end{pmatrix} \begin{pmatrix} 0,6 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,36 \\ 0 \end{pmatrix}, \dots$$

Podobně, odezvy na jednotkový impuls v čase  $t = 0$  ve druhé rovnici, tj. pro  $\varepsilon_{2,0} = 0$ , budou následující:

$$\begin{pmatrix} Y_{1,0} \\ Y_{2,0} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,0} \\ \varepsilon_{2,0} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} Y_{1,1} \\ Y_{2,1} \end{pmatrix} = \begin{pmatrix} 0,6 & 0,2 \\ 0 & 0,3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0,2 \\ 0,3 \end{pmatrix},$$

$$\begin{pmatrix} Y_{1,2} \\ Y_{2,2} \end{pmatrix} = \begin{pmatrix} 0,6 & 0,2 \\ 0 & 0,3 \end{pmatrix} \begin{pmatrix} 0,2 \\ 0,3 \end{pmatrix} = \begin{pmatrix} 0,18 \\ 0,09 \end{pmatrix}, \dots$$

*Poznámka.* Při výpočtech jsme předpokládali, že  $\mathbf{Y}_t$  je nulový vektor.

Je zřejmé, že v obou případech odezvy v průběhu času odeznívají. Odezva vysvětlované proměnné  $Y_2$  na jednotkový impuls v první rovnici je nulová, což lze vysvětlit tím, že parametr  $\varphi_{21}$  je roven nule.

## 10 Prostorová regrese

Podobně jako v regresi s časovou autokorelací, vyjádřenou pomocí autokorelační funkce, musíme zde podchytit autokorelaci prostorovou. Časová autokorelace závisí na časové vzdálenosti dvou okamžiků. Prostorová autokorelace pak závisí na vzdálenosti dvou měření v prostoru. Body blíže k sobě předpokládáme, že jsou více korelované, než body vzdálené.

Variogram

$$\gamma(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2} E[(Z(\mathbf{x}_1) - Z(\mathbf{x}_2))^2]$$

Vyjadřuje právě takovou prostorovou korelaci mezi dvěma body v prostoru. Malé hodnoty odpovídají závislosti, velké nezávislosti. Za předpokladu slabé stacionarity, tudíž konstantnosti prvního a druhého momentu počítáme variogram pro vzdálenost  $h$  dvou bodů v prostoru.

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{Var}[(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))].$$

Uvědomme si, že za slabé stacionarity je střední hodnota  $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}) = 0$ .

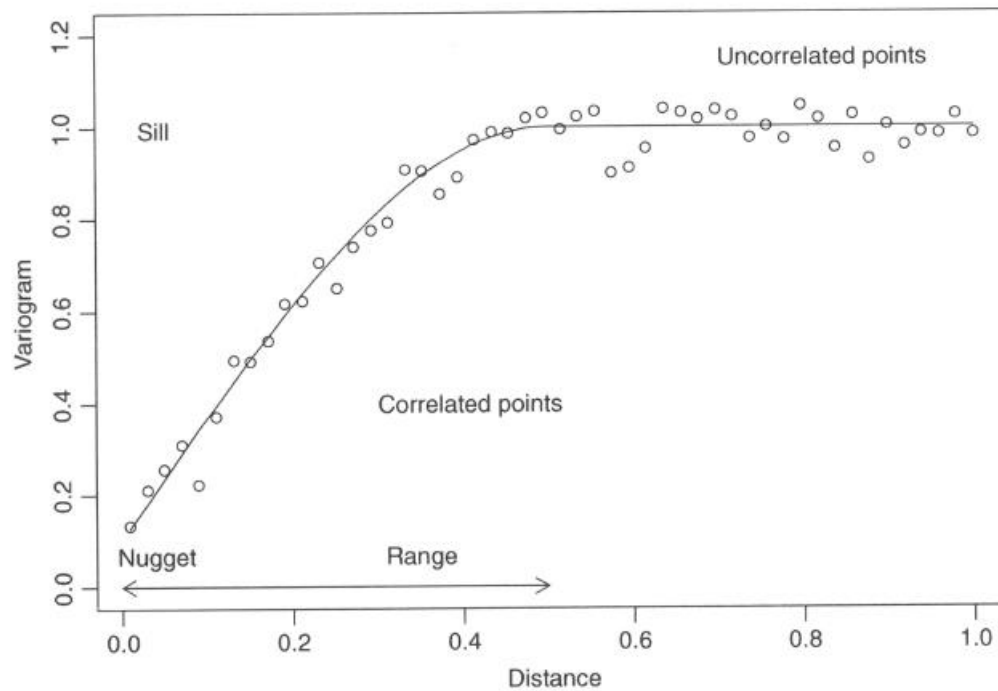
Výběrový variogram je pak

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2 N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{x}_i + \mathbf{h}) - z(\mathbf{x}_i)]^2$$

$N(\mathbf{h})$  udává počet dvojic bodů ve vzdálenosti  $h$ . Těch je obvykle velmi málo a proto se používají tzv. jádrové odhady variogramu.

Obvykle se také předpokládá isotropie, tudíž že posunutí ve všech směrech je ekvivalentní, tedy že  $h$  je čistě vzdálenost a ne směrový vektor.

Příklad variogramu



**Fig. 7.2** Variogram with fitted line. The sill is the asymptotic value and the range is the distance where this value occurs. Pairs of points that have a distance larger than the range are uncorrelated. The nugget effect occurs if  $\hat{\gamma}(\mathbf{h})$  is far from 0 for small  $\mathbf{h}$

Nugget effect reprezentuje nespojitost proměnné způsobené prostorovou strukturou ve vzdálenostech menších než je minimální vzdálenost mezi dvěma měřeními.

Isotropii je možné prozkoumat vytvořením variogramu pro různé směrové úhly.

```
#####Spatial correlation
setwd("C:/Users/Tomas/Disk Google/Vyuka/Stochasticke Modely")
Boreality <- read.table(file = "Boreality.txt", header = TRUE, dec = ".")
Boreality$nBor #Pocet druhu patricich do Boreal Coenosis
Boreality$nTot #Celkovy pocet druhu
qqnorm((Boreality$nBor+1)/Boreality$nTot)
qqline((Boreality$nBor+1)/Boreality$nTot)
Boreality$Bor <- sqrt(1000*(Boreality$nBor+1)/Boreality$nTot)
qqnorm(Boreality$Bor)
qqline(Boreality$Bor)
plot(Boreality$Bor)

#linearni model Boreality ~ a+ b*wetness
B.lm <- lm(Bor ~ Wet, data=Boreality)
summary(B.lm)

plot(B.lm) #Vypada dobre

#Ale co prostorovy efekt?
#Vytvorme graf residui v zavislosti na prostoru

E <- rstandard(B.lm)
library(gstat)
mydata <- data.frame(E=E, x=Boreality$x, y=Boreality$y)
```



```
library(sp)
coordinates(mydata) = ~x+y
bubble(mydata)
#Residuals <- SpatialPointsDataFrame(cbind(Boreality$x, Boreality$y),data.frame(E))
#bubble(Residuals)
#Jasna prostorova nehomogenita residui je patrna, to znamena ze bud chybi covariat,
nebo je tam prostorova corelace

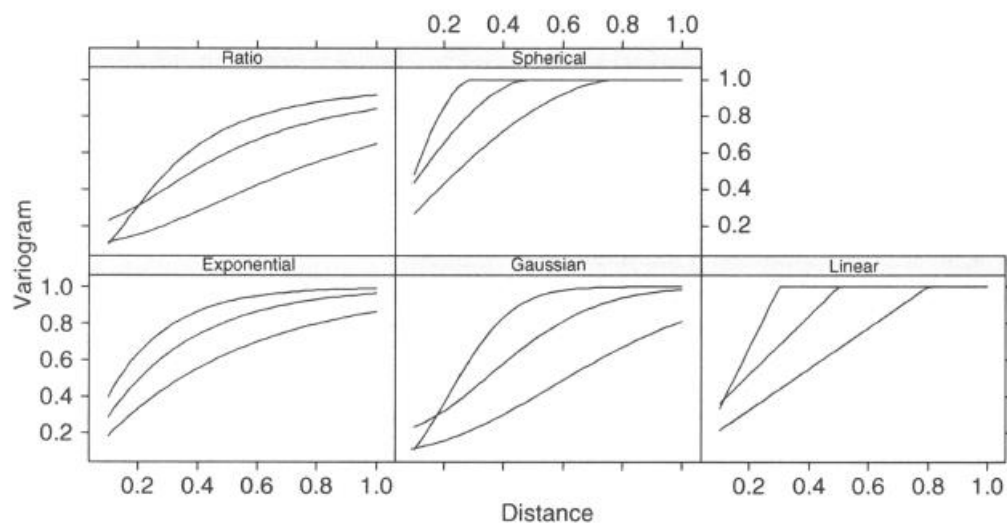
#####Variogram
vario1 <- variogram(E~1,data=mydata)
plot(vario1)

#####4 směrové Variogramy pro 4 kvadranty.

vario2 <- variogram(E~1,data=mydata, alpha = c(0,45,90,135))
plot(vario2)
```

Chceme-li modelovat prostorovou autokorelaci, musíme zvolit vhodný model (Podobně jako AR(1), ARMA(1,1) u časové autokorelace. Možnosti jsou:

- Exponential correlation using the function `corExp`.
- Gaussian correlation using the function `corGaus`.
- Linear correlation using the function `corLin`.
- Rational quadratic correlation using the function `corRatio`.
- Spherical correlation using the function `corSpher`.



**Fig. 7.4** Different variogram patterns. The three lines in the same panel were obtained using different values for the range and nugget

Exponenciální korelogram je:

$$\gamma(s, \rho) = 1 - e^{-\frac{s}{\rho}}$$

Rho je range a s je vzdálenost.

S nugget effectem dostaneme tvar:

$$\gamma(s, \rho) = \begin{cases} c_0 + (1 - c_0)(1 - e^{-\frac{s}{\rho}}) & \text{if } s > 0 \\ 0 & \text{if } s = 0 \end{cases}$$

Variogramy se dostanou vynásobením korelogramu rozptylem. Gausovský korelogram je podobný Exponenciálnímu jenom s tím rozdílem, že je  $1 - \exp(-(s/\rho)^2)$

### Prostorová autokorelace uvnitř gls

```
#Variogram spocteny primo z modelu gls ( z jeho rezidui) pomoci nlme package
#resType = "pearson" specifikuje, ze chceme standardizovana rezidua
```

```
library(nlme)
f1 <- formula(Bor ~ wet)
B1.gls <- gls(f1, data=Boreality)
Vario.gls <- Variogram(B1.gls, form = ~ x+y, robust = TRUE, maxDist = 2000, resType =
"pearson")
plot(Vario.gls,smooth = TRUE)
```

#Pokud bychom videli horizontalni caru fitu, pak zrejme se jedna o prostorovou nezavislost. To neni pripad zde.

```
B1A <- gls(f1, correlation = corSpher(form = ~x+y, nugget = TRUE), data = Boreality)
B1B <- gls(f1, correlation = corLin(form = ~x+y, nugget = TRUE), data = Boreality)
B1C <- gls(f1, correlation = corRatio(form = ~x+y, nugget = TRUE), data = Boreality)
B1D <- gls(f1, correlation = corGaus(form = ~x+y, nugget = TRUE), data = Boreality)
```

```
B1E <- gls(f1, correlation = corExp(form = ~x+y, nugget = TRUE), data = Boreality)
AIC(B1.gls, B1A, B1C,B1D,B1E)
BIC(B1.gls, B1A, B1C,B1D,B1E)
anova(B1.gls,B1E) #Likelihood ratio test comparing two models.
```

```
#Odstranili jsme prostorovou autokorelaci z rezidui?
VarioE <- Variogram(B1E, form = ~ x+y, robust = TRUE, maxDist = 2000, resType =
"pearson")
plot(VarioE,smooth = TRUE)
```

```
#normalizovany rezidua jsou nyní horizontální, tudiz jsme odstranili prostorovou
autokorelaci z modelu
VarioEN <- Variogram(B1E, form = ~ x+y, robust = TRUE, maxDist = 2000, resType =
"normalized")
plot(VarioEN,smooth = TRUE)
```

```
summary(B1E)
```

```
#####
data(meuse)
head(meuse)
```