

MOŽNOSTI ANALÝZY ČASOVÝCH RADOV V JAZYKU R

Michal Páleš

1 ÚVOD

Cieľom príspevku je podať základnú informácie o analýze časových radov v jazyku R v kontexte publikácie (Šoltés a kol., 2015, **príklad 6.2**, s. 270). V postupnom slede uvádzame zodpovedajúce riešenie všetkých úloh predmetného príkladu v jazyku R (v komparácii s publikáciou, ktorá obsahuje riešenia v softvéroch *SAS Enterprise Guide* a *Statgraphics Centurion*). V príspevku sa zameriavame len na technickú stránku výpočtu a tieto poznatky uvádzame v kontexte s možným využitím jazyka R a analýzy časových radov v aktuárskej praxi (Páleš, 2016, 2017) a tiež ako pomôcku pre študentov k predmetom *Úvod do aktuárstva* a *Softvérové aplikácie pre aktuárov*.

2 VSTUPNÉ ÚDAJE

Využívame modelové údaje podľa tab. 6.12 (upozorňujeme, že v publikácii uvedenej vyššie nekorešponujú hodnoty tab. 6.11 a 6.12), ktoré sa vzťahujú na štvrt'roky sledovaného obdobia rokov 2012 a 2015. Jedná sa krátkodobý intervalový časový rad (ďalej aj ČR) extenzitnej neodvodenej veličiny.

```
y<-c(20,17,43,55,42,34,65,76,56,50,85,100,75,70,108,123)
yts<-ts(y,frequency=4,start=c(2012,1))
```

V prípade potreby môžeme tieto načítať zo súboru príkazom `read.csv`, resp. `read.csv2`. Do prostredia jazyka R však možno importovať časové rady nielen textového formátu (s koncovkou `***.CSV`), ale aj formátu SAS, štruktúrovaných formátov, ako je DBF, resp. DB, apod.

3 FORMULÁCIA ÚLOH

- a) *Vypočítame kľzavé a centrované kľzavé priemery ČR. Graficky znázorníme vývoj ČR a jeho centrovaných kľzavých priemerov.*
- b) *Pre analýzu sezónne-náhodného kolísania v štvrt'rokoch sledovaných rokov uskutočnime sezónnu dekompozíciu časového radu.*
- c) *Vypočítame sezónne očistený ČR a odhadneme trend jeho vývoja regresným modelom. Graficky znázorníme tento rad a jeho regresný model trendu.*
- d) *Odhadneme náhodnú zložku ČR.*
- e) *Extrapoláciou určíme prognózu na všetky štvrt'roky roku 2016 bez aj so zohľadnením sezónnosti.*
- f) *Graficky zobrazíme skutočné a sezónne zaťažené hodnoty vyrovnané regresným modelom spolu so sezónne korigovanými prognózami na štvrt'roky roku 2016.*

4 RIEŠENIE

a)

Kĺzavé a centrované kĺzavé priemery vypočítame príkazom (upozorňujeme, že v publikácii uvedenej vyššie (s. 273) sú pri vzťahoch zamenené vypočítané výsledné hodnoty KP a CKP)

```
KP<-filter(yts, rep(1/4, 4))
```

```
CKP<-filter(KP, rep(1/2, 2))
```

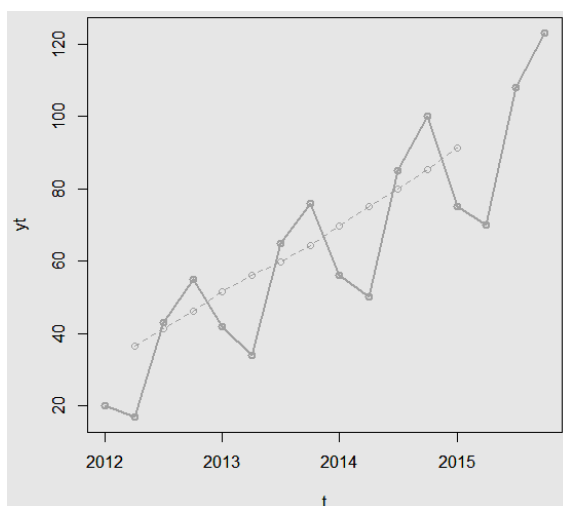
a dostávame výstup

```
      Qtr1  Qtr2  Qtr3  Qtr4
2012     NA 33.75 39.25 43.50
2013 49.00 54.25 57.75 61.75
2014 66.75 72.75 77.50 82.50
2015 88.25 94.00     NA     NA

      Qtr1  Qtr2  Qtr3  Qtr4
2012     NA 36.500 41.375 46.250
2013 51.625 56.000 59.750 64.250
2014 69.750 75.125 80.000 85.375
2015 91.125     NA     NA     NA
```

Pre grafický výstup nižšie volíme príkazy (tento možno vizuálne upraviť pomocou syntaxe funkcie plot)

```
plot(yts, type="o", lwd=2, col="grey", xlab="t", ylab="yt")
lines(CKP, type="o", lty=2, col="grey")
```



Predbežným odhadom trendovej zložky je teda časový rad centrovaných kĺzavých priemerov. Nevýhodou ČR CKP je jeho skrátenie o 4 hodnoty oproti pôvodnému časovému radu (čo môžeme aj pozorovať na obrázku vyššie, resp. vo výstupe CKP) a nemôžeme ho preto použiť na prognózovanie budúcich hodnôt trendu.

b)

Pretože so zmenou trendu sa systematicky mení rozsah sezónneho kolísania pri meraní sezónnosti budeme vychádzať z *multiplikatívneho modelu*. Cyklická zložka sa vzhľadom na malú dĺžku ČR nedá identifikovať. Výstup pre takúto sezónnu dekompozíciu získame v R pomocou funkcie `decompose`

```
SD<-decompose(yts,type="multiplicative")
```

Pričom pre komplexný R výstup je potrebné zavolať premennú `SD`. Z nej pomocou príkazu

```
s_ind<-SD$figure*100
```

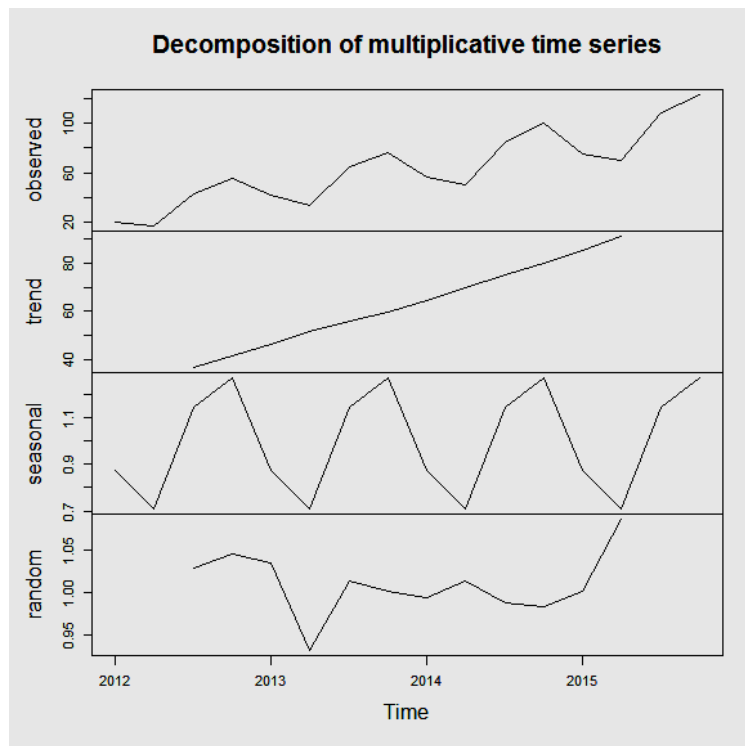
môžeme následne získať hodnoty sezónnych indexov (v %)

```
[1] 87.70476 70.72710 114.49820 127.06993
```

Pomocou príkazu

```
plot(SD)
```

získame veľmi užitočný grafický výstup všetkých zložiek ČR (je potrebné si však opätovne uvedomiť, že sa jedná o krátkodobý ČR).



c)

Sezónne očistený časový rad získame ako

```
yadj<-y/rep(SD$figure,4)
```

```
[1] 22.80378 24.03605 37.55517 43.28325 47.88794 48.07210 56.76945 59.80958  
[9] 63.85058 70.69426 74.23697 78.69682 85.51417 98.97196 94.32462 96.79709
```

Ak uvažujeme s regresným modelom *lineárneho, kvadratického a exponenciálneho trendu* volíme príkazy

```
x<-c(1:length(yadj)) # postupnosť jednotlivých hodnôt 1:16  
modell<-lm(yadj ~ x)  
summary(modell)
```

a dostávame výstup

```
Call:  
lm(formula = yadj ~ x)  
  
Residuals:  
    Min       1Q   Median       3Q      Max   
-5.4859 -1.6774 -0.7885  2.0227  8.1862  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)      
(Intercept)  19.3113     1.7950   10.76 3.75e-08 ***  
x              5.1053     0.1856   27.50 1.38e-13 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 3.423 on 14 degrees of freedom  
Multiple R-squared:  0.9818,    Adjusted R-squared:  0.9805  
F-statistic: 756.4 on 1 and 14 DF,  p-value: 1.382e-13
```

```
model2<-lm(yadj ~ poly(x,2,row=T)) # nie ortogonálny polynóm  
summary(model2)
```

```
Call:  
lm(formula = yadj ~ poly(x, 2, raw = T))  
  
Residuals:  
    Min       1Q   Median       3Q      Max   
-4.6733 -2.2225 -0.5796  1.4148  8.5345  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)      
(Intercept)  17.33788     2.95550    5.866 5.53e-05 ***  
poly(x, 2, raw = T)1  5.76312     0.80018    7.202 6.92e-06 ***  
poly(x, 2, raw = T)2 -0.03869     0.04576   -0.846  0.413
```

```

---
Signif. codes:  0  '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.458 on 13 degrees of freedom
Multiple R-squared:  0.9828,    Adjusted R-squared:  0.9801
F-statistic: 370.8 on 2 and 13 DF,  p-value: 3.429e-12

```

Poznámka.

Pre exponenciálny trend môžeme využiť napr. funkciu pre *zovšeobecnené lineárne regresné modely* glm v tvare: `glm(yadj ~ x, family = gaussian(link = 'log'))`. Tento model však v našom prípade považujeme za najmenej relevantný.

Ako vhodný model sezónne očisteného časového radu vyberieme model lineárneho trendu. Jeho odhadom je priamka

$$\hat{y}_t = 19,311 + 5,105 \cdot t$$

Má neskreslenú štandardnú chybu rezíduí (*RMSE*, vo výstupe pozri `Residual standard error`) nižšiu v porovnaní s kvadratickým (aj exponenciálnym) trendom. Všimnime si tiež, že posledný parameter kvadratického trendu je štatisticky nevýznamný (vo výstupe pozri `Signif. codes`).

Pre prognózu aj špecifický grafický výstup je vhodné využiť knižnicu `forecast`

```

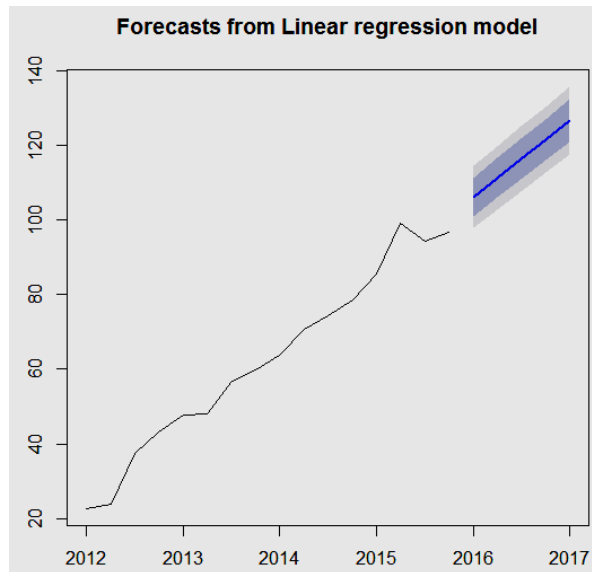
library(forecast)
yadjts<-ts(yadj,frequency=4,start=c(2012,1))
fit <- tslm(yadjts ~ trend)
nsf<-forecast(fit,h=4)

```

čím dostávame sezónne nezaťažené prognózy na 4 štvrťroky nasledujúceho roka podľa zvoleného lineárneho trendu ČR vrátane 80 % a 95 % intervalu spoľahlivosti

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2016 Q1	106.1017	100.9031	111.3003	97.81202	114.3914
2016 Q2	111.2070	105.9016	116.5124	102.74698	119.6670
2016 Q3	116.3123	110.8907	121.7340	107.66695	124.9577
2016 Q4	121.4176	115.8709	126.9643	112.57288	130.2624

a príkazom `plot(forecast(fit,h=5))` grafický výstup (kde lineárny trend je automaticky zobrazovaný až v prognóze).



Pre ešte detailnejší výstup možno voliť `summary(nsf)`.

d)

Odhad náhodnej zložky bol súčasťou výstupu (aj grafického) sezónnej dekompozície (pozri časť b)). Výstup získame ako

```
rand<-as.vector(na.omit(SD$random))
```

```
[1] 1.0289089 1.0461209 1.0354149 0.9311786 1.0137402 1.0009972 0.9937834
[8] 1.0135377 0.9881793 0.9837103 1.0016301 1.0861121
```

Normalitu rozdelenia reziduálnej zložky možno ďalej posudzovať graficky, resp. testovať (aj v R) pomocou rôznych štatistických testov.

e)

Sezónne nezaťažené štvrťročné prognózy sme uviedli v časti c) s využitím knižnice `forecast`. Po *zohľadnení sezónnosti* tieto matematicky (bez použitia ďalších doplnkových knižníc) vypočítame jednoducho prostredníctvom kódu

```
b0<-modell$coefficients[1]
b1<-modell$coefficients[2]
xp<-c((length(x)+1):(length(x)+4))
sf<-(b0+b1*xp)*(s_ind/100)
```

a dostávame výstup pre jednotlivé štvrťroky roka 2016

```
[1] 93.05623 78.65349 133.17552 154.28532
```

Poznámka.

Pri nezohľadnení sezónnosti je rozdiel v kóde len v časti `nsf<-(b0+b1*xp)`.

f)

Nižšie uvádzame kód (bez použitia ďalších doplnkových knižníc) pre jednoduchý a prehľadný komparatívny grafický výstup (čitateľ jednotlivé grafy môže rozlíšiť podľa názvu osi y)

```
par(mfrow=c(2,2)) # vzhľad výstupu
```

```
plot(yts,type="o",lwd=2,col="grey",xlab="t",ylab="yt")
```

```
nsf<-(b0+b1*xp)
```

```
yf2<-c(yadj,nsf)
```

```
yfts<-ts(yf2,frequency=4,start=c(2012,1))
```

```
plot(yfts,type="o",lwd=2,col="darkblue",xlab="t",ylab="yt+seas  
_adj+forecast_non_seas");abline(v=2016,col="red",lwd=2)
```

```
yf3<-c(yadj,sf)
```

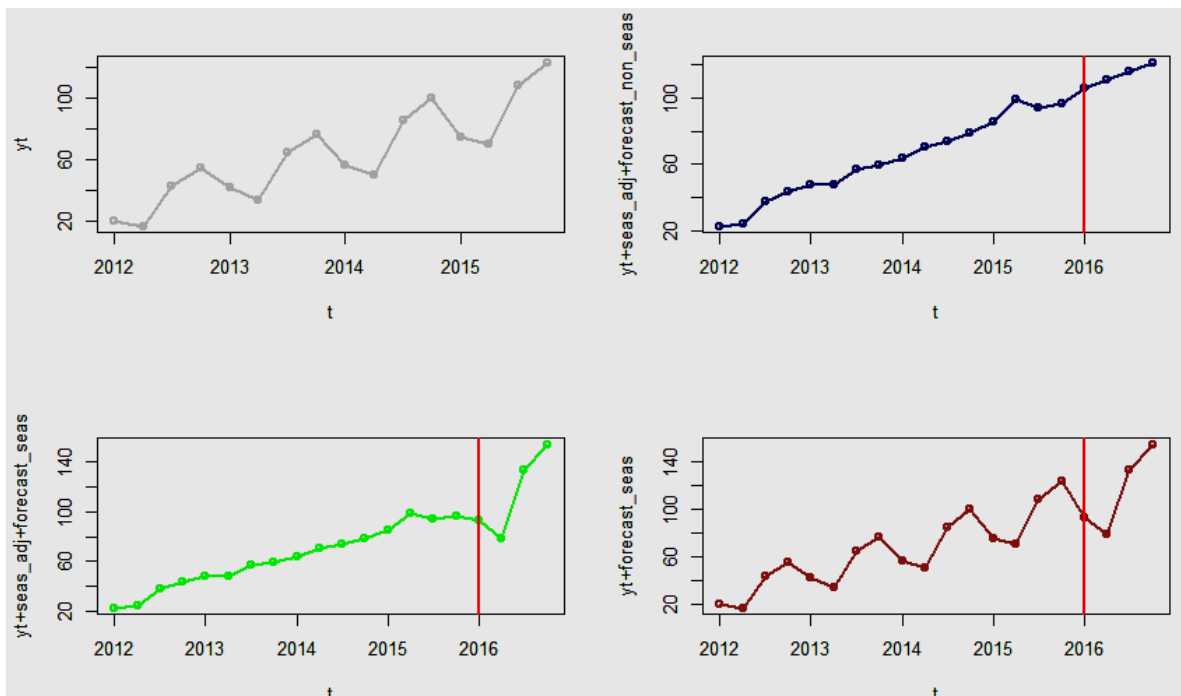
```
yfts<-ts(yf3,frequency=4,start=c(2012,1))
```

```
plot(yfts,type="o",lwd=2,col="green",xlab="t",ylab="yt+seas_ad  
j+forecast_seas");abline(v=2016,col="red",lwd=2)
```

```
yf4<-c(yts,sf)
```

```
yfts<-ts(yf4,frequency=4,start=c(2012,1))
```

```
plot(yfts,type="o",lwd=2,col="brown",xlab="t",ylab="yt+forecas  
t_seas");abline(v=2016,col="red",lwd=2)
```



5 ZÁVER

Základná funkcionálna jazyka R (knížnica `stats`) je dodávaná s množstvom funkcií užitočných pre analýzu časových radov. Ďalšie môže používateľ nájsť v doplnkových knižniciach (*add-on package*), ktoré sú vybavené pokročilými funkcionalitami pre túto oblasť, pričom tieto neobsahujú iba funkcie pre samotné časové rady, ale aj nástroje z hľadiska postavenia časových radov v ekonometrii, financiách i v aktuárstve, ktoré sa do značnej miery so svojimi nástrojmi pre časové rady so všeobecnou analýzou časových radov prekrývajú. Ide teda tak o klasické metódy, ako aj o veľmi pokročilé techniky.

Článok ponúka využitie jazyka R pri riešení fundamentálnych úloh (tu najmä sezónna dekompozícia) analýzy časových radov. Ďalej možno jazyk R využiť pre Boxovu-Jenkinsovu metodológiu, analýzu nestacionarity, pokročilé regresné metódy, štrukturálne modely, diagnostiku viacrozmerných časových radov, spektrálnu analýzu, ARIMA a GARCH modely a i. Pre využitie **ARIMA** modelov pozri R napr. funkcie `pacf`, `arima`, `tsdiag`, `predict`, pre spektrálnu analýzu funkciu `spectrum` a pre GARCH modely knižnicu `fGarch`.

Cieľom článku bolo predovšetkým upozorniť na obrovský potenciál, jednoduchosť aj praktické (grafické i programátorské) možnosti používania jazyka R pri modelovaní časových radov. Využitie špecializovaných knižníc R na metódy analýzy a prognózovania časových radov dovoľuje zvládať problémy, ktoré sú riešiteľné ostatnými komerčnými štatistickými softvérmi. Neustály vývoj nových knižníc, na ktorých vznikajú spravidla podieľajú významní odborníci aj skúsení praktici v tejto oblasti, zaisťuje možnosť prístupu k moderným metódam a technikám analýzy časových radov a ich rozšírenie celej komunite používateľov jazyka R. Jazyk R tak získava stále väčší význam nielen v komerčnej a akademickej sfére, resp. vo vedecko-výskumných pracoviskách, ale nadobúda na dôležitosť aj pri tvorbe výstupov štátnej štatistiky v národných štatistických úradoch. V súčasnej dobe sa jazyk R používa pri analýzach národných štatistických úradov Rakúska, Talianska, Holandska, Rumunska a čiastočne aj SR a iných štátov EÚ, ako aj v samotnom Eurostat-e (Pavelka, 2017).

Poznámka.

Jednou z popredných vlastností jazyka R je jednoduché **zdieľanie kódu** pre potreby ďalších používateľov (napr. v akademickej, resp. vedeckej sfére). Potom kód programu na výpočet sezónne zaťažených štvrťročných prognóz, ktoré sa priamo aj graficky zobrazia na grafe skutočného časového radu by bol nasledovný.

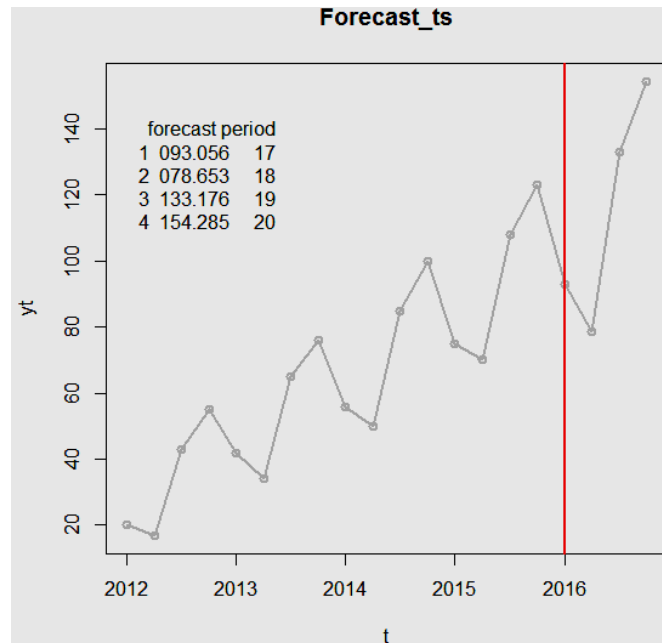
```
#### vstupné hodnoty
start_y<-2012; start_q<-1; s<-4
y<-c(20,17,43,55,42,34,65,76,56,50,85,100,75,70,108,123)
###
yts<-ts(y,frequency=s,start=c(start_y,start_q))
SD<-decompose(yts,type="multiplicative")
s_ind<-SD$figure
yadj<-y/rep(SD$figure,s)
x<-c(1:length(yadj))
model<-lm(yadj ~ x)
b0<-model$coefficients[1];b1<-model$coefficients[2]
xp<-c((length(x)+1):(length(x)+s))
sf<-(b0+b1*xp)*s_ind
yf<-c(yts,sf)
```



```

yf.ts<-ts(yf,frequency=s,start=c(start_y,start_q))
plot(yf.ts,type="o",lwd=2,col="grey",xlab="t",main="Forecast_ts",ylab="yt")
abline(v=start_y+s,col="red",lwd=2)
## kód pre vloženie výstupu do grafu (voliteľné)
df<-
data.frame("forecast"=sprintf("%07.3f",sf),"period"=c((length(x)+1):(length(x)+s)))
text(start_y,max(y),paste(capture.output(df),collapse='\n'),pos=4)

```



ZDROJE

- [1] PÁLEŠ, M.: *Aktuárstvo v režime Solventnosť II (S riešenými príkladmi v jazyku R)*. Bratislava : Vydavateľstvo EKONÓM, 2016.
- [2] PÁLEŠ, M.: *Jazyk R v aktuárskych analýzach*. Bratislava : Vydavateľstvo EKONÓM, 2017.
- [3] PAVELKA, R.: *Analýzy časových řad pomocí programovacího jazyka R*. In *Slovenská štatistika a demografia č. 2/2017*. Bratislava : Štatistický úrad Slovenskej republiky, 2017.
- [4] ŠOLTÉS, E. a kol.: *Štatistické metódy pre ekonómov (zbierka príkladov)*. Bratislava: Wolters Kluwer, 2015.
- [5] R CORE TEAM: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017 <<http://www.R-project.org/>>
- [6] <https://www.r-bloggers.com/>
- [7] <https://stackoverflow.com/questions/>

KONTAKTNÉ ÚDAJE

Páleš, Michal, Ing., PhD., Katedra matematiky a aktuárstva, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave, Dolnozemska cesta 1, 852 35 Bratislava, tel. +421 2/672 95 841, e-mail: pales.euba@gmail.com