

ARIMA modely – poznámky k softvéru R

Beáta Stehlíková, Časové rady FMFI UK, 2011

Budeme potrebovať:

- balík **tseries**,
- zo stránky www.stat.pitt.edu/stoffer/tsa3 súbor **tsa3.rda** - obsahuje užitočné funkcie pre prácu s ARIMA modelmi. Pri použití týchto funkcií treba uviesť referenciu, skripty nie sú súčasťou žiadneho balíka R.

Nastavíme pracovný adresár, načítame **tseries** a **tsa3.rda** (príkazom **load('tsa3.rda')**). Budeme pracovať s dátami krátkodobej a dlhodobej úrokovej miery z cvičenia. Dáta sú v súboroch **RSQ.txt** a **R20Q.txt**.

Načítanie dát

Začíname štandardne:

```
R Console
> rs=read.table('RSQ.txt')
> r20=read.table('R20Q.txt')
> spread=r20-rs
> |
```

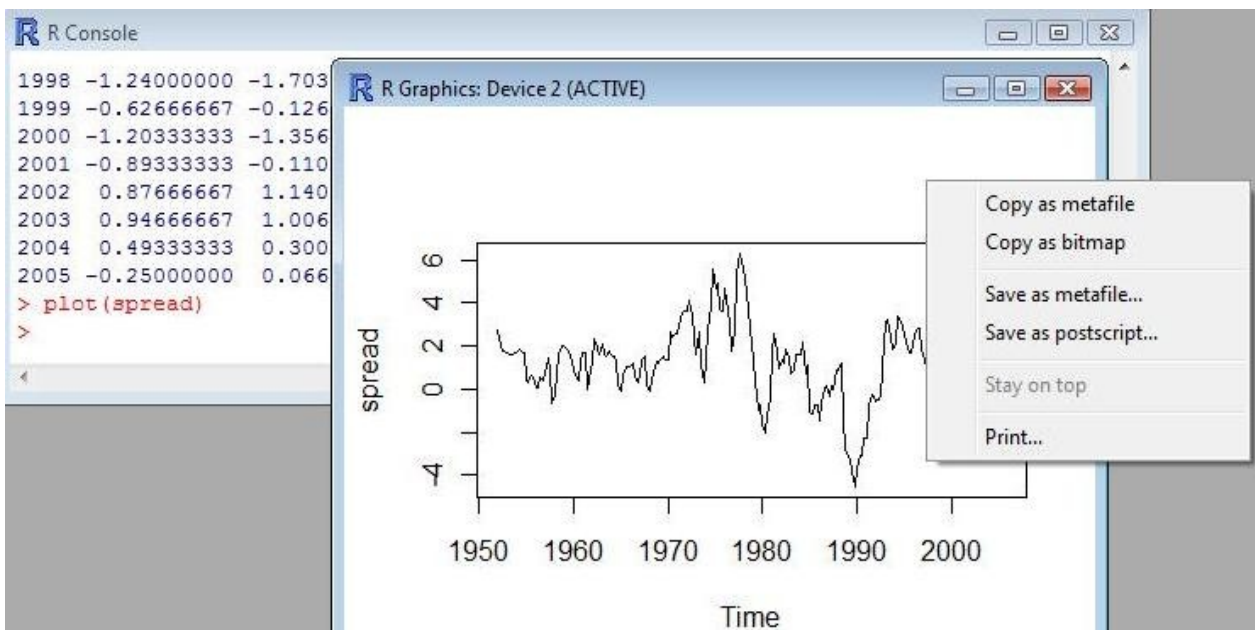
Teraz pridáme k číselným hodnotám štruktúru časového radu, ide o kvartálne dáta, prvé pozorovanie je z prvého kvartálu 1952:

```
> spread=ts(spread,frequency=4,start=c(1952,1),names='spread')
```

K dátam sa tak priradia časy:

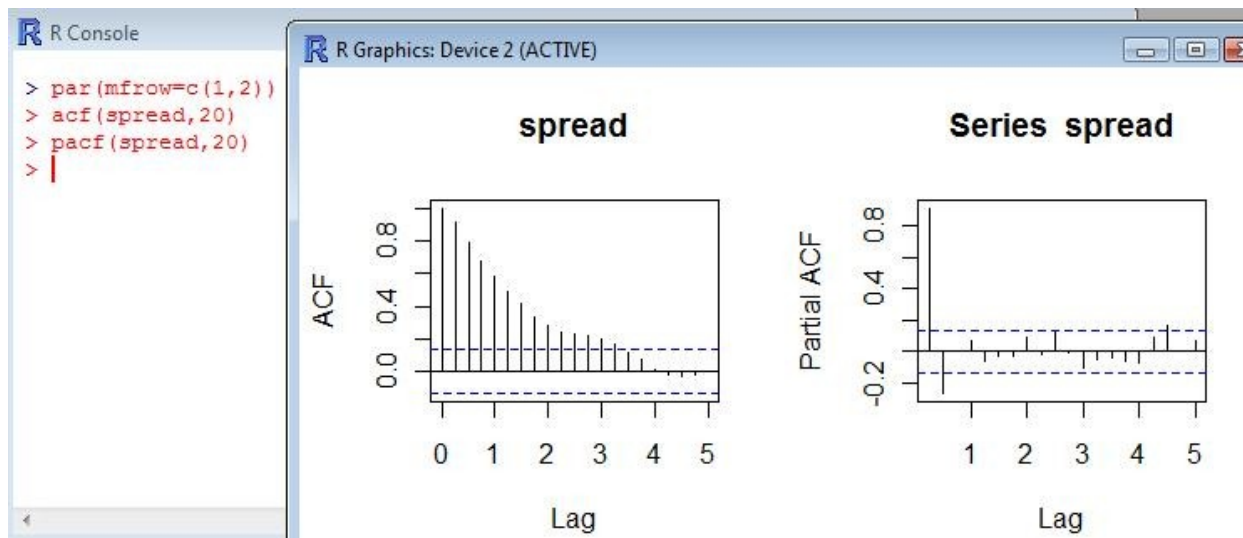
```
> spread
      Qtr1      Qtr2      Qtr3      Qtr4
1952  2.77065278  2.01327778  1.74994444  1.77894444
1953  1.69069444  1.59805556  1.62783333  1.69477778
1954  1.68829167  1.86754167  1.75465278  1.71620833
1955  0.45219444  0.29266667  0.64733333  0.53100000
1956  0.09200000  0.04033333  0.57500000  0.41700000
1957  0.70666667  1.46700000  0.81066667  0.63033333
```

Priebeh:

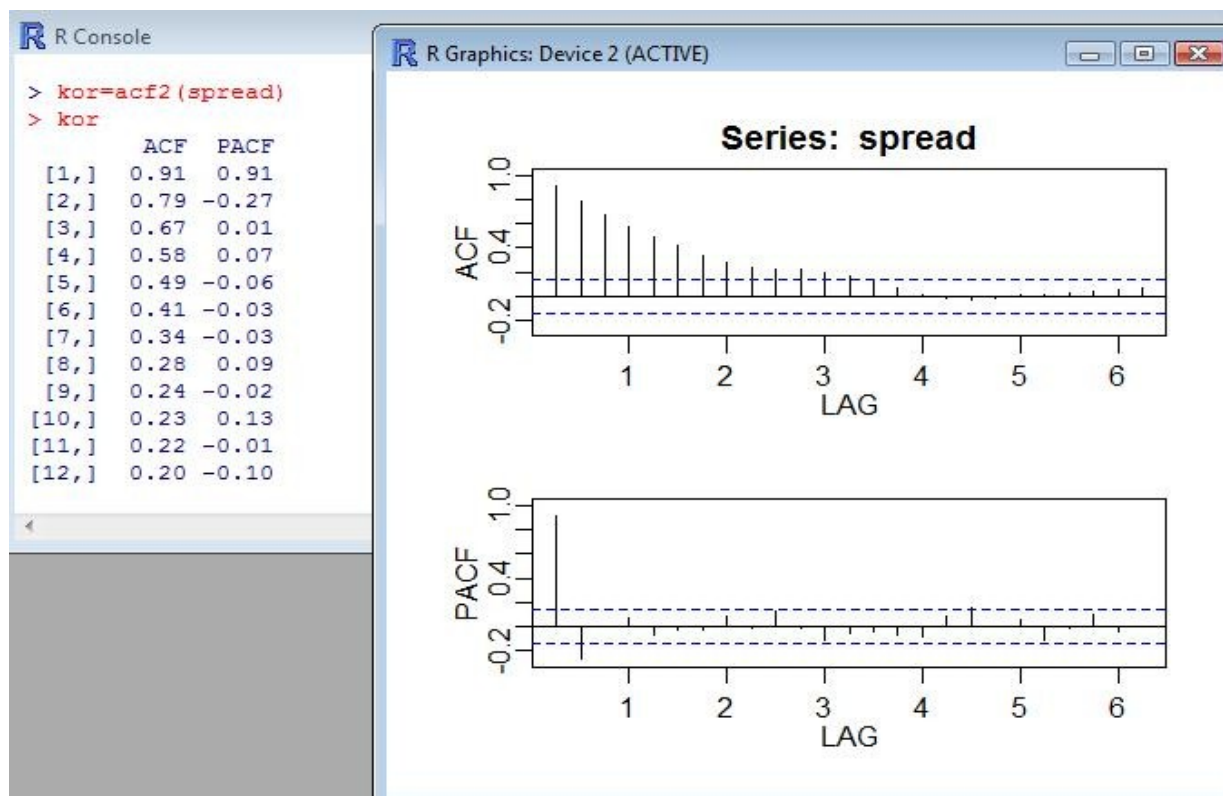


Autokorelačná a parciálna autokorelačná funkcia

Získame ich príkazmi `acf(spread)`, `pacf(spread)`, môžeme dodať ďalší parameter – koľko autokorelácií sa má vykresliť, napr. `acf(spread, 20)`. V jednom obrázku:



Lag na x-ovej osi je v rokoch, ak ho chceme mať spôsobom 1, 2, 3 ..., zadáme napríklad `acf(ts(spread,frequency=1),20)`. Výhrady D. Stoffera: ACF začína od lagu 0 (čo je zbytočné, lebo autokorelácia časového radu so sebou je vždy rovná 1) a PACF od 1, a majú rôzne mierky na y-ovej osi. Preto je v `tsa3.rda` funkcia `acf2`:



Odhadovanie a testovanie ARIMA modelu

V R sú určité problémy s ARIMA modelmi – kedy sa vypisuje stredná hodnota procesu a kedy intercept rovnice, zahrnutie konštanty do modelu pri diferencovaných dátach, P hodnota Ljung-Boxovej štatistiky (podrobnejšie na stránke D. Stoffera) – toto je vyriešené v `tsa3.rda`.

Model odhadujeme pomocou príkladu **sarima**. Zadávame názov premennej, a typ procesu: **p,k,q**, ak pre k-te diferencie odhadujeme ARMA(p,q) model. V našom prípade máme AR(2) model pre pôvodné dáta, parametrom **details=FALSE** zrušíme vypisovanie iterácií pri odhadovaní:

```
> sarima(spread,2,0,0,details=FALSE)
$fit

Call:
arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
      xreg = xmean, include.mean = FALSE, optim.control = list(trace = trc, REPORT = 1,
        reltol = tol))

Coefficients:
      ar1      ar2    xmean
  1.1809 -0.2886  1.0449
s.e.  0.0650  0.0651  0.4212

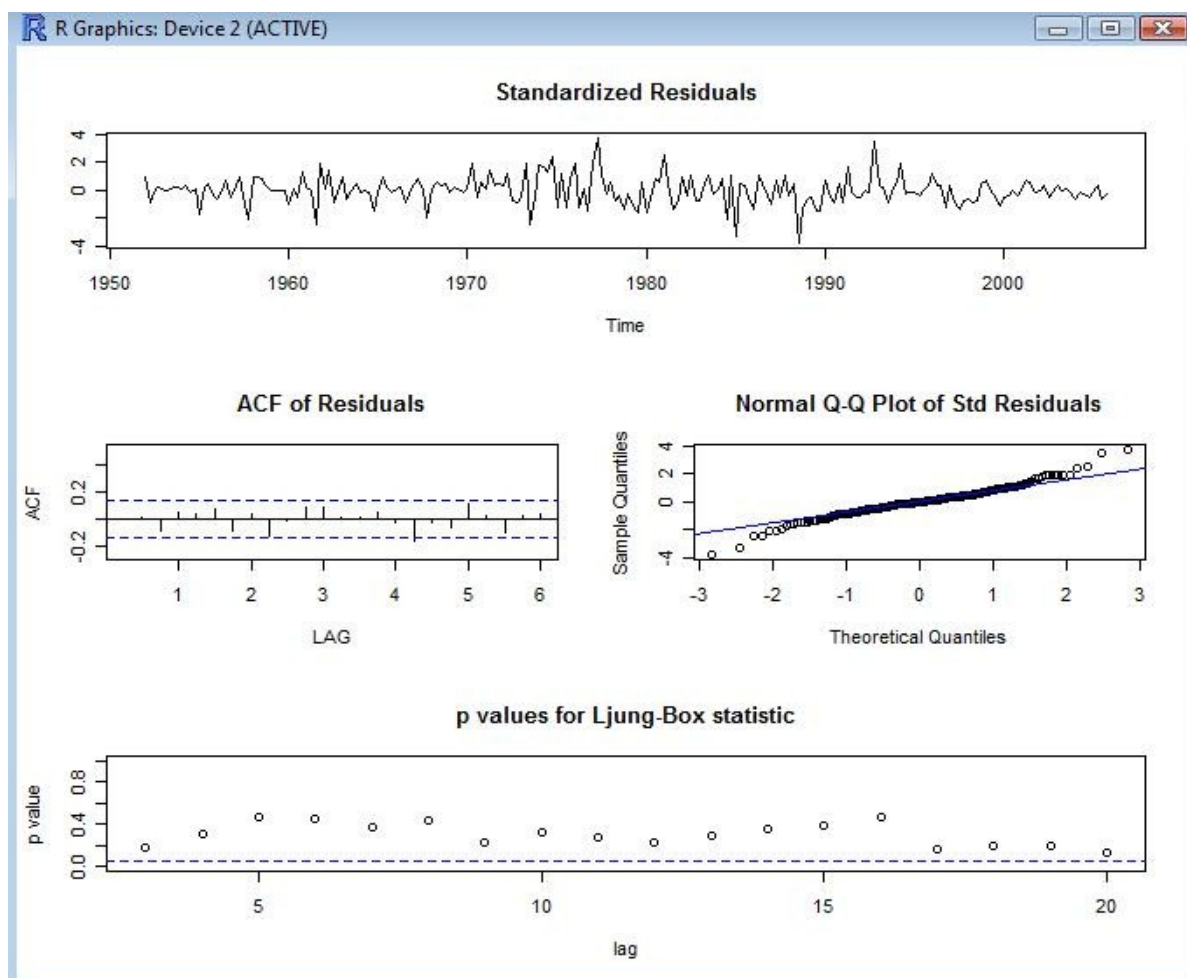
sigma^2 estimated as 0.4677:  log likelihood = -225.42,  aic = 458.84

$AIC
[1] 0.2678181

$AICc
[1] 0.277955

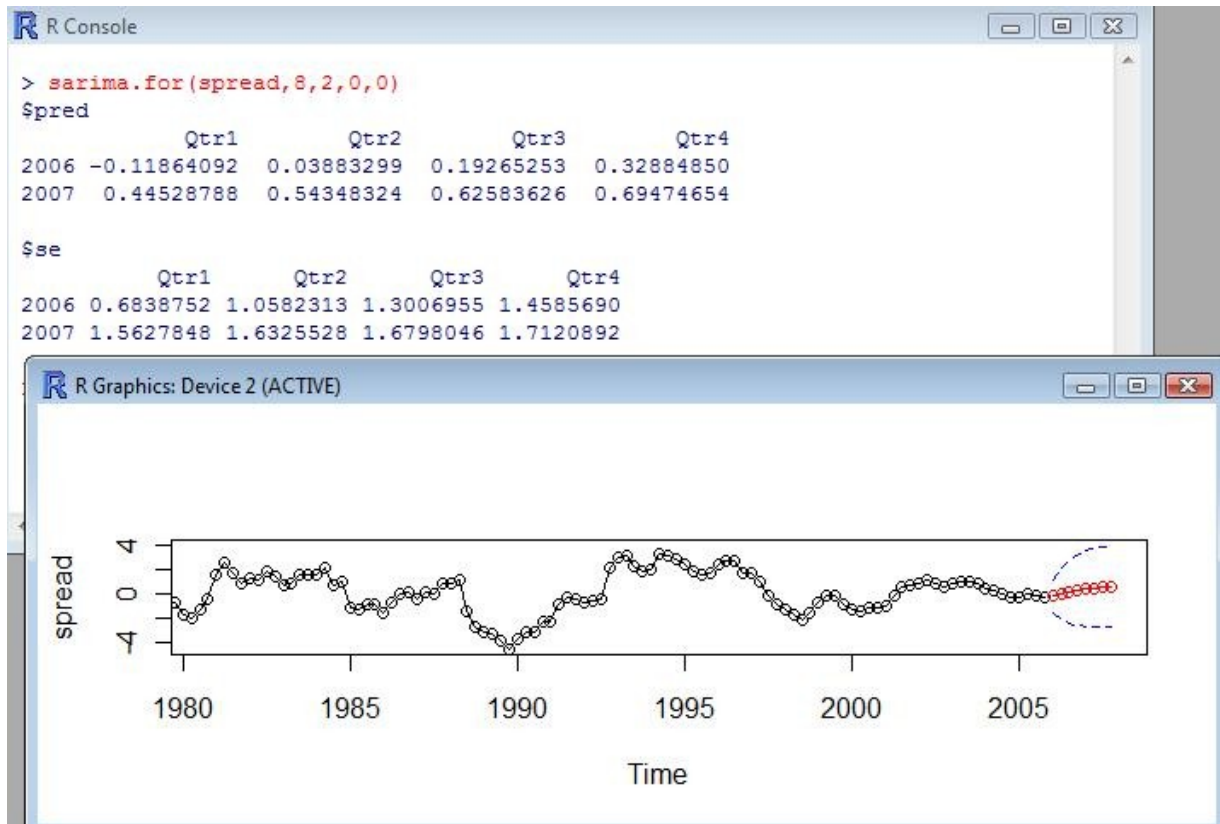
$BIC
[1] -0.6853031
```

Dostaneme odhady AR a MA členov; či odhadnutá konštanta predstavuje konštantu v definícii procesu alebo strednú hodnotu procesu vidíme z toho, či je označená ako **constant** (vtedy je to tá konštanta) alebo **mean** (vtedy je to stredná hodnota). Okrem tohto výstupu dostaneme nasledovné grafy:



Predikcie

Ak máme dobrý model, môžeme z neho robiť predikcie. V **tsa3.rda** je na to príkaz **sarima.for**. Parametrami je premenná, počet období, na ktoré robíme predikcie (v našom prípade na obrázku dolu 8, t. j. 8 kvartálov) a použitý model (**p,k,q** znovu znamená ARMA(p,q) model pre k-te diferencie, t. j. ARIMA(p,k,q) model).



Ak chceme zopakovať to, čo bolo na cvičení – vynechať niekoľko posledných dát, odhadnúť model bez nich a porovnať predikcie so skutočnými hodnotami:

