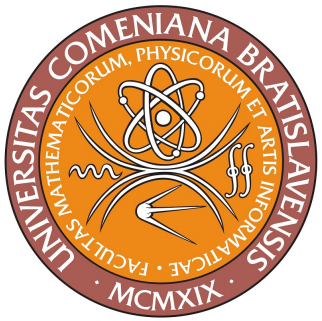
Predikovanie úmrtnostných mier pomocou balíka demography v softvéri R

Gábor Szűcs

e-mail: Gabor.Szucs@fmph.uniba.sk

Fakulta matematiky, fyziky a informatiky Univerzity Komenského v Bratislave



25. medzinárodný seminár

Výpočtová štatistika

(Bratislava, 1. decembra 2016)

Obsah prezentácie

1. Úvod
2. Leeov-Carterov model
3. Balík `demography` v softvéri R
4. Predikovanie úmrtnostných mier pomocou údajov Štatistického úradu SR
5. Záver

Motivácia

Motivácia vytvárania predikcií úmrtnostných mier

- Meniaca sa populácia, meniace sa miery úmrtnosti,
- predlžovanie dĺžky života ludí vo svete aj na Slovensku.

Motivácia

Motivácia vytvárania predikcií úmrtnostných mier

- Meniaca sa populácia, meniace sa miery úmrtnosti,
- predlžovanie dĺžky života ľudí vo svete aj na Slovensku.

Aplikácie v praxi

- ⋆ Pri oceňovaní dlhodobých životných poistení,
- ⋆ pri valuácii doživotných dôchodkov vyplácaných z úspor v starobnom dôchodkovom sporení na Slovensku.

Oceňovanie dôchodkov

Bežný výpočet súčasnej hodnoty polehotného dôchodku

$$a_{x:\overline{n}|} = \sum_{k=1}^{n} v^k {}_k p_x = v^1 {}_1 p_x + v^2 {}_2 p_x + v^3 {}_3 p_x + \cdots + v^n {}_n p_x =$$

$$= v^1 p_x + v^2 p_x \; p_{x+1} + v^3 p_x \; p_{x+1} \; p_{x+2} + \cdots + v^n p_x \; p_{x+1} \; p_{x+2} \cdots p_{x+n-1}$$

Oceňovanie dôchodkov

Bežný výpočet súčasnej hodnoty polehotného dôchodku

$$a_{x:\overline{30}|} = \sum_{k=1}^{30} v^k {}_k p_x = v^1 {}_1 p_x + v^2 {}_2 p_x + v^3 {}_3 p_x + \cdots + v^{30} {}_{30} p_x =$$

$$= v^1 p_x + v^2 p_x \; p_{x+1} + v^3 p_x \; p_{x+1} \; p_{x+2} + \cdots + v^{30} p_x \; p_{x+1} \; p_{x+2} \cdots p_{x+29}$$

Oceňovanie dôchodkov

Bežný výpočet súčasnej hodnoty polehotného dôchodku

$$a_{x:\overline{30}|} = \sum_{k=1}^{30} v^k {}_k p_x = v^1 {}_1 p_x + v^2 {}_2 p_x + v^3 {}_3 p_x + \cdots + v^{30} {}_{30} p_x =$$

$$= v^1 p_x + v^2 p_x \; p_{x+1} + v^3 p_x \; p_{x+1} \; p_{x+2} + \cdots + v^{30} p_x \; p_{x+1} \; p_{x+2} \cdots p_{x+29}$$

Dynamický výpočet súčasnej hodnoty polehotného dôchodku

$$a^{dyn}_{x:\overline{30}|} = v^1 p_x^{2016} + v^2 p_x^{2016} \; p_{x+1}^{2017} + v^3 p_x^{2016} \; p_{x+1}^{2017} \; p_{x+2}^{2018} + \cdots +$$

$$+ \cdots + v^{30} p_x^{2016} \; p_{x+1}^{2017} \; p_{x+2}^{2018} \cdots p_{x+29}^{2045}$$

Leeov-Carterov model

Jedným najznámejším dynamickým modelom úmrtnostných kriviek je tzv. **Leeov-Carterov model** (*Lee-Carter model*, LC model) publikovaný v článku:

Lee, R.D., Carter, L.R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, **87** (419, September).

Definícia

$$m_{x,t} = \exp\left\{a_x + b_x k_t + \varepsilon_{x,t}\right\}, \qquad x \in \mathcal{X},\ t \in \mathcal{T}, \tag{1}$$

$m_{x,t}$ = centrálna miera úmrtnosti osoby vo veku x v roku t,

a_x, b_x = vekovo špecifické parametre modelu (nezávislé od času),

k_t = časovo premenlivý dynamický faktor,

$\varepsilon_{x,t}$ = náhodný šum s nulovou strednou hodnotou a disperziou σ^2.

Leeov-Carterov model

Jedným najznámejším dynamickým modelom úmrtnostných kriviek je tzv. **Leeov-Carterov model** (*Lee-Carter model*, LC model) publikovaný v článku:

Lee, R.D., Carter, L.R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, **87** (419, September).

Definícia

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \qquad x \in \mathcal{X},\ t \in \mathcal{T}. \tag{2}$$

$m_{x,t}$ = centrálna miera úmrtnosti osoby vo veku x v roku t,

a_x, b_x = vekovo špecifické parametre modelu (nezávislé od času),

k_t = časovo premenlivý dynamický faktor,

$\varepsilon_{x,t}$ = náhodný šum s nulovou strednou hodnotou a disperziou σ^2.

Balík demography v softvéri R

```
library(demography)
```

Autor **Rob J Hyndman**

Hyndman, R.J. with contributions from Booth, H., Tickle, L. and Maindonald, J. (2014). *demography: Forecasting mortality, fertility, migration and population data*. R package version 1.18.

Adresy

https://cran.r-project.org/package=demography

https://cran.r-project.org/web/packages/demography/demography.pdf

Požiadavky balíka demography

Pri našich analýzach sme používali verziu 3.3.2 softvéru R.

Požadované balíky

`forecast`, `zoo`, `timeDate`, `rainbow`,
`MASS`, `pcaPP`, `ftsa`, `sde`, `stats4`,
`fda`, `splines`, `Matrix`, `cobs`

Balík `demography`, resp. balík `ftsa` vyžaduje používanie minimálne verzie 3.2.4 softvéru R.

Predikovanie úmrtnostných mier pomocou údajov ŠÚ SR

Zdroj údajov

Štatistický úrad Slovenskej republiky. (ŠÚ SR, online). *Obyvateľstvo a migrácia: Tabuľky života*, Úmrtnostné tabuľky (roky 1996-2015) pre Slovenskú republiku.

`https://slovak.statistics.sk/wps/wcm/connect/3e426ca5-9aca-4dc7-bb33-92726b6cd3b3/Umrtnostne_tabulky_SR_1996_2015.zip?MOD=AJPERES`

Príprava podkladových dát

1. krok

Pripraviť podkladové dáta do vhodnej podoby a načítať ich do softvéru R

```
# úmrtnostné miery: mx = Dx/Px
    Y1996           Y1997           Y1998           ... Y2015
62 0.0193709549222 0.0180029738076 0.0186474217204 ... 0.0143242925912
63 0.0206199311187 0.0214274131900 0.0213327109347 ... 0.0159887304102
64 0.0215138742536 0.0200018142231 0.0215529478565 ... 0.0173032347600
... ...            ...             ...             ... ...
99 0.1145374449339 0.1066666666667 0.0759162303665 ... 0.2076271186440

svk.mx <- read.table("D://data//ut62//svkmx62.txt", header=TRUE)
```

Príprava podkladových dát

1. krok

Pripraviť podkladové dáta do vhodnej podoby a načítať ich do softvéru R

```
# celková populácia
   Y1996 Y1997 Y1998 ... Y2014 Y2015
62 44035 43715 42955 ... 72047 72395
63 45005 43169 42845 ... 68812 70988
64 45552 44096 42268 ... 63464 67733
...   ...   ...   ... ...   ...   ...
99    227   300   382 ...   244   236

svk.pop <-read.table("D://data//ut62//popul62.txt", header=TRUE)
```

Vytvorenie objektu demogdata

2. krok

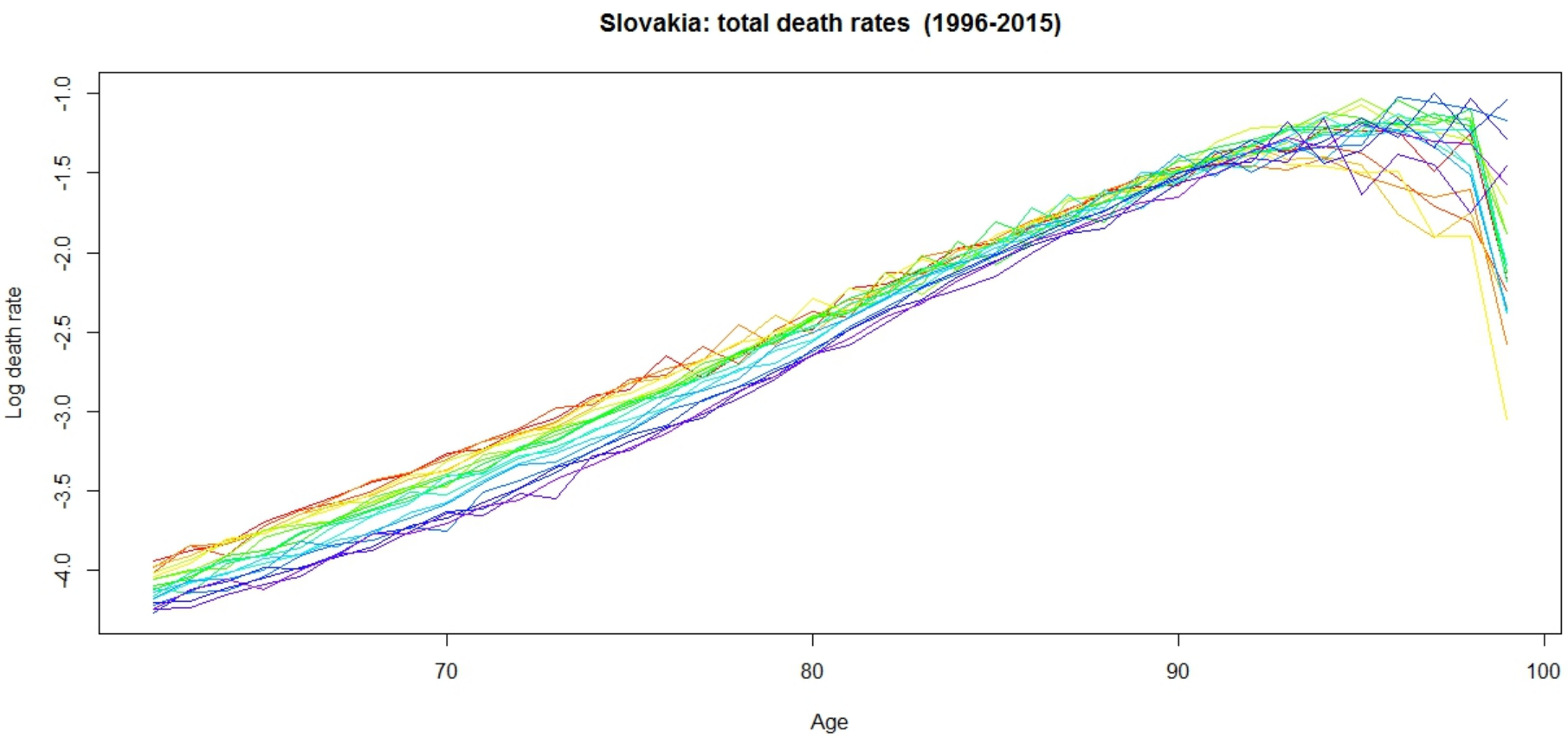
Načítať knižnicu demography a vytvoriť objekt demogdata

```
# načítanie knižnice 'demography' library(demography)

# vytvorenie objektu typu 'demogdata'
SVK.data <- demogdata(data=svk.mx, pop=svk.pop, ages=62:99,
years=1996:2015, type="mortality", label="Slovakia",
name="total", lambda=0)

# grafické znázornenie logaritmických mier úmrtnosti
plot(SVK.data, ages=62:99, years=1996:2015)
```

Logaritmické miery úmrtnosti (bez vyhladzovania)



Vyhladzovanie úmrtnostných kriviek

3. krok

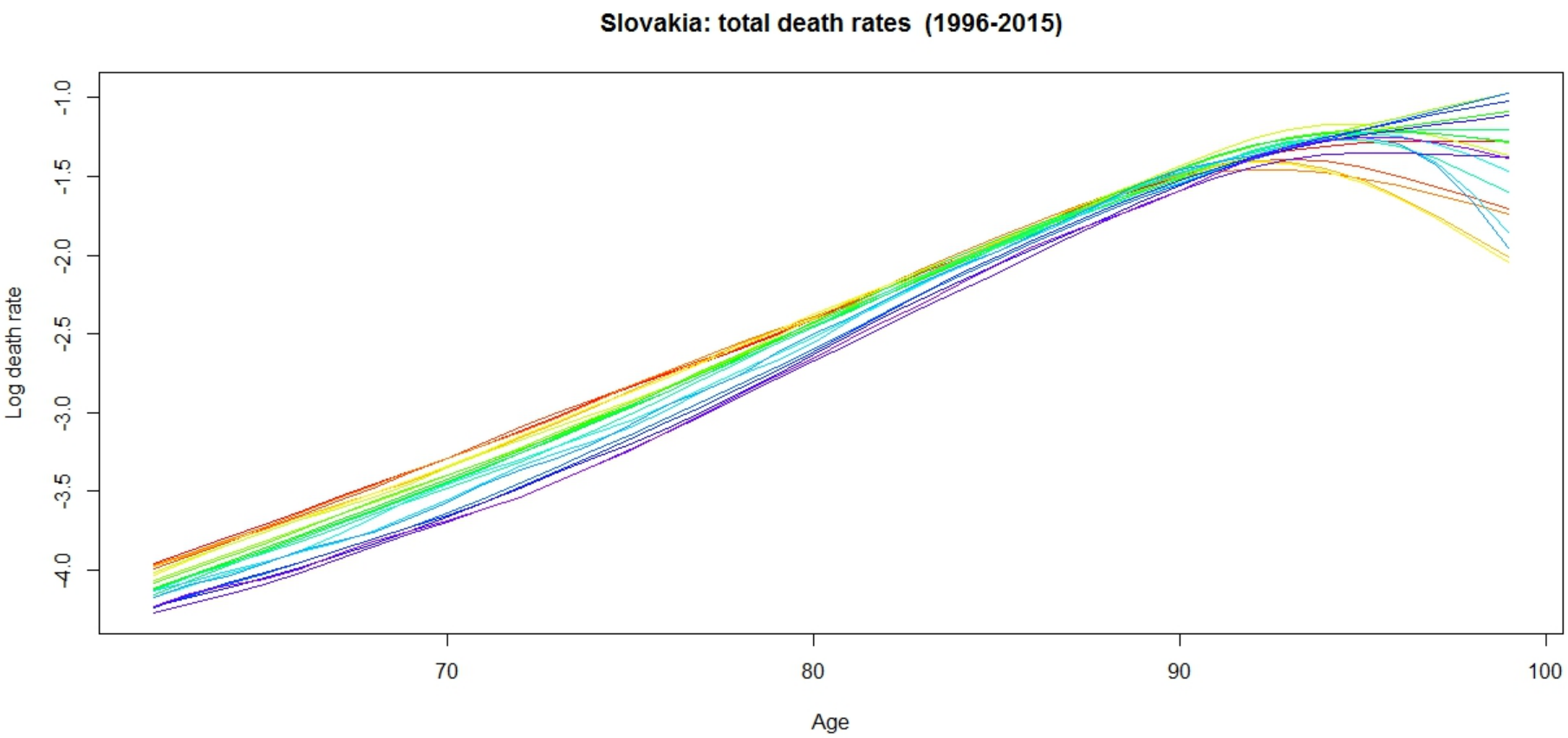
```
# vyhladzovanie úmrtnostných kriviek

SVK.data.sm <- smooth.demogdata(data=SVK.data, method="spline")


# grafické znázornenie vyhladzovaných log-mier úmrtnosti

plot(SVK.data.sm, ages=62:99, years=1996:2015)
```

Vyhladzované logaritmické miery úmrtnosti



Konštrukcia LC modelu pre nevyhladzované dáta

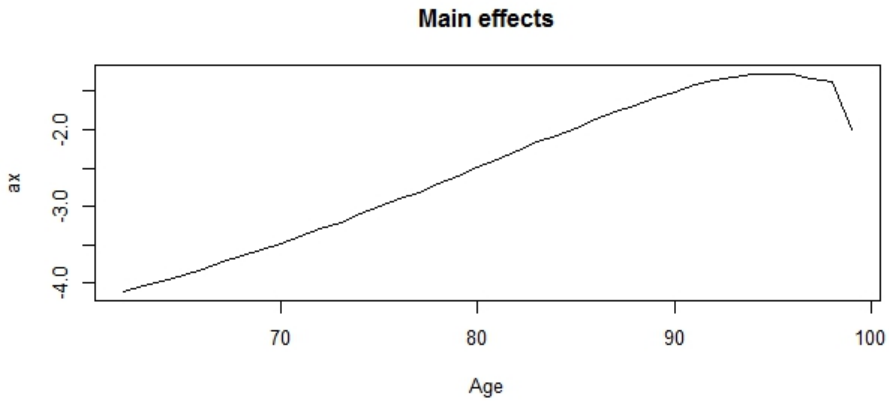
4. krok

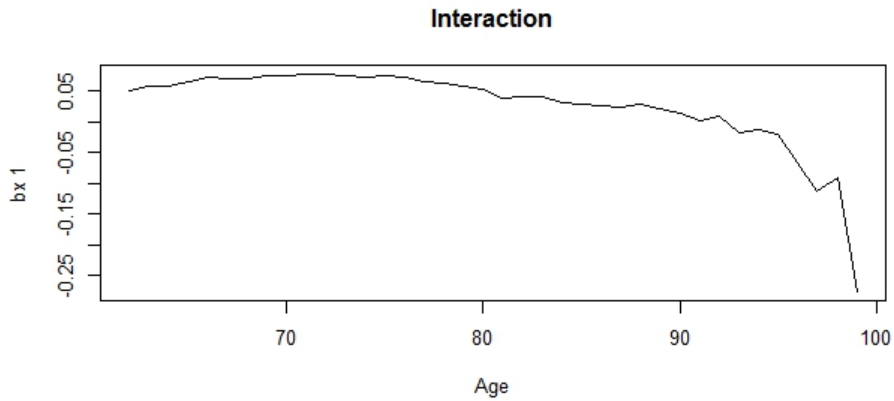
```
# Konštrukcia Leeovho-Carterovho modelu pre dáta ŠÚ SR
# za roky 1996:2015, pre veky=62:99, pre celkovú populáciu
SVK.lca <- lca(data=SVK.data, series="total")

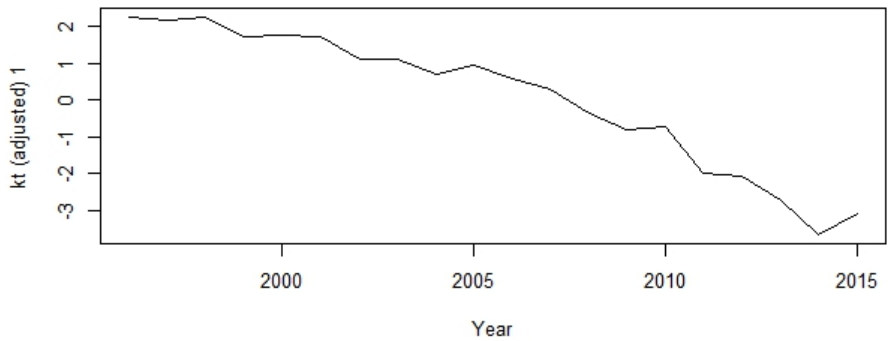
# grafické znázornenie funkcií Leeovho-Carterovho modelu
plot(SVK.lca)

# grafické znázornenie fitovaných log-mier pomocou LC modelu
plot(SVK.lca$fitted)
```

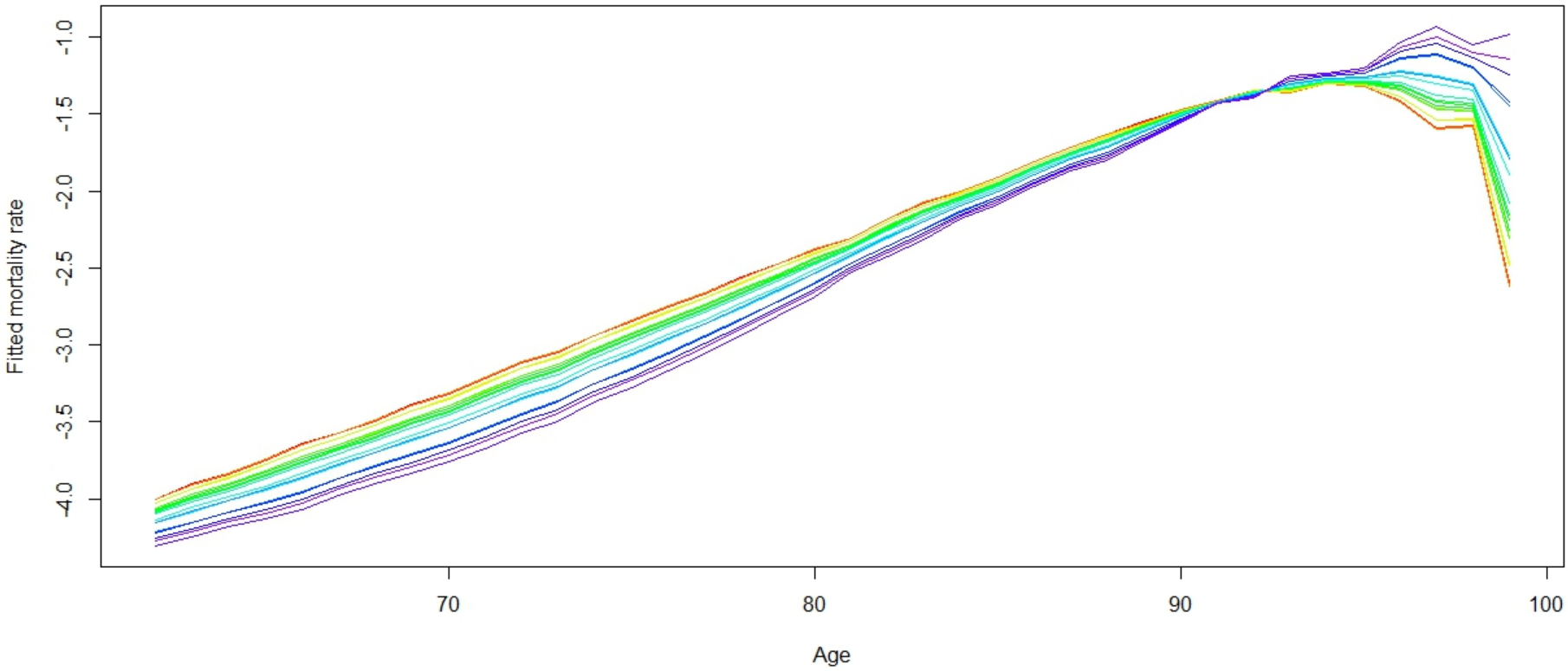
Parametre LC modelu: a(x), b(x), k(t)







Fitované log-miery pomocou nevyhladzovaného LC modelu



Konštrukcia LC modelu pre vyhladzované dáta

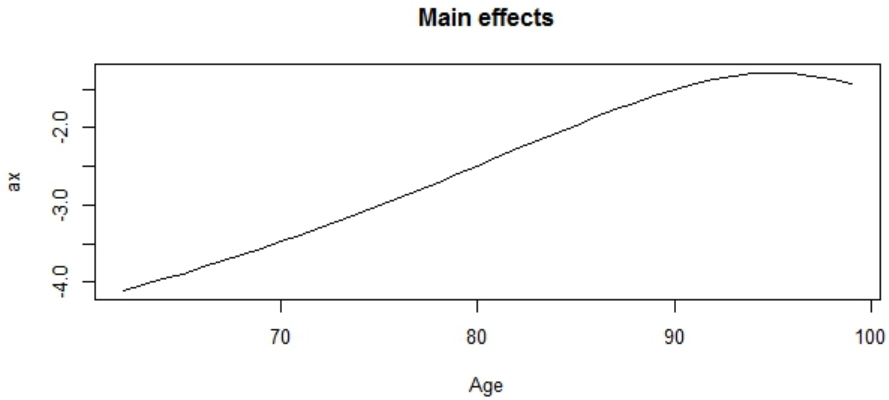
5. krok

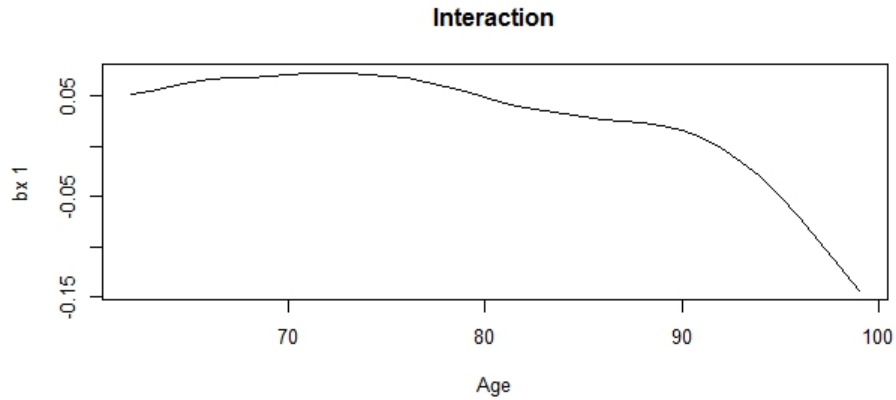
```
# Konštrukcia Leeovho-Carterovho modelu
# pre vyhladzované dáta ŠÚ SR
# za roky 1996:2015, pre veky=62:99, pre celkovú populáciu
SVK.lca.sm <- lca(data=SVK.data.sm, series="total")

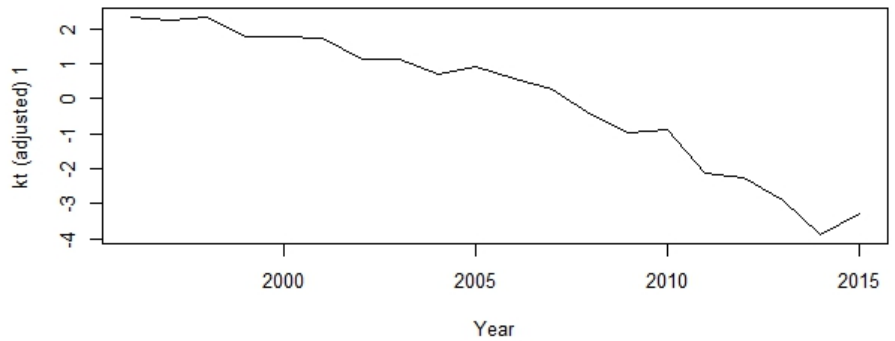
# grafické znázornenie funkcií Leeovho-Carterovho modelu
plot(SVK.lca.sm)

# grafické znázornenie fitovaných log-mier pomocou LC modelu
plot(SVK.lca.sm$fitted)
```

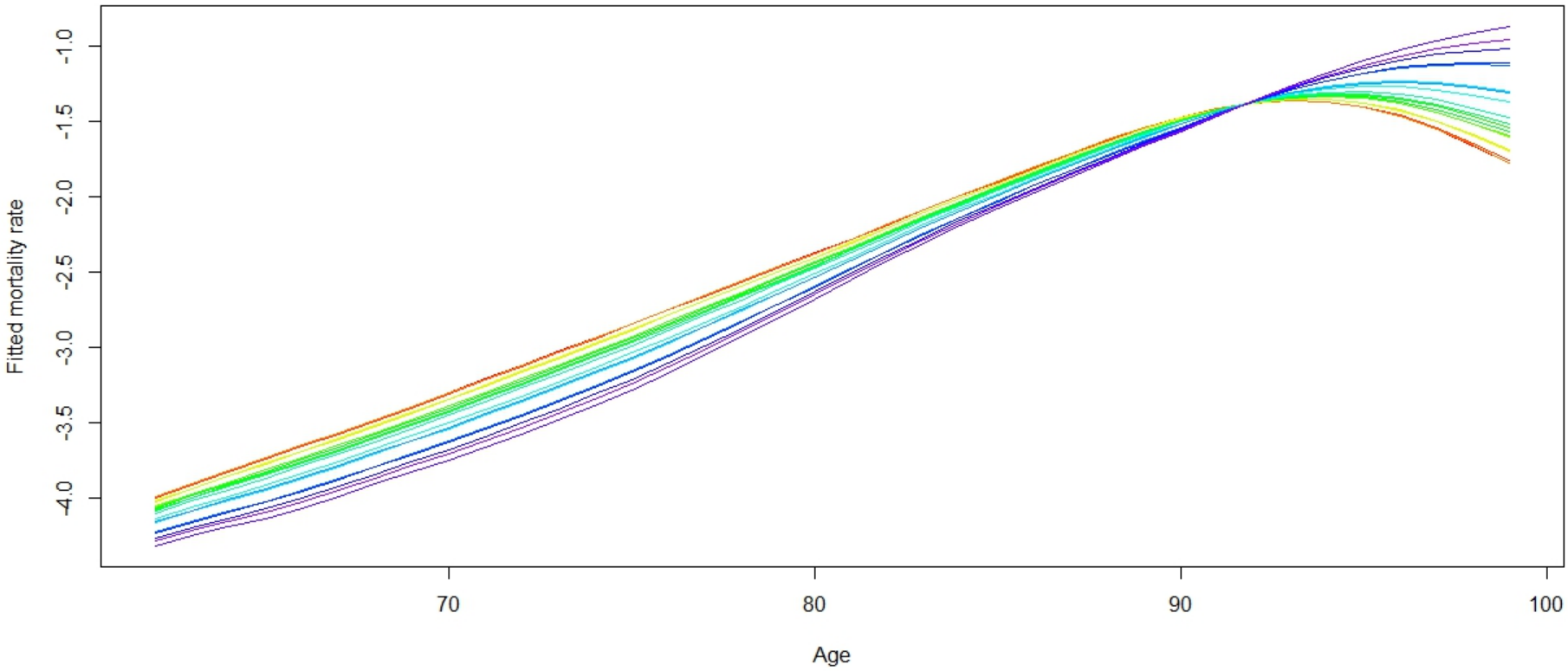
Parametre vyhladzovaného LC modelu: a(x), b(x), k(t)







Fitované log-miery pomocou vyhladzovaného LC modelu



Predikcie pomocou LC modelu pre nevyhladzované dáta

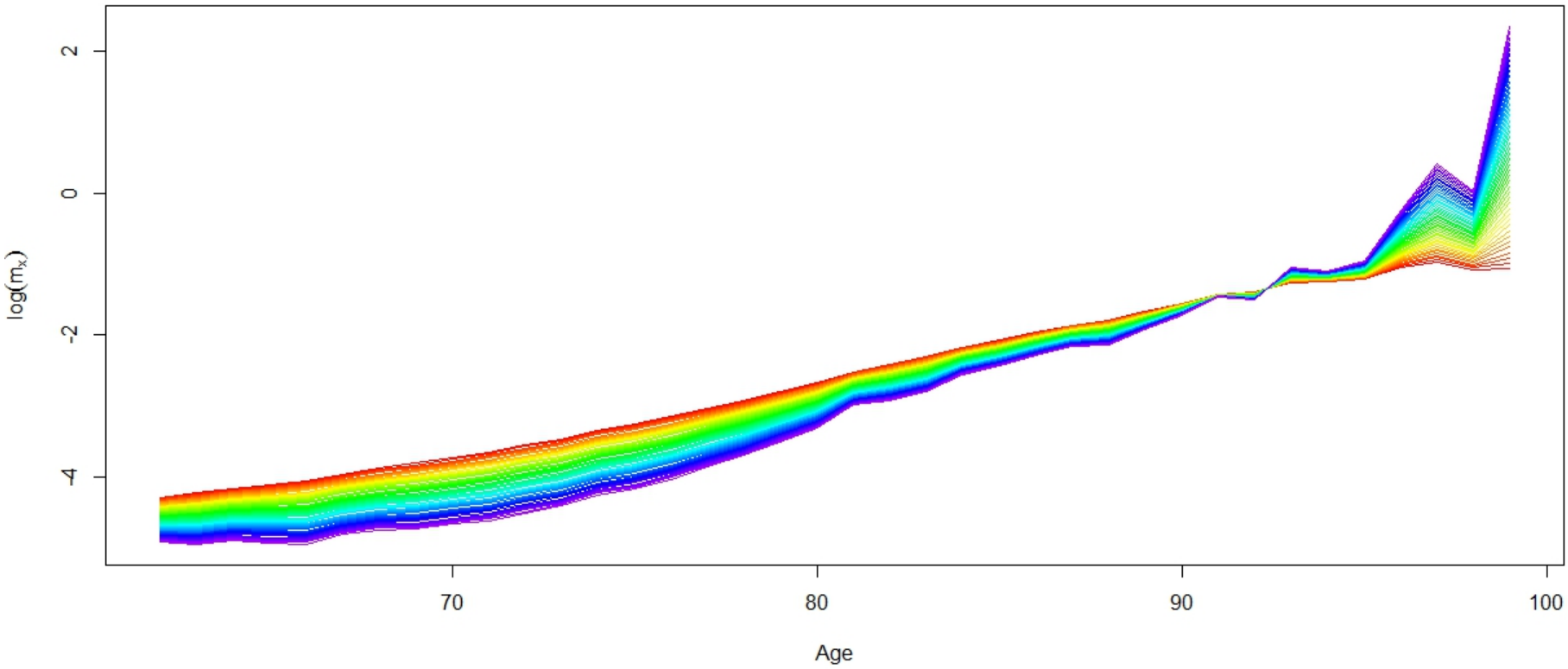
6. krok

```
# predikcie log-mier úmrtnosti pre obdobie 2016-2060
# (nevyhladzované dáta)
SVK.pred <-forecast(SVK.lca, h=45, level=90)

# grafické znázornenie predikovaných log-úmrtnostných kriviek
plot(SVK.pred, xlab="roky", ylab=expression(log(m[x])))

# predikované úmrtnostné miery a interval spoľahlivosti
SVK.pred$rate
SVK.pred$rate$lower
SVK.pred$rate$upper
```

Predikované log-miery pre nevyhladzované dáta



Predikcie pomocou LC modelu pre vyhladzované dáta

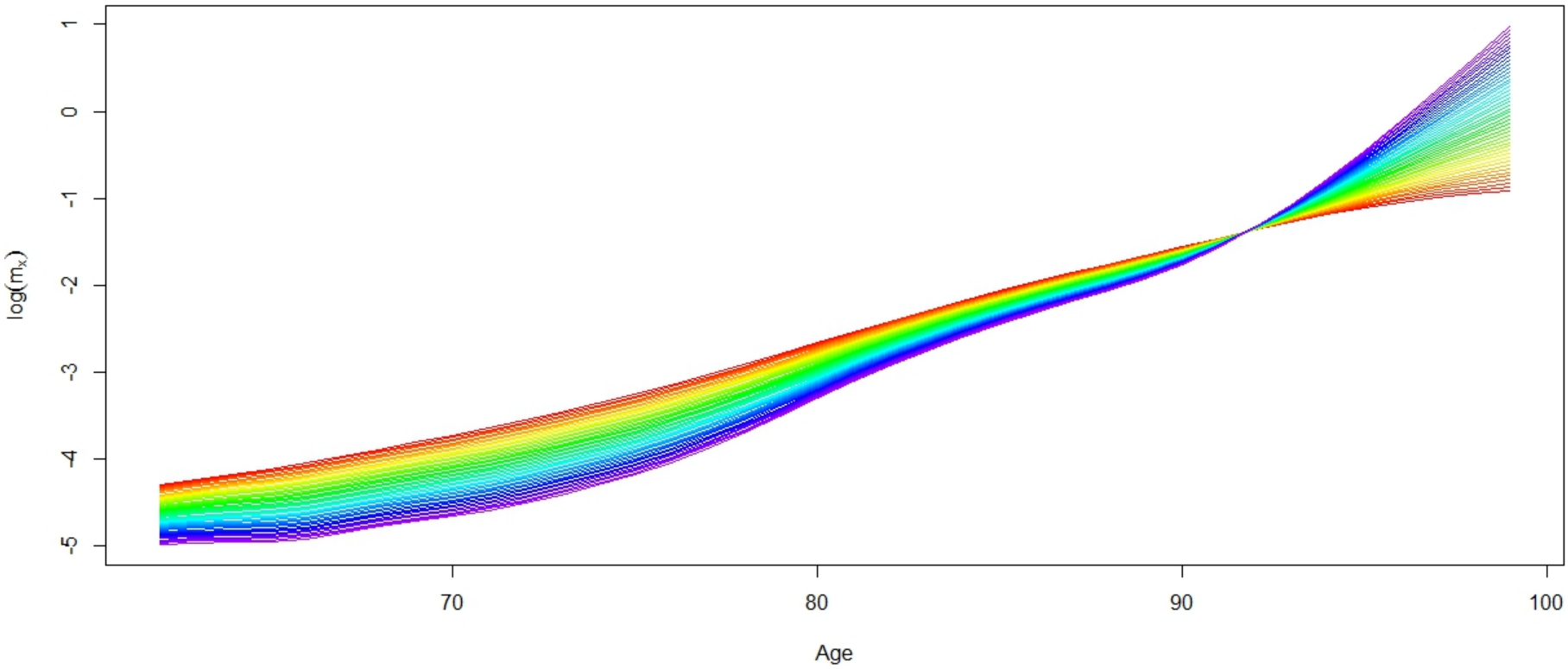
7. krok

```
# predikcie log-mier úmrtnosti pre obdobie 2016-2060
# (vyhladzované dáta)
SVK.pred.sm <-forecast(SVK.lca.sm, h=45, level=90)

# grafické znázornenie predikovaných log-úmrtnostných kriviek
plot(SVK.pred.sm, xlab="roky", ylab=expression(log(m[x])))

# predikované úmrtnostné miery a interval spoľahlivosti
SVK.pred.sm$rate
SVK.pred.sm$rate$lower
SVK.pred.sm$rate$upper
```

Predikované log-miery pre vyhladzované dáta



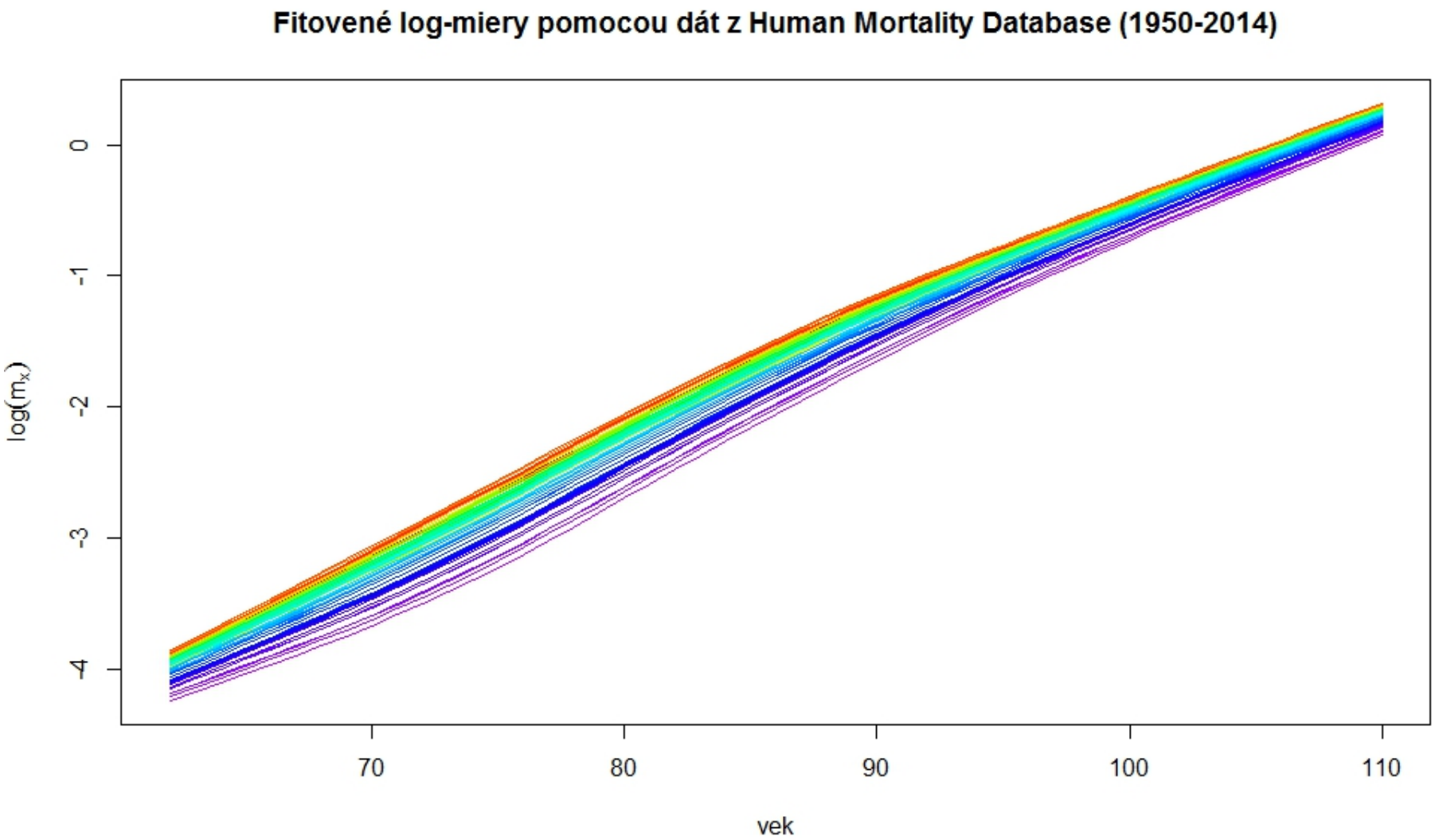
Záver

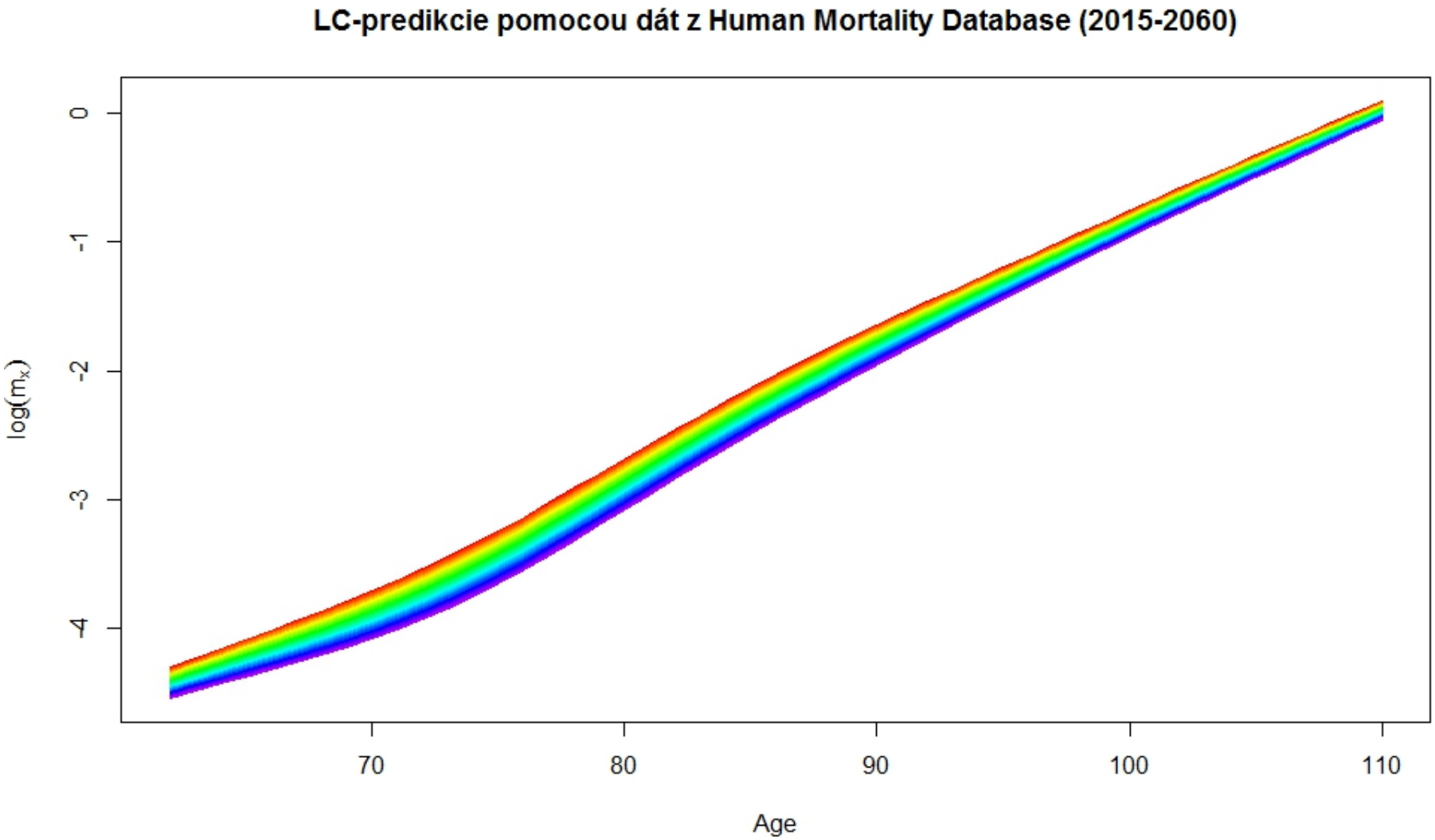
Čo ponúka balík `demography`?

+ Odhad parametrov Leeovho-Carterovho modelu,
+ vyhladzovanie a fitovanie úmrtnostných kriviek,
+ vytvorenie predikcií pre budúce úmrtnostné miery,
+ konštrukciu intervalov spoľahlivosti pre predikcie,
+ predpovedanie strednej budúcej dĺžky života.

− citlivosť na vstupné dáta,

− nesprávne zachytené trendy vo vývoji úmrtnostných mier.


Fitovené log-miery pomocou dát z Human Mortality Database (1950-2014)
log(m_x)
0
-1
-2
-3
-4
70
80
90
100
110
vek


LC-predikcie pomocou dát z Human Mortality Database (2015-2060)
log(m_x)
0
-1
-2
-3
-4
70
80
90
100
110
Age

Ďakujem vám za pozornosť.

Tento výskum bol podporený grantmi VEGA 2/0047/15, VEGA 1/0251/16 a APVV-0465-12.