

CoDa R Package

Marius Pascariu

2018-07-19

Description

The CoDa R package contains the implementation of the Compositional Data Mortality Model (CoDa). This is a Lee-Carter (1992) type method that is used to modelling and forecasting the life table distribution of deaths (dx) using Principal Component Analysis. In the context of mortality forecasting the CoDa method was first used in Bergeron-Boucher et al. (2017). The package includes functions for fitting the model, analysing its goodness-of-fit and performing mortality projections.

Installation

1. Make sure you have the most recent version of R
2. Run the following code in your R console

```
install.packages("CoDa")
```

Updating to the latest version of the package

You can track and contribute to the development of CoDa on GitHub. To install it:

1. Install the release version of devtools from CRAN with `install.packages("devtools")`.
2. Make sure you have a working development environment.
 - **Windows:** Install Rtools.
 - **Mac:** Install Xcode from the Mac App Store.
 - **Linux:** Install a compiler and various development libraries (details vary across different flavors of Linux).
3. Install the development version of CoDa.

```
devtools::install_github("mpascariu/CoDa")
```

Help

All functions are documented in the standard way, which means that once you load the package using `library(CoDa)` you can just type `?coda` to see the help file.

Examples

```
library(CoDa)
```

Fit CoDa model

The model can be fitted using function `coda` and using a dataset containing mortality data (dx distributions) in for of a matrix or data.frame with ages as row and time as column. `CoDa.data` is an example of such a data set.

```
CoDa.data[1:5, 1:5]
```

```
##           1960           1961           1962           1963           1964
## 0 0.063768765 0.062408776 0.058475531 0.055569918 0.054402102
## 1 0.004484745 0.004205882 0.004092247 0.003889991 0.003846039
## 2 0.002571836 0.002508007 0.002415691 0.002317961 0.002175572
## 3 0.001976766 0.001851735 0.001814390 0.001787156 0.001714041
## 4 0.001615381 0.001571251 0.001416598 0.001463843 0.001364048
```

`CoDa.data` is containing distribution of deaths for US female population between 1960 and 2014. The data is provided in the package for testing purposes only. By the time you are using it, it may be outdated. Download actual demographic data free of charge from Human Mortality Database. Once a username and a password is created on the website the MortalityLaws R package can be used to extract data directly into your R console.

```
M <- coda(dx = CoDa.data, x = 0:110, y = 1960:2014)
M
```

```
##
## Compositional Data Mortality Model fit - CoDa (Oeppen 2008)
## Model with predictor:  $\text{clr } d[x] = a[x] + b[x]k[t]$ 
## Call: coda(dx = CoDa.data, x = 0:110, y = 1960:2014)
## Ages in fit: 0 - 110
## Years in fit: 1960 - 2014
```

Output objects

The output is an object of class `coda` with the components:

- `input` – List with arguments provided in input. Saved for convenience;
- `coefficients` – Estimated coefficients;
- `fitted` – Fitted values of the estimated CoDa model;
- `residuals` – Deviance residuals;
- `x` – Vector of ages used in the fitting;
- `y` – Vector of years used in the fitting;
- `call` – The unevaluated expression of the defined `coda` function.

```
ls(M)
```

```
## [1] "call"           "coefficients"  "fitted"       "input"
## [5] "residuals"     "x"             "y"
```

Summary

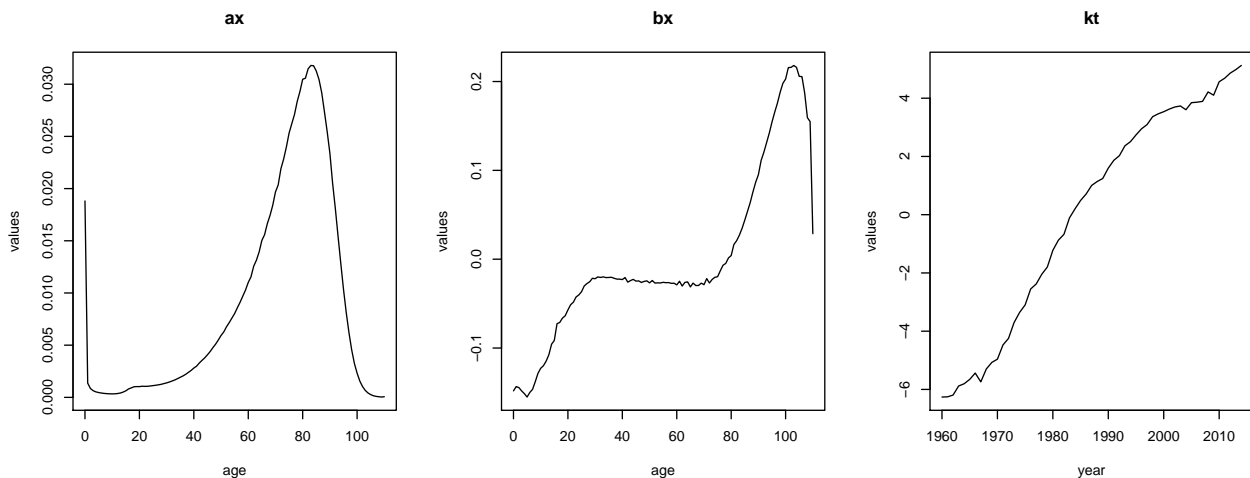
```
summary(M)
```

```
##
## Compositional Data Model fit - CoDa (Oeppen 2008)
## Model with predictor:  $\text{clr } d[x] = a[x] + b[x]k[t]$ 
##
## Coefficients:
##           ax      bx      .      y      kt
## 0  0.01882 -0.14822 | 1960 -6.2615
## 1  0.00136 -0.14334 | 1961 -6.25469
## 2  0.00085 -0.14457 | 1962 -6.19413
## 3  0.00064 -0.14831 | 1963 -5.88283
## 4  0.00053 -0.15134 | 1964 -5.80534
## 5  0.00046 -0.15515 | 1965 -5.65574
## ...      ...      ... <NA> ...      ...
## 105 0.00026 0.20582 | 2009 4.10337
## 106 0.00015 0.20556 | 2010 4.56493
## 107 9e-05 0.18714 | 2011 4.68676
## 108 6e-05 0.15935 | 2012 4.86257
## 109 3e-05 0.1549 | 2013 4.98146
## 110 6e-05 0.0287 | 2014 5.12442
```

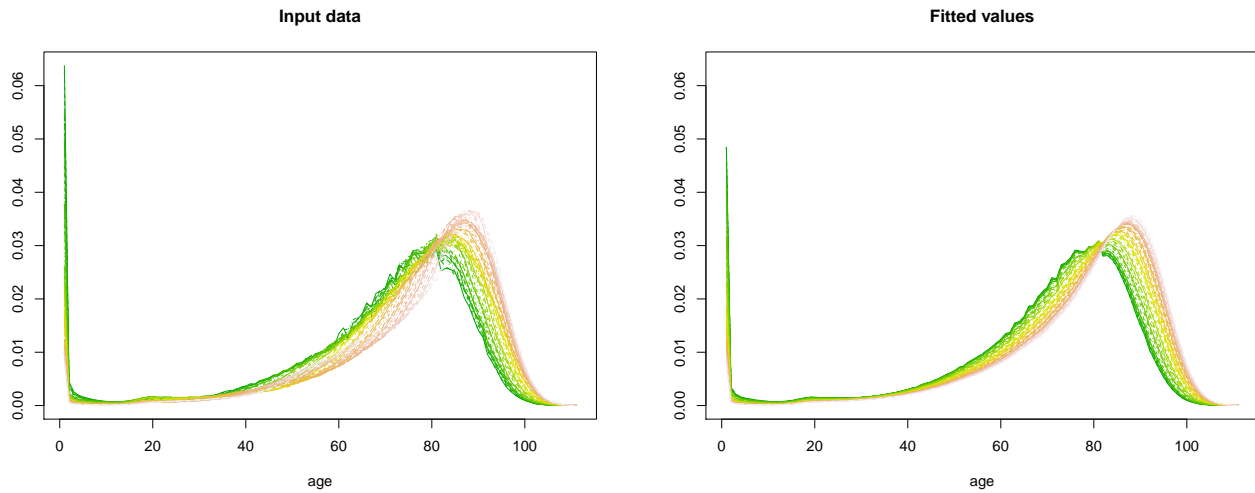
How to plot fitted parameters and fitted values of a CoDa mortality model

Two types of plots are available: "coef" to obtain representations of the three estimated series of parameters and "data" for visualising the input and fitted values.

```
plot(M, plotType = "coef", ylab = "values")
```



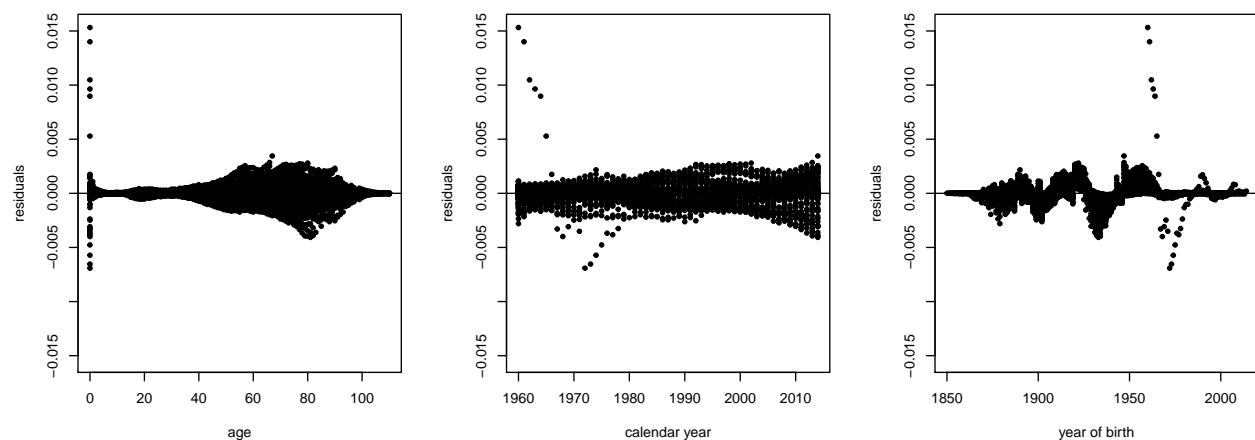
```
plot(M, plotType = "data")
```



Plot Residuals

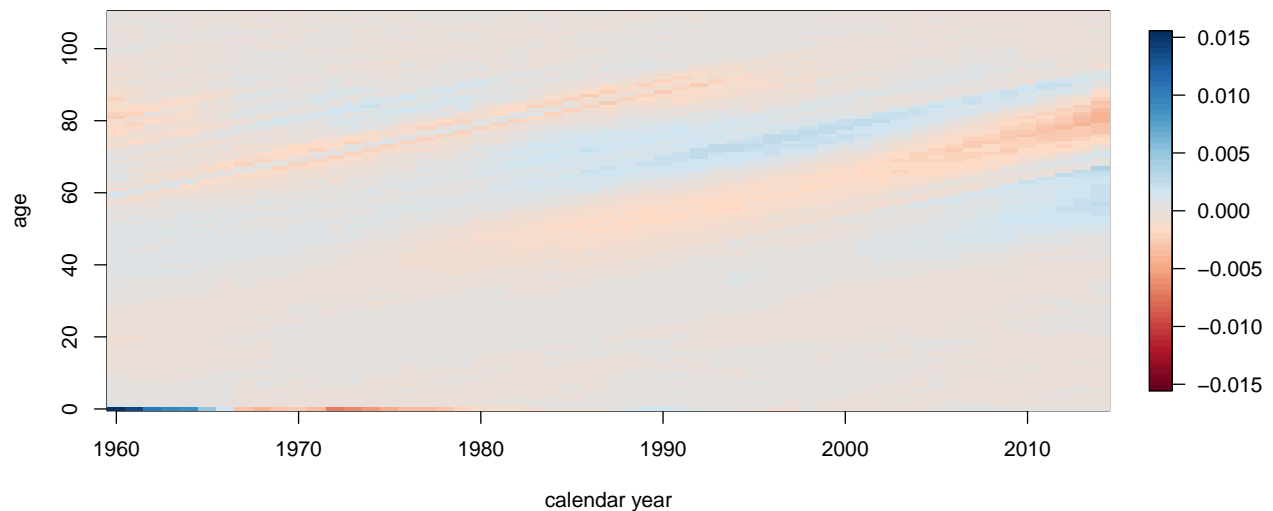
From the resulted deviance residuals, `resid(M)`, three types of figures can be obtained. When `plotType = "scatter"` scatter plots of the residuals against age, calendar year and cohort (year of birth) are produced.

```
plot(resid(M), plotType = "scatter")
```



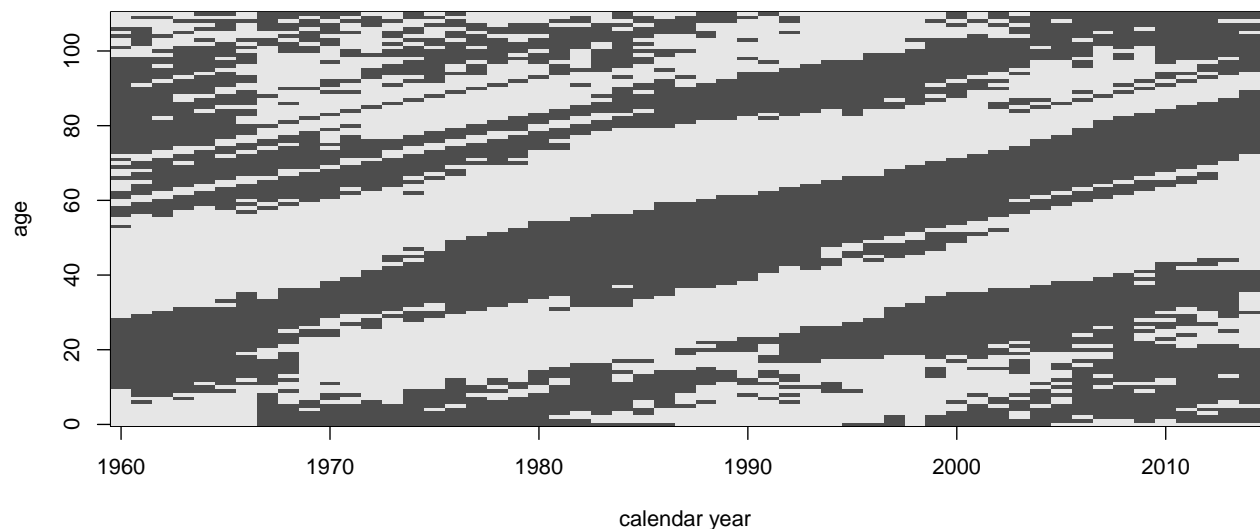
When `plotType = "colourmap"` a two dimensional colour map of the residuals is plotted. This is produced using function `image.plot`. See `image.plot` for further parameters that can be passed to this type of plots.

```
plot(resid(M), plotType = "colourmap")
```



When `plotType = "signplot"` a two dimensional black and white map of the residuals is plotted with dark grey representing negative residuals and light grey representing positive residuals. This is produced using function `image.default`.

```
plot(resid(M), plotType = "signplot")
```



Mortality projections

Mortality projections can be obtained using function `predict`. The example below shows how a 30 year mortality forecast is realised using the fitted `coda` model. For the computation of the `jumpchoice` there are two alternatives: `actual` (uses actual rates from final year) and `fit` (uses fitted rates).

```
P <- predict(M, h = 30, jumpchoice = 'actual')
P

##
## Compositional Data Mortality Model forecast
## Time series model (kt): ARIMA(2,2,2)
## Ages in forecast: 0 - 110
## Years in forecast: 2015 - 2044
```

```
# list of objects in predict
ls(P)
```

```
## [1] "kt"          "predicted.values" "ts.model"
## [4] "x"           "y"
```

```
# Predicted distribution of death
head(P$predicted.values$mean, 3)
```

```
##           2015           2016           2017           2018           2019
## 0 0.0075851408 0.0075122888 0.0071610721 0.0070100479 0.0067843320
## 1 0.0004882349 0.0004836694 0.0004616400 0.0004521574 0.0004379734
## 2 0.0003116650 0.0003087307 0.0002945754 0.0002884838 0.0002793739
##           2020           2021           2022           2023           2024
## 0 0.0065498611 0.0063901695 0.0061528080 0.0059875747 0.0057904885
## 1 0.0004232242 0.0004131699 0.0003982116 0.0003877887 0.0003753454
## 2 0.0002699034 0.0002634490 0.0002538487 0.0002471609 0.0002391785
##           2025           2026           2027           2028           2029
## 0 0.0056057935 0.0054392728 0.0052553604 0.0050972852 0.0049295726
## 1 0.0003636732 0.0003531400 0.0003414958 0.0003314780 0.0003208397
## 2 0.0002316926 0.0002249387 0.0002174742 0.0002110538 0.0002042374
##           2030           2031           2032           2033           2034
## 0 0.0047720332 0.0046201851 0.0044673466 0.0043248650 0.0041815764
## 1 0.0003108373 0.0003011874 0.0002914654 0.0002823939 0.0002732625
## 2 0.0001978299 0.0001916496 0.0001854247 0.0001796176 0.0001737735
##           2035           2036           2037           2038           2039
## 0 0.0040449341 0.0039117279 0.0037813077 0.0036562608 0.0035333042
## 1 0.0002645465 0.0002560419 0.0002477074 0.0002397089 0.0002318368
## 2 0.0001681966 0.0001627563 0.0001574260 0.0001523117 0.0001472795
##           2040           2041           2042           2043           2044
## 0 0.0034149202 0.0032996082 0.0031875662 0.0030792190 0.0029736429
## 1 0.0002242505 0.0002168542 0.0002096610 0.0002026986 0.0001959080
## 2 0.0001424311 0.0001377052 0.0001331102 0.0001286637 0.0001243278
```

References

1. Bergeron-Boucher, M-P., Canudas-Romo, V., Oeppen, J. and Vaupel, W.J. 2017. Coherent forecasts of mortality with compositional data analysis. Demographic Research, Volume 17, Article 17, Pages 527–566.
2. Oeppen, J. 2008. Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data. Paper presented at the European Population Conference 2008, Barcelona, Spain, July 9-12, 2008.
3. Aitchison, J. 1986. The Statistical Analysis of Compositional Data. London: Chapman and Hall. 2015.
4. Ronald D. Lee and Lawrence R. Carter. 1992. Modeling and Forecasting U.S. Mortality, Journal of the American Statistical Association, 87:419, 659–671.
5. Wikipedia. Compositional data