

Package ‘CvmortalityMult’

March 27, 2024

Type Package

Title Cross-Validation for Multi-Population Mortality Models

Version 0.0.1

Date 2024-03-20

Description Implementation of cross-validation method for testing the forecasting accuracy of several multi-population mortality models. The family of multi-population includes several multi-population mortality models proposed through the actuarial and demography literature. The package includes functions for fitting and forecast the mortality rates of several populations. Additionally, we include functions for testing the forecasting accuracy of different multi-population models.

References.

Atance, D., Debon, A., and Navarro, E. (2020) <[doi:10.3390/math8091550](https://doi.org/10.3390/math8091550)>.

Bergmeir, C. & Benitez, J.M. (2012) <[doi:10.1016/j.ins.2011.12.028](https://doi.org/10.1016/j.ins.2011.12.028)>.

Debon, A., Montes, F., & Martinez-Ruiz, F. (2011) <[doi:10.1007/s13385-011-0043-z](https://doi.org/10.1007/s13385-011-0043-z)>.

Lee, R.D. & Carter, L.R. (1992) <[doi:10.1080/01621459.1992.10475265](https://doi.org/10.1080/01621459.1992.10475265)>.

Russolillo, M., Giordano, G., & Haberman, S. (2011) <[doi:10.1080/03461231003611933](https://doi.org/10.1080/03461231003611933)>.

Santolino, M. (2023) <[doi:10.3390/risks11100170](https://doi.org/10.3390/risks11100170)>.

License MIT + file LICENSE

URL <https://github.com/davidAtance/CvmortalityMult>

BugReports <https://github.com/davidAtance/CvmortalityMult/issues>

LazyLoad yes

LazyData true

RoxygenNote 7.2.3

Depends R (>= 3.50)

Imports StMoMo, forecast, gnm, tmap, sf

Encoding UTF-8

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author David Atance [aut, cre] (<<https://orcid.org/0000-0001-5860-0584>>),

Ana Debón [aut] (<<https://orcid.org/0000-0002-5116-289X>>)

Maintainer David Atance <david.atance@uah.es>

Repository CRAN

Date/Publication 2024-03-27 09:40:05 UTC

R topics documented:

fit_additive.LC.multi	2
fit_multiplicative.LC.multi	4
for_additive.LC.multi	6
for_multiplicative.LC.multi	8
MAE	11
MAPE	12
MSE	13
multipopulation_cv	15
multipopulation_loocv	18
plotLC.multi	21
regions	22
SpainMap	23
SpainNat	24
SpainRegions	25
SSE	27
Index	29

fit_additive.LC.multi *FUNCTION TO FIT ADDITIVE MULTI-POPULATION MORTALITY MODEL*

Description

R function for fitting additive multi-population mortality model developed by: Debon et al. (2011). This model follows the structure of the well-known Lee-Carter model (Lee and Carter, 1992) but including an additive parameter to capture the behavior of each population considered. This parameter seeks to capture the individual behavior of every population considered. It should be mentioned that in case that this function is developed for fitting several populations. However, in case you only consider one population, the function will fit the one-population Lee-Carter model.

Usage

```
fit_additive.LC.multi(qxt, periods, ages, nPop, lxt = NULL)
```

Arguments

<code>qxt</code>	mortality rates used to fit the additive multipopulation mortality model. This rates can be provided in matrix or in <code>data.frame</code> .
<code>periods</code>	periods considered in the fitting in a vector way <code>c(minyear:maxyear)</code> .
<code>ages</code>	vector with the ages considered in the fitting. If the mortality rates provide from an abridged life tables, it is necessary to provide a vector with the ages, see the example.
<code>nPop</code>	number of population considered for fitting.
<code>lxt</code>	survivor function considered for every population, not necessary to provide.

Value

A list with different components of the fitting process:

- `ax` parameter that captures the average shape of the mortality curve in all considered populations.
- `bx` parameter that explains the age effect x with respect to the general trend `kt` in the mortality rates of all considered populations.
- `kt` represent the national tendency of multi-mortality populations during the period.
- `Ii` gives an idea of the differences in the pattern of mortality in any region i with respect to Region 1.
- `formula` additive multi-population mortality formula used to fit the mortality rates.
- `data` used mortality rates used to fit the data.
- `qxt.real` real mortality rates.
- `qxt.fitted` fitted mortality rates using the additive multi-population mortality model.
- `logit.qxt.fitted` fitted mortality rates in logit way.
- `Ages` provided ages to fit the data.
- `Periods` provided periods to fit the periods.
- `nPop` provided number of populations to fit the periods.

References

- Debon, A., Montes, F., and Martinez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. *European Actuarial Journal*, 1, 291-308.
- Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659–671.

See Also

[fit_multiplicative.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [multipopulation_cv](#), [multipopulation_loocv](#), [plotLC.multi](#)

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#The data that we are going to use:
SpainRegions
library(gnm)
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
#In this case, we fit for males providing the lxt
additive_Spainmales <- fit_additive.LC.multi(qxt = SpainRegions$qx_male,
      periods = c(1991:2020),
      ages = c(ages),
      nPop = 18,
      lxt = SpainRegions$lx_male)
#Once, we have fit the data, it is possible to see the ax, bx, kt, and Ii
#provided parameters for the fitting.
plotLC.multi(additive_Spainmales)

#As we mentioned in the details of the function, if we only provide the data
#from one-population the function fit_additive.LC.multi()
#will fit the Lee-Carter model for single populations.
LC_Spainmales <- fit_additive.LC.multi(qxt = SpainNat$qx_male,
      periods = c(1991:2020),
      ages = ages,
      nPop = 1)

plotLC.multi(LC_Spainmales)
```

```
fit_multiplicative.LC.multi
```

*FUNCTION TO FIT MULTIPLICATIVE MULTI-POPULATION
MORTALITY MODEL*

Description

R function for fitting multiplicative multi-population mortality model developed by: Russolillo et al. (2011). This model follows the structure of the well-known Lee-Carter model (Lee and Carter, 1992) but including a multiplicative parameter to capture the behavior of each population considered. This parameter seeks to capture the individual behavior of every population considered. It should be mentioned that in case that this function is developed for fitting several populations. However, in case you only consider one population, the function will fit the one-population Lee-Carter model.

Usage

```
fit_multiplicative.LC.multi(qxt, periods, ages, nPop, lxt = NULL)
```

Arguments

qxt	mortality rates used to fit the additive multipopulation mortality model. This rates can be provided in matrix or in data.frame.
periods	periods considered in the fitting in a vector way c(minyear:maxyear).
ages	vector with the ages considered in the fitting. If the mortality rates provide from an abridged life tables, it is required to provide a vector with the ages, see the example.
nPop	number of population considered for fitting.
lxt	survivor function considered for every population, not necessary to provide.

Value

A list with different components of the fitting process:

- ax parameter that captures the average shape of the mortality curve in all considered populations.
- bx parameter that explains the age effect x with respect to the general trend kt in the mortality rates of all considered populations.
- kt represent the national tendency of multi-mortality populations during the period.
- Ii gives an idea of the differences in the pattern of mortality in any region i with respect to Region 1.
- formula additive multi-population mortality formula used to fit the mortality rates.
- data . used mortality rates used to fit the data.
- qxt . real real mortality rates.
- qxt . fitted fitted mortality rates using the additive multi-population mortality model.
- logit . qxt . fitted fitted mortality rates in logit way.
- Ages provided ages to fit the data.
- Periods provided periods to fit the periods.
- nPop provided number of populations to fit the periods.

References

- Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659–671.
- Russolillo, M., Giordano, G., & Haberman, S. (2011). Extending the Lee–Carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, 2011(2), 96-117.

See Also

[fit_additive.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [multipopulation_cv](#), [multipopulation_loocv](#)

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#The data that we are going to use:
SpainRegions
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
#In this case, we fit for males providing the lxt
library(gnm)
multiplicative_Spainmales <- fit_multiplicative.LC.multi(qxt = SpainRegions$qx_male,
  periods = c(1991:2020),
  ages = c(ages),
  nPop = 18,
  lxt = SpainRegions$lx_male)

#Once, we have fit the data, it is possible to see the ax, bx, kt, and It
#provided parameters for the fitting.
plotLC.multi(multiplicative_Spainmales)

#As we mentioned in the details of the function, if we only provide the data
#from one-population the function fit_multiplicative.LC.multi()
#will fit the Lee-Carter model for single populations.
LC_Spainmales <- fit_multiplicative.LC.multi(qxt = SpainNat$qx_male,
  periods = c(1991:2020),
  ages = ages,
  nPop = 1)

plotLC.multi(LC_Spainmales)
```

for_additive.LC.multi *FUNCTION TO FORECAST ADDITIVE MULTI-POPULATION
MORTALITY MODEL*

Description

R function for forecasting additive multi-population mortality model developed by: Debon et al (2011). This model follows the structure of the well-known Lee-Carter model (Lee and Carter, 1992) but including an additive parameter to capture the behavior of each population considered. This parameter seeks to capture the individual behavior of every population considered. It should be mentioned that in case that this function is developed for forecasting several populations. However, in case you only consider one population, the function will forecast the Lee-Carter model for one population.

Usage

```
for_additive.LC.multi(
  fitted.obj,
```

```

    nahead,
    ktmethod = c("Arimapdq", "arima010"),
    kt_include.cte = TRUE
  )

```

Arguments

`fitted.obj` object developed using function `fit_additive.LC.multi()`.

`nahead` number of periods ahead to forecast.

`ktmethod` method used to forecast the value of `kt` Arima(p,d,q) or ARIMA(0,1,0); c("Arimapdq", "arima010").

`kt_include.cte` if you want that `kt` include constant in the arima process.

Value

A list with different components of the forecasting process:

- `ax` parameter that captures the average shape of the mortality curve in all considered populations.
- `bx` parameter that explains the age effect `x` with respect to the general trend `kt` in the mortality rates of all considered populations.
- `arimakt` the arima selected for the `kt` time series.
- `kt.fitted` obtained values for the tendency behavior captured by `kt`.
- `kt.fut` projected values of `kt` for the `nahead` periods ahead.
- `kt.futintervals` arima selected and future values of `kt` with the different intervals, lower and upper, 80\
- `Ii` parameter that captures the differences in the pattern of mortality in any region `i` with respect to Region 1.
- `formula` additive multi-population mortality formula used to fit the mortality rates.
- `qxt.real` real mortality rates.
- `qxt.fitted` fitted mortality rates using the additive multi-population mortality model.
- `logit.qxt.fitted` fitted mortality rates in logit way estimated with the additive multi-population mortality model.
- `qxt.future` future mortality rates estimated with the additive multi-population mortality model.
- `logit.qxt.future` future mortality rates in logit way estimated with the additive multi-population mortality model.
- `nPop` provided number of populations to fit the periods.

References

- Debon, A., Montes, F., & Martinez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. *European Actuarial Journal*, 1, 291-308.
- Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659–671.

See Also

[fit_additive.LC.multi](#), [fit_multiplicative.LC.multi](#), [for_multiplicative.LC.multi](#), [multipopulation_cv](#),

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#First, we present the data that we are going to use
SpainRegions
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
#Before forecast the future value of the mortality rates for different populations
#we need the object the fitted values of the additive multi-population mortality model.
library(gnm)
additive_Spainmales <- fit_additive.LC.multi(qxt = SpainRegions$qx_male,
                                           periods = c(1991:2020),
                                           ages = c(ages),
                                           nPop = 18,
                                           lxt = SpainRegions$lx_male)

#Once, we have fit the data, it is possible to forecast 10 years ahead
#in this case for the additive multi-population mortality model
fut_additive_Spainmales <- for_additive.LC.multi(fitted.obj = additive_Spainmales, nahead = 10,
                                               ktmethod = "Arimapdq", kt_include.cte = TRUE)

#As we mentioned in the details of the function, if we only provide the data
#from one-population the function fit_additive.LC.multi()
#will fit the Lee-Carter model for single populations.
LC_Spainmales <- fit_additive.LC.multi(qxt = SpainNat$qx_male,
                                     periods = c(1991:2020),
                                     ages = ages,
                                     nPop = 1)

#To forecast future values of the single population using LC
fut_LC_Spainmales <- for_additive.LC.multi(fitted.obj = LC_Spainmales,
                                          nahead = 10, ktmethod = "Arimapdq",
                                          kt_include.cte = TRUE)
```

```
for_multiplicative.LC.multi
```

*FUNCTION TO FORECAST MULTIPLICATIVE MULTI-
POPULATION MORTALITY MODEL*

Description

R function for forecasting multiplicative multi-population mortality model developed by: Russo-lillo et al (2011). This model follows the structure of the well-known Lee-Carter model (Lee and Carter, 1992) but including a multiplicative parameter to capture the behavior of each population considered. This parameter seeks to capture the individual behavior of every population considered.

It should be mentioned that in case that this function is developed for forecasting several populations. However, in case you only consider one population, the function will forecast the Lee-Carter model for one population.

Usage

```
for_multiplicative.LC.multi(
  fitted.obj,
  nahead,
  ktmethod = c("Arimapdq", "arima010"),
  kt_include.cte = TRUE
)
```

Arguments

`fitted.obj` object developed using function `fit_mutllicative.LC.multi()`

`nahead` number of periods ahead to forecast

`ktmethod` method used to forecast the value of `kt` Arima(p,d,q) or ARIMA(0,1,0); c("Arimapdq", "arima010").

`kt_include.cte` if you want that `kt` include constant in the arima process.

Value

A list with different components of the forecasting process:

- `ax` parameter that captures the average shape of the mortality curve in all considered populations.
- `bx` parameter that explains the age effect `x` with respect to the general trend `kt` in the mortality rates of all considered populations.
- `arimakt` the arima selected for the `kt` time series.
- `kt.fitted` obtained values for the tendency behavior captured by `kt`.
- `kt.fut` projected values of `kt` for the `nahead` periods ahead.
- `kt.futintervals` arima selected and future values of `kt` with the different intervals, lower and upper, 80\
- `Ii` gives an idea of the differences in the pattern of mortality in any region `i` with respect to Region 1.
- `formula` additive multi-population mortality formula used to fit the mortality rates.
- `qxt.real` real mortality rates.
- `qxt.fitted` fitted mortality rates using the additive multi-population mortality model.
- `logit.qxt.fitted` fitted mortality rates in logit way estimated with the additive multi-population mortality model.
- `qxt.future` future mortality rates estimated with the additive multi-population mortality model.
- `logit.qxt.future` future mortality rates in logit way estimated with the additive multi-population mortality model.
- `nPop` provided number of populations to fit the periods.

MAE *MEAN ABSOLUTE ERRORS (MAE)*

Description

R function to estimate the mean absolute errors (MAE) for the mortality rates;

$$\frac{1}{n} \sum_x \sum_t |qxt1 - qxt2|$$

. where qxt1 is the real mortality rates qxt_re, and qxt2 is the adjusted mortality rates qxt_aju. You only have to provide the real value and the fitted or forecasted value for your mortality rates. However, the function is constructed to provide the real value and the fitted or forecasted value of your independent variable. These variables must have the same dimensions to be compared.

Usage

```
MAE(qxt_re, qxt_aju, wxt)
```

Arguments

qxt_re	real mortality rates used to check the goodness of fit measure.
qxt_aju	adjusted mortality rates using a specific mode.
wxt	weights of the mortality rates or data provided.

Value

A value of MAE for the data provided.

References

Atance, D., Debón, A., & Navarro, E. (2020). A comparison of forecasting mortality models using resampling methods. *Mathematics*, 8(9), 1550.

See Also

[fit_additive.LC.multi](#), [fit_multiplicative.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [SSE](#), [MSE](#), [MAPE](#), [multipopulation_cv](#), [multipopulation_loocv](#).

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest
```

```
#To show how the function works, we need to provide fitted or forecasted data and the real data.
#In this case, we employ the following data of the library:
SpainRegions
```

```

library(gnm)
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
#In this case, we fit for males providing the lxt
multiplicative_Spainmales <- fit_multiplicative.LC.multi(qxt = SpainRegions$qx_male,
  periods = c(1991:2020),
  ages = c(ages),
  nPop = 18,
  lxt = SpainRegions$lx_male)

#Once, we have the fitted data, we will obtain the SSE for the first population.
#We need to obtain wxt (weight of the mortality rates or data provided) using a
library(StMoMo)
wxt_1pob <- genWeightMat(ages = ages, years = c(1991:2020), clip = 0)
MAE(qxt_re = multiplicative_Spainmales$qxt.real$pob1,
  qxt_aju = multiplicative_Spainmales$qxt.fitted$pob1,
  wxt = wxt_1pob)

```

MAPE

MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

Description

R function to estimate the mean absolute percentage error (MAPE) for the mortality rates:

$$\frac{1}{n} \sum_x \sum_t \left| \frac{(qxt1 - qxt2)}{qxt2} \right|$$

where qxt1 is the real mortality rates qxt_re, and qxt2 is the adjusted mortality rates qxt_aju. You only have to provide the real value and the fitted or forecasted value for your mortality rates. However, the function is constructed to provide the real value and the fitted or forecasted value of your independent variable. These variables must have the same dimensions to be compared.

Usage

```
MAPE(qxt_re, qxt_aju, wxt)
```

Arguments

qxt_re	real mortality rates used to check the goodness of fit measure.
qxt_aju	adjusted mortality rates using a specific mode.
wxt	weights of the mortality rates or data provided.

Value

A value of MAPE for the data provided.

References

Atance, D., Debón, A., & Navarro, E. (2020). A comparison of forecasting mortality models using resampling methods. *Mathematics*, 8(9), 1550.

See Also

[fit_additive.LC.multi](#), [fit_multiplicative.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [MAE](#), [MSE](#), [SSE](#), [multipopulation_cv](#), [multipopulation_loocv](#).

Examples

```
##The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#To show how the function works, we need to provide fitted or forecasted data and the real data.
#In this case, we employ the following data of the library:
SpainRegions
library(gnm)
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
#In this case, we fit for males providing the lxt
multiplicative_Spainmales <- fit_multiplicative.LC.multi(qxt = SpainRegions$qx_male,
  periods = c(1991:2020),
  ages = c(ages),
  nPop = 18,
  lxt = SpainRegions$lx_male)

#Once, we have the fitted data, we will obtain the SSE for the first population.
#We need to obtain wxt (weight of the mortality rates or data provided) using a
library(StMoMo)
wxt_1pob <- genWeightMat(ages = ages, years = c(1991:2020), clip = 0)
MAPE(qxt_re = multiplicative_Spainmales$qxt.real$pob1,
  qxt_aju = multiplicative_Spainmales$qxt.fitted$pob1,
  wxt = wxt_1pob)
```

MSE

MEAN SQUARED ERRORS (MSE)

Description

R function to estimate the mean squared errors (MSE) for the mortality rates:

$$\frac{1}{n} \sum_x \sum_t (qxt1 - qxt2)^2 = \frac{1}{n} SSE$$

where $qxt1$ is the real mortality rates qxt_re , and $qxt2$ is the adjusted mortality rates qxt_aju . You only have to provide the real value and the fitted or forecasted value for your mortality rates. However, the function is constructed to provide the real value and the fitted or forecasted value of your independent variable. These variables must have the same dimensions to be compared.

Usage

```
MSE(qxt_re, qxt_aju, wxt)
```

Arguments

```
qxt_re      real mortality rates used to check the goodness of fit measure.
qxt_aju     adjusted mortality rates using a specific mode.
wxt        weights of the mortality rates or data provided.
```

Value

A value of MSE for the data provided.

References

Atance, D., Debón, A., & Navarro, E. (2020). A comparison of forecasting mortality models using resampling methods. *Mathematics*, 8(9), 1550.

See Also

[fit_additive.LC.multi](#), [fit_multiplicative.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [MAE](#), [SSE](#), [MAPE](#), [multipopulation_cv](#), [multipopulation_loocv](#).

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#To show how the function works, we need to provide fitted or forecasted data and the real data.
#In this case, we employ the following data of the library:
SpainRegions
library(gnm)
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
#In this case, we fit for males providing the lxt
multiplicative_Spainmales <- fit_multiplicative.LC.multi(qxt = SpainRegions$qx_male,
  periods = c(1991:2020),
  ages = c(ages),
  nPop = 18,
  lxt = SpainRegions$lx_male)

#Once, we have the fitted data, we will obtain the SSE for the first population.
#We need to obtain wxt (weight of the mortality rates or data provided) using a
library(StMoMo)
wxt_1pob <- genWeightMat(ages = ages, years = c(1991:2020), clip = 0)
MSE(qxt_re = multiplicative_Spainmales$qxt.real$pob1,
  qxt_aju = multiplicative_Spainmales$qxt.fitted$pob1,
  wxt = wxt_1pob)
```

multipopulation_cv *FUNCTION TO APPLY CROSS-VALIDATION TECHNIQUES
FOR TESTING THE FORECASTING ACCURACY OF MULTI-
POPULATION MORTALITY MODELS*

Description

R function for testing the accuracy out-of-sample of different multi-population mortality models, Additive (Debon et al., 2011) and Multiplicative (Russolillo et al., 2011). We provide a R function that employ the cross-validation techniques for data panel-time series (Atance et al. 2020) to test the forecasting accuracy. These techniques consist on split the database in two parts: training set (to run the model) and test set (to check the forecasting accuracy of the model). This procedure is repeated several times trying to check the forecasting accuracy in different ways. With this function, the user can provide its own mortality rates for different populations. The function will split the database chronologically (Bergmeir and Benitez, 2012) based on the nahead which consist on the length of the training set. We have include the following Figure 1 to understand how the R function works. " alt" alt It should be mentioned that this function is developed for cross-validation the forecasting accuracy of several populations. However, in case you only consider one population, the function will forecast the Lee-Carter model for one population. To test the forecasting accuracy of the selected model, the function provides five different measures: SSE, MSE, MAE, MAPE or All. Depending on how you want to check the forecasting accuracy of the model you could select one or other. In this case, the measures will be obtained using the mortality rates in the normal scale as recommended by Santolino (2023) against the log scale.

Usage

```

multipopulation_cv(
  qxt,
  model = c("additive", "multiplicative"),
  periods,
  ages,
  nPop,
  lxt = NULL,
  nahead,
  ktmethod = c("Arimapdq", "arima010"),
  kt_include.cte = TRUE,
  measures = c("SSE", "MSE", "MAE", "MAPE", "All")
)

```

Arguments

qxt	mortality rates used to fit the multi-population mortality models. This rates can be provided in matrix or in data.frame.
model	choose the multi-population mortality model to fit the mortality rates c("additive", "multiplicative")
periods	periods considered in the fitting in a vector way c(minyear:maxyear).

ages	vector with the ages considered in the fitting. If the mortality rates provide from an abridged life tables, it is necessary to provide a vector with the ages, see the example.
nPop	number of population considered for fitting.
lxt	survivor function considered for every population, not necessary to provide.
nahead	is a vector specifying the number of periods to block in the blocked CV. The function operates by using the sum of the periods in nahead and three (the minimum number of years required to construct a time series), as the initial training set. This ensures that the first train set has sufficient observations to forecast the initial test set, which will be of length nahead.
ktmethod	method used to forecast the value of kt Arima(p,d,q) or ARIMA(0,1,0); c("Arimapdq", "arima010").
kt_include.cte	if you want that kt include constant in the arima process.
measures	choose the non-penalized measure of forecasting accuracy that you want to use; c("SSE", "MSE", "MAE", "MAPE", "A11"). Check the function

Value

A list with different components of the cross-validation process:

- `ax` parameter that captures the average shape of the mortality curve in all considered populations.
- `bx` parameter that explains the age effect x with respect to the general trend `kt` in the mortality rates of all considered populations.
- `kt.fitted` obtained values for the tendency behavior captured by `kt`.
- `kt.future` future values of `kt` for every iteration in the cross-validation.
- `kt.arima` the arima selected for each `kt` time series.
- `Ii` parameter that captures the differences in the pattern of mortality in any region i with respect to Region 1.
- `formula` multi-population mortality formula used to fit the mortality rates.
- `nPop` provided number of populations to fit the periods.
- `qxt.real` real mortality rates.
- `qxt.future` future mortality rates estimated with the multi-population mortality model.
- `logit.qxt.future` future mortality rates in logit way estimated with the multi-population mortality model.
- `meas_ages` measure of forecasting accuracy through the ages of the study.
- `meas_periods` measure of forecasting accuracy in every forecasting period(s) of the study.
- `meas_pop` measure of forecasting accuracy through the populations considered in the study.
- `meas_total` a global measure of forecasting accuracy through the ages, periods and populations of the study.

References

- Atance, D., Debon, A., and Navarro, E. (2020). A comparison of forecasting mortality models using resampling methods. *Mathematics* 8(9): 1550.
- Bergmeir, C. & Benitez, J.M. (2012) On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
- Debon, A., & Atance, D. (2022). Two multi-population mortality models: A comparison of the forecasting accuracy with resampling methods. in *Contributions to Risk Analysis: Risk 2022*. Fundacion Mapfre
- Debon, A., Montes, F., & Martinez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. *European Actuarial Journal*, 1, 291-308.
- Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659–671.
- Russolillo, M., Giordano, G., & Haberman, S. (2011). Extending the Lee–Carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, 96-117.
- Santolino, M. (2023). Should Selection of the Optimum Stochastic Mortality Model Be Based on the Original or the Logarithmic Scale of the Mortality Rate?. *Risks*, 11(10), 170.

See Also

[multipopulation_loocv](#), [fit_additive.LC.multi](#), [fit_multiplicative.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [plotLC.multi](#), [SSE](#), [MAE](#), [MAPE](#).

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#We present a cross-validation method for spanish male regions

ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40,
          45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
library(gnm)
#Let start with a simple nahead=5 CV method obtaining the SSE forecasting measure of accuracy
cv_Spainmales_addit <- multipopulation_cv(qxt = SpainRegions$qx_male,
                                         model = c("additive"),
                                         periods = c(1991:2020), ages = c(ages),
                                         nPop = 18, lxt = SpainRegions$lx_male,
                                         nahead = 5,
                                         ktmethod = c("Arimapdq"),
                                         kt_include.cte = TRUE,
                                         measures = c("SSE"))

#Once, we have run the function we can check the result in different ways:
cv_Spainmales_addit$meas_ages
cv_Spainmales_addit$meas_periods fut
```

```
cv_Spainmales_addit$meas_pop
cv_Spainmales_addit$meas_total
```

```
multipopulation_loocv FUNCTION TO APPLY LEAVE-ONE-OUT CROSS-VALIDATION
TECHNIQUE FOR TESTING THE FORECASTING ACCURACY OF
MULTI-POPULATION MORTALITY MODELS
```

Description

R function for testing the accuracy out-of-sample of different multi-population mortality models, Additive (Debon et al., 2011) and Multiplicative (Russolillo et al., 2011). We provide a R function that employ the leave-one-out cross-validation technique for data panel-time series (Atance et al. 2020) to test the forecasting accuracy of one-multipopulation mortality model. This technique consists on split the database in two parts: training set (to run the model) and test set (to check the forecasting accuracy of the model) with only one data (one period in this case). This procedure is repeated several times trying to check the forecasting accuracy in different ways enlarging the training set one period-ahead. With this function, the user can provide its own mortality rates for different populations. The function will split the database chronologically (Bergmeir and Benitez, 2012) based on the `trainset1` which consist on the length of the first training set. We have include the following Figure 2 to understand how the R function works. " alt" alt It should be mentioned that this function is developed for testing the the forecasting accuracy of several populations using leave-one-out cross-validation . However, in case you only consider one population, the function will forecast the Lee-Carter model for one population. To test the forecasting accuracy of the selected model, the function provides five different measures: SSE, MSE, MAE, MAPE or All. Depending on how you want to check the forecasting accuracy of the model you could select one or other. In this case, the measures will be obtained using the mortality rates in the normal scale as recommended by Santolino (2023) against the log scale.

Usage

```
multipopulation_loocv(
  qxt,
  model = c("additive", "multiplicative"),
  periods,
  ages,
  nPop,
  lxt = NULL,
  ktmethod = c("Arimapdq", "arima010"),
  kt_include.cte = TRUE,
  measures = c("SSE", "MSE", "MAE", "MAPE", "All"),
  trainset1
)
```

Arguments

<code>qxt</code>	mortality rates used to fit the multi-population mortality models. This rates can be provided in matrix or in data.frame.
<code>model</code>	choose the multi-population mortality model to fit the mortality rates <code>c("additive", "multiplicative")</code>
<code>periods</code>	periods considered in the fitting in a vector way <code>c(minyear:maxyear)</code> .
<code>ages</code>	vector with the ages considered in the fitting. If the mortality rates provide from an abridged life tables, it is necessary to provide a vector with the ages, see the example.
<code>nPop</code>	number of population considered for fitting.
<code>lxt</code>	survivor function considered for every population, not necessary to provide.
<code>ktmethod</code>	method used to forecast the value of <code>kt</code> Arima(p,d,q) or ARIMA(0,1,0); <code>c("Arimapdq", "arima010")</code> .
<code>kt_include.cte</code>	if you want that <code>kt</code> include constant in the arima process.
<code>measures</code>	choose the non-penalized measure of forecasting accuracy that you want to use; <code>c("SSE", "MSE", "MAE", "MAPE", "A11")</code> . Check the function
<code>trainset1</code>	vector with the periods for the first training set. This value must be greater than 2 to meet the minimum time series size (Hyndman and Khandakar, 2008).

Value

A list with different components of the cross-validation process:

- `ax` parameter that captures the average shape of the mortality curve in all considered populations.
- `bx` parameter that explains the age effect `x` with respect to the general trend `kt` in the mortality rates of all considered populations.
- `kt.fitted` obtained values for the tendency behavior captured by `kt`.
- `kt.future` future values of `kt` for every iteration in the cross-validation.
- `kt.arima` the arima selected for each `kt` time series.
- `Ii` parameter that captures the differences in the pattern of mortality in any region `i` with respect to Region 1.
- `formula` multi-population mortality formula used to fit the mortality rates.
- `nPop` provided number of populations to fit the periods.
- `qxt.real` real mortality rates.
- `qxt.future` future mortality rates estimated with the multi-population mortality model.
- `logit.qxt.future` future mortality rates in logit way estimated with the multi-population mortality model.
- `meas_ages` measure of forecasting accuracy through the ages of the study.
- `meas_periods` measure of forecasting accuracy in every forecasting period(s) of the study.
- `meas_pop` measure of forecasting accuracy through the populations considered in the study.
- `meas_total` a global measure of forecasting accuracy through the ages, periods and populations of the study.

References

- Atance, D., Debon, A., and Navarro, E. (2020). A comparison of forecasting mortality models using resampling methods. *Mathematics* 8(9): 1550.
- Bergmeir, C. & Benitez, J.M. (2012) On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
- Debon, A., & Atance, D. (2022). Two multi-population mortality models: A comparison of the forecasting accuracy with resampling methods. in *Contributions to Risk Analysis: Risk 2022*. Fundacion Mapfre
- Debon, A., Montes, F., & Martinez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. *European Actuarial Journal*, 1, 291-308.
- Hyndman, R.J. & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical. Software*, 26, 1–22.
- Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659–671.
- Russolillo, M., Giordano, G., & Haberman, S. (2011). Extending the Lee–Carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, 96-117.
- Santolino, M. (2023). Should Selection of the Optimum Stochastic Mortality Model Be Based on the Original or the Logarithmic Scale of the Mortality Rate?. *Risks*, 11(10), 170.

See Also

[multipopulation_cv](#), [fit_additive.LC.multi](#), [fit_multiplicative.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [plotLC.multi](#), [SSE](#), [MAE](#), [MAPE](#).

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#We present the leave-one-out cross-validation (LOOCV) method for spanish male regions
#The idea is to get the same results as in the short paper published in Risk Congress 2023
SpainRegions
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40,
          45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
library(gnm)
#Let start with a simple trainset1 = 10 CV method obtaining the SSE forecasting measure of accuracy
loocv_Spainmales_addit <- multipopulation_loocv(qxt = SpainRegions$qx_male,
                                               model = c("additive"),
                                               periods = c(1991:2020), ages = c(ages),
                                               nPop = 18, lxt = SpainRegions$lx_male,
                                               ktmethod = c("Arimapdq"),
                                               kt_include.cte = TRUE,
                                               measures = c("SSE"),
                                               trainset1 = 10)

#Once, we have run the function we can check the result in different ways:
```

```
loocv_Spainmales_addit$meas_ages  
loocv_Spainmales_addit$meas_periodsfut  
loocv_Spainmales_addit$meas_pop  
loocv_Spainmales_addit$meas_total
```

plotLC.multi

FUNCTION TO PLOT MULTI-POPULATION MORTALITY MODEL

Description

R function to plot the parameters for the Additive (Debon et al., 2011) and Multiplicative (Russolillo et al., 2011) Multi-Population mortality model. It should be mentioned that in case that this function is developed for fitting several populations. However, in case you only consider one population, the function will fit the one-population Lee-Carter model (Lee and Carter, 1992).

Usage

```
plotLC.multi(fitted.obj)
```

Arguments

fitted.obj object developed using function `fit.additive.LC.multi()` and `fit.multiplicative.LC.multi()`.

Value

plot the different parameters for the multi-population mortality models a_x , b_x , k_t and I_i . This function is valid for both approaches Additive and Multiplicative multi-population mortality models.

References

Debon, A., Montes, F., & Martinez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. *European Actuarial Journal*, 1, 291-308.

Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659–671.

Multi-population mortality model developed by: Russolillo, M., Giordano, G., & Haberman, S. (2011). Extending the Lee–Carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, 2011(2), 96-117.

See Also

[fit_additive.LC.multi](#), [fit_multiplicative.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [multipopulation_cv](#), [multipopulation_loocv](#)

Examples

```

#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

SpainRegions
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
#In this case, we fit for males providing the lxt
library(gnm)
multiplicative_Spainmales <- fit_multiplicative.LC.multi(qxt = SpainRegions$qx_male,
  periods = c(1991:2020),
  ages = c(ages),
  nPop = 18,
  lxt = SpainRegions$lx_male)

#Once, we have fit the data, it is possible to see the ax, bx, kt, and Ii
#provided parameters for the fitting.
plotLC.multi(multiplicative_Spainmales)

#In case you only provide one population the model will fit
#the single population version of the Lee-Carter model
LC_Spainmales <- fit_additive.LC.multi(qxt = SpainNat$qx_male,
  periods = c(1991:2020),
  ages = ages,
  nPop = 1)

plotLC.multi(LC_Spainmales)

```

regions

regions

Description

Data from the Spanish region of Spain which are provided to plot an indicator. This dataset contains a plot with the information of Spain regions (geometry and name of every region).

Usage

```
regions
```

Format

A data frame with 600 rows and 9 columns with the following information

- `Codigo` a vector containing the code of every region of Spain.
- `Texto` a vector containing the name of every region of Spain.
- `Texto_Alt` a vector containing the long name of every region of Spain.
- `Ii` a vector containing a possible value of one indicator to be shown.
- `geometry` the dimension of every region of Spain. This vector allows to plot the regions of Spain.

Value

a plot with the Spain regions colored by the indicator provided.

References

Spanish National Institute of Statistics (INE) (2023). Tablas de mortalidad, metodologia. Technical report, Instituto Nacional de Estadistica

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#In this case, we show the region dataset applying it to a multipopulation model.
#First, we present the dataset
regions
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
library(gnm)
#Then, we fit the multiplicative model
multiplicative_Spainmales <- fit_multiplicative.LC.multi(qxt = SpainRegions$qx_male,
                                                       periods = c(1991:2020),
                                                       ages = c(ages),
                                                       nPop = 18,
                                                       lxt = SpainRegions$lx_male)

#To show the values of the population indicator in the Spanish map.
SpainMap(regionvalue = multiplicative_Spainmales$Ii[2:18],
         main = c("Multiplicative for males"),
         name = c("Ii"))
```

SpainMap

Spain National map information

Description

This data contains information to plot the percentiles plot in Spanish regions. Therefore, the users only have to provide a specific variable to show in regions of Spain.

Usage

```
SpainMap(regionvalue, main, name)
```

Arguments

regionvalue	vector with the values that you want to plot in percentiles in the Spain map.
main	the specific title of the map plot
name	the assigned name for the legend in map plot.

Value

a map from the regions of Spain colored with the variable provided by the user.

References

Spanish National Institute of Statistics (INE) (2023). Tablas de mortalidad, metodologia. Technical report, Instituto Nacional de Estadistica

Examples

```
name <- c("Ii")
main <- c("Multiplicative for males")
regionvalue <- c(0.131867619, -0.063994652, 0.088094096,
                 0.019685552, 0.064671498, 0.012212161,
                 -0.088864474, -0.146079884, -0.017703536,
                 0.050376721, 0.052476852, -0.022871202,
                 -0.093952332, 0.049266816, -0.101224890,
                 0.001481333, -0.078523511)
```

```
SpainMap(regionvalue, main, name)
```

SpainNat

Spain National Mortality data

Description

Data from the Spanish national of Spain from the Spanish National Institute of Statistics (INE) for both genders years 1991-2020 and abridged ages from 0 to 90. This dataset contains mortality rates for the total national population of Spain. Additionally, the dataset includes the number of people alive (lxt) for each age and period.

Usage

```
SpainNat
```

Format

A data frame with 600 rows and 9 columns with the following information

- ccaa a vector containing all the regions of Spain. Indeed, the column takes the following information: Spain.
- years a vector containing the periods of the dataset from 1991 to 2020.
- ages a vector containing the abridged ages considered in the dataset, 0, <1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, and 90-94.
- qx_male mortality rates for the males in the Spain Nation.

- qx_female mortality rates for the females in the Spain Nation.
- lx_male survivor function considered for the males of Spain Nation.
- lx_female survivor function considered for the females of Spain Nation.
- series information for the series of data provided.
- label the assigned tag to the data frame.

References

Spanish National Institute of Statistics (INE) (2023). Tablas de mortalidad, metodologia. Technical report, Instituto Nacional de Estadística

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#In this case, we show the region dataset applying it to a multipopulation model.
#First, we present the dataset
SpainNat
#An example to how the additive multi-population model fits the data
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
library(gnm)
LC_Spainmales <- fit_additive.LC.multi(qxt = SpainNat$qx_male,
                                     periods = c(1991:2020),
                                     ages = ages,
                                     nPop = 1)
```

SpainRegions

Spain Regions Mortality data

Description

Data from the Spanish region of Spain from the Spanish National Institute of Statistics (INE) for both genders years 1991-2020 and abridged ages from 0 to 90. This dataset contains mortality rates (qxt) from 18 different regions of Spain. Additionally, the dataset includes the number of people alive (lxt) for each age and period.

Usage

SpainRegions

SSE *SUM of SQUARED ERRORS (SSE)*

Description

R function to estimate the sum of squared errors for the mortality rates:

$$\sum_x \sum_t (qxt1 - qxt2)^2$$

where `qxt1` is the real mortality rates `qxt_re`, and `qxt2` is the adjusted mortality rates `qxt_aju`. You only have to provide the real value and the fitted or forecasted value for your mortality rates. However, the function is constructed to provide the real value and the fitted or forecasted value of your independent variable. These variables must have the same dimensions to be compared.

Usage

```
SSE(qxt_re, qxt_aju, wxt)
```

Arguments

<code>qxt_re</code>	real mortality rates used to check the goodness of fit measure.
<code>qxt_aju</code>	adjusted mortality rates using a specific mode.
<code>wxt</code>	weights of the mortality rates or data provided.

Value

A value of SSE for the data provided.

References

Atance, D., Debón, A., & Navarro, E. (2020). A comparison of forecasting mortality models using resampling methods. *Mathematics*, 8(9), 1550.

See Also

[fit_additive.LC.multi](#), [fit_multiplicative.LC.multi](#), [for_additive.LC.multi](#), [for_multiplicative.LC.multi](#), [MAE](#), [MSE](#), [MAPE](#), [multipopulation_cv](#), [multipopulation_loocv](#).

Examples

```
#The example takes more than 5 seconds because it includes
#several fitting and forecasting process and hence all
#the process is included in donttest

#To show how the function works, we need to provide fitted or forecasted data and the real data.
#In this case, we employ the following data of the library:
#First, we present the dataset
```

```
#To show how the function works, we need to provide fitted or forecasted data and the real data.
#In this case, we employ the following data of the library:
SpainRegions
ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
#In this case, we fit for males providing the lxt
library(gnm)
multiplicative_Spainmales <- fit_multiplicative.LC.multi(qxt = SpainRegions$qx_male,
               periods = c(1991:2020),
               ages = c(ages),
               nPop = 18,
               lxt = SpainRegions$lx_male)

#Once, we have the fitted data, we will obtain the SSE for the first population.
#We need to obtain wxt (weight of the mortality rates or data provided) using a
library(StMoMo)
wxt_1pob <- genWeightMat(ages = ages, years = c(1991:2020), clip = 0)
SSE(qxt_re = multiplicative_Spainmales$qxt.real$pob1,
    qxtaju = multiplicative_Spainmales$qxt.fitted$pob1,
    wxt = wxt_1pob)
```

Index

* datasets

regions, [22](#)

SpainNat, [24](#)

SpainRegions, [25](#)

fit_additive.LC.multi, [2](#), [5](#), [8](#), [10](#), [11](#), [13](#),
[14](#), [17](#), [20](#), [21](#), [27](#)

fit_multiplicative.LC.multi, [3](#), [4](#), [8](#), [10](#),
[11](#), [13](#), [14](#), [17](#), [20](#), [21](#), [27](#)

for_additive.LC.multi, [3](#), [5](#), [6](#), [10](#), [11](#), [13](#),
[14](#), [17](#), [20](#), [21](#), [27](#)

for_multiplicative.LC.multi, [3](#), [5](#), [8](#), [8](#),
[11](#), [13](#), [14](#), [17](#), [20](#), [21](#), [27](#)

MAE, [11](#), [13](#), [14](#), [17](#), [20](#), [27](#)

MAPE, [11](#), [12](#), [14](#), [17](#), [20](#), [27](#)

MSE, [11](#), [13](#), [13](#), [27](#)

multipopulation_cv, [3](#), [5](#), [8](#), [10](#), [11](#), [13](#), [14](#),
[15](#), [20](#), [21](#), [27](#)

multipopulation_loocv, [3](#), [5](#), [11](#), [13](#), [14](#), [17](#),
[18](#), [21](#), [27](#)

plotLC.multi, [3](#), [17](#), [20](#), [21](#)

regions, [22](#)

SpainMap, [23](#)

SpainNat, [24](#)

SpainRegions, [25](#)

SSE, [11](#), [13](#), [14](#), [17](#), [20](#), [27](#)