AVERAGE STATES ARIMAX MODEL - Influenza hospitalization data - no log-back transformation

Victor Felix

April 01, 2025

This function utilizes ensembles and single automatic ARIMAX models which have mean hospitalization on U.S. states as exogenous variables. The function fits on a rolling window of previous 104 weeks for the state under analysis and a rolling windows with the same size with 1 week-lag for the exogenous variables to generate forecasts. It return some metrics that evaluate the performance of the mod els:target_end_date, abs_error, cases, forecast, 'N_of_models", weighted interval score (WIS), predictive quantiles. The user can choose a single best automatic ARIMAXs (auto=TRUE), or ensembles of 27 permutations of 0,1,2 pdq's (ES27=TRUE) or 64 permutations of 0,1,2,3 pdq's (ES64=TRUE). The user also chooses the number of weeks ahead for each forecast, and the size of the rolling window which is set as 2 years (104 weeks).

```
knitr::opts_chunk$set(echo = TRUE)
```

!!!!!!!!!!!!!!!!!! LOADING THE DATASET !!!!!!!!!!!!!!!!!!

```
library("tidyr")
library("MMWRweek")
library("data.table")
library("caret")
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
library("purrr")
```

```
##
## Attaching package: 'purrr'
```

```
## The following object is masked from 'package:caret':
##
##     lift
```

```
## The following object is masked from 'package:data.table':
##
##     transpose
```

```r
library("dplyr")
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library("tseries")
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
library("gtools")
library("forecast")
library("scoringutils")
```

```
## Note: scoringutils is currently undergoing major development changes (with an update planned for the
```

```r
library("covidHubUtils")
library("parallel")
library("future")#https://cran.r-project.org/web/packages/future/vignettes/future-4-issues.html
```

```
##
## Attaching package: 'future'

## The following object is masked from 'package:tseries':
##
##     value

## The following object is masked from 'package:caret':
##
##     cluster
```

```r
library("listenv")
```

```
##
## Attaching package: 'listenv'
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
library("epitools")
```

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! LOADING DATASET AND FUNCTIONS !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

```
##########################################
#       LOADING AND CLEANING THE DATASET  #
##########################################

# Loads the AVERAGE states models
source("ES_AVERAGE_48_nolog.R", local = TRUE, chdir = TRUE)

# Loads the ILI dataset
my_data = read.csv("treated_influenza_hosp_dataframe_v2.csv")
my_data$target_end_date<-as.Date(my_data$target_end_date) # set the dates as dates

list_of_states <- split(my_data, my_data$state_name)
```

AUTO AVERAGE WEEK1 - ALL STATES

IF YOU WANT A FAST RESULT RUN ONLY THE AUTO_example, the others take longer.

```
start_time <- Sys.time()

AUTO_AVERAGE_WEEK1_list <- mclapply(list_of_states, ES_AVERAGE, auto=TRUE, n_weeks_ahead=1,list_of_stat
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 8.420558 mins
```

```
# FINAL DATAFRAME
AUTO_AVERAGE_WEEK1 <- bind_rows(AUTO_AVERAGE_WEEK1_list, .id = "State")
```

AUTO AVERAGE WEEK2 - ALL STATES

```
start_time <- Sys.time()

# RUN MODEL
AUTO_AVERAGE_WEEK2_list <- mclapply(list_of_states, ES_AVERAGE, auto=TRUE, n_weeks_ahead=2,list_of_stat
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 8.358378 mins
```

```
# FINAL DATAFRAME
AUTO_AVERAGE_WEEK2 <- bind_rows(AUTO_AVERAGE_WEEK2_list, .id = "State")
```

AUTO AVERAGE WEEK3 - ALL STATES

```
start_time <- Sys.time()

AUTO_AVERAGE_WEEK3_list <- mclapply(list_of_states, ES_AVERAGE, auto=TRUE, n_weeks_ahead=3,list_of_stat
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 8.145184 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
AUTO_AVERAGE_WEEK3 <- bind_rows(AUTO_AVERAGE_WEEK3_list, .id = "State")
```

AUTO AVERAGE WEEK4 - ALL STATES

```
start_time <- Sys.time()

AUTO_AVERAGE_WEEK4_list <- mclapply(list_of_states, ES_AVERAGE, auto=TRUE, n_weeks_ahead=4,list_of_stat
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 8.214522 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
AUTO_AVERAGE_WEEK4 <- bind_rows(AUTO_AVERAGE_WEEK4_list, .id = "State")
```

```
save.image("AVERAGE_MODELS_influenza_hospitalization_nolog.Rdata")
```

ES27 AVERAGE WEEK1 - ALL STATES

```
start_time <- Sys.time()

ES27_AVERAGE_WEEK1_list <- mclapply(list_of_states, ES_AVERAGE, ES27=TRUE, n_weeks_ahead=1,list_of_stat
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 11.93137 mins
```

```
# FINAL DATAFRAME
ES27_AVERAGE_WEEK1 <- bind_rows(ES27_AVERAGE_WEEK1_list, .id = "State")
```

ES27 AVERAGE WEEK2 - ALL STATES

```
start_time <- Sys.time()

ES27_AVERAGE_WEEK2_list <- mclapply(list_of_states, ES_AVERAGE, ES27=TRUE, n_weeks_ahead=2,list_of_state
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 11.91673 mins
```

```
# FINAL DATAFRAME
ES27_AVERAGE_WEEK2 <- bind_rows(ES27_AVERAGE_WEEK2_list, .id = "State")
```

ES27 AVERAGE WEEK3 - ALL STATES

```
start_time <- Sys.time()

ES27_AVERAGE_WEEK3_list <- mclapply(list_of_states, ES_AVERAGE, ES27=TRUE, n_weeks_ahead=3,list_of_state
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 12.05412 mins
```

```
# FINAL DATAFRAME
ES27_AVERAGE_WEEK3 <- bind_rows(ES27_AVERAGE_WEEK3_list, .id = "State")
```

ES27 AVERAGE WEEK4 - ALL STATES

```
start_time <- Sys.time()

ES27_AVERAGE_WEEK4_list <- mclapply(list_of_states, ES_AVERAGE, ES27=TRUE, n_weeks_ahead=4,list_of_state
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 12.26663 mins
```

```r
# FINAL DATAFRAME
ES27_AVERAGE_WEEK4 <- bind_rows(ES27_AVERAGE_WEEK4_list, .id = "State")
```

```r
save.image("AVERAGE_MODELS_influenza_hospitalization_nolog.Rdata")
```

ES64 AVERAGE WEEK1 - ALL STATES

```r
start_time <- Sys.time()

ES64_AVERAGE_WEEK1_list <- mclapply(list_of_states, ES_AVERAGE, ES64=TRUE, n_weeks_ahead=1,list_of_state
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 33.60045 mins
```

```r
# FINAL DATAFRAME
ES64_AVERAGE_WEEK1 <- bind_rows(ES64_AVERAGE_WEEK1_list, .id = "State")
```

ES64 AVERAGE WEEK2 - ALL STATES

```r
start_time <- Sys.time()

ES64_AVERAGE_WEEK2_list <- mclapply(list_of_states, ES_AVERAGE, ES64=TRUE, n_weeks_ahead=2,list_of_state
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 33.53258 mins
```

```r
# Combine the list of data frames into a single data frame with names as a column
ES64_AVERAGE_WEEK2 <- bind_rows(ES64_AVERAGE_WEEK2_list, .id = "State")
```

ES64 AVERAGE WEEK3 - ALL STATES

```r
start_time <- Sys.time()

ES64_AVERAGE_WEEK3_list <- mclapply(list_of_states, ES_AVERAGE, ES64=TRUE, n_weeks_ahead=3,list_of_state
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 33.07142 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
ES64_AVERAGE_WEEK3 <- bind_rows(ES64_AVERAGE_WEEK3_list, .id = "State")
```

ES64 AVERAGE WEEK4 - ALL STATES

```
start_time <- Sys.time()

ES64_AVERAGE_WEEK4_list <- mclapply(list_of_states, ES_AVERAGE, ES64=TRUE, n_weeks_ahead=4,list_of_stat
  setNames(names(list_of_states))

end_time <- Sys.time()
run_time <- end_time - start_time

print(run_time)
```

```
## Time difference of 32.12816 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
ES64_AVERAGE_WEEK4 <- bind_rows(ES64_AVERAGE_WEEK4_list, .id = "State")
```

```
save.image("AVERAGE_MODELS_influenza_hospitalization_nolog.Rdata")
```