TEMPERATURE ARIMAX MODELS - without log-back transformations

Victor Felix

November 12, 2025

Here we run ensembles and single automatic ARIMAX models for forecasting weekly hospitalizations in 48 states on the contiguous U.S. These model use mean temperature 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. These models do not include log-back transformations. It return some metrics that evaluate the performance of the models: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 PACKAGES !!!!!!!!!!!!!!!!!!!

```
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 model
source("ES_TEMPERATURE_nolog.R", local = TRUE, chdir = TRUE)

# Loads the 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)
```

Loading temperature data for 2010_2024 and filtering only the correct dates

```
# ERA5-based temperature dataframe
temperature_data<-read.csv("final_temperature_data_2010_2024.csv")
temperature_data$date<-as.Date(temperature_data$date)

correct_dates <- as.Date(list_of_states$Alabama$target_end_date)

# Filter temperature_data to keep only rows with correct dates
temperature_data <- temperature_data %>%
  filter(date %in% correct_dates)%>%
  select(-X,-date, -year,-epi_week)

head(temperature_data)
```

```
##    Alabama  Arizona Arkansas California Colorado Connecticut Delaware  Georgia
## 1 294.7937 303.9116 292.8267   298.0692 293.8906    288.2363 290.7262 294.4187
## 2 291.9592 293.7813 283.4730   293.7103 286.6248    280.7051 285.9110 293.7008
## 3 288.9502 301.6525 288.6751   296.0819 298.2114    282.7926 286.1192 289.5816
## 4 293.4747 289.9025 290.2545   285.5204 287.3689    287.5749 289.5262 294.5395
## 5 287.7191 298.6074 288.2169   290.7583 293.0266    278.7808 282.2132 287.8822
## 6 288.9218 291.9539 284.8963   287.5686 287.3409    281.1481 284.3456 290.0707
##      Idaho Illinois  Indiana     Iowa   Kansas Kentucky Louisiana    Maine
## 1 288.0447 284.6617 287.1833 277.7939 284.2681 291.7439  295.5982 282.9460
## 2 285.3451 279.0915 280.8556 276.0407 280.1079 285.2236  289.4352 275.2234
## 3 293.8576 286.2373 285.3423 287.9893 292.3379 286.3802  290.4244 278.9045
## 4 279.3297 285.5153 286.4789 281.9026 286.3925 289.4908  294.0563 284.0830
## 5 284.6930 282.4915 282.1890 281.2386 288.8831 284.2888  290.4109 274.9901
## 6 280.4953 279.7755 280.3660 277.8544 283.1184 283.5929  290.5063 274.2842
##   Maryland Massachusetts Michigan Minnesota Mississippi Missouri  Montana
## 1 290.9559      287.3740 280.0109  273.3173    294.8278 285.4641 276.8006
```

```
## 2 285.3042      280.0638 276.3995  273.1810    289.0763 279.7253 280.8457
## 3 285.7628      282.3675 284.3017  285.2937    289.0766 288.0899 291.8694
## 4 289.3039      287.5642 283.1108  277.2524    292.7960 286.5102 276.7364
## 5 282.2478      278.4986 277.7030  274.7542    288.6080 285.3930 283.2804
## 6 283.7669      280.1390 276.1560  274.3039    288.0228 281.5584 281.0216
##   Nebraska   Nevada New.Hampshire New.Jersey New.Mexico New.York North.Carolina
## 1 280.0342 297.9702      285.6007   289.6138   301.4300 287.1761       293.2653
## 2 279.4342 291.6445      277.2538   283.5736   290.2074 279.1964       291.2903
## 3 294.1473 298.1492      280.6518   284.7843   299.3307 282.5183       288.0366
## 4 283.3590 283.5382      286.4781   288.6801   292.8631 286.3851       292.5314
## 5 287.4419 290.9069      276.4031   280.4836   297.4077 277.9691       285.6968
## 6 281.9991 285.2120      276.6867   283.1411   291.5381 278.8004       287.6327
##   North.Dakota     Ohio Oklahoma   Oregon Pennsylvania Rhode.Island
## 1     272.0458 288.5318 289.7114 288.2784     289.7960     287.7229
## 2     275.1247 281.6339 281.1126 287.1505     281.9363     280.8590
## 3     286.6325 285.0062 291.3670 290.6956     284.3264     283.0342
## 4     275.4913 287.2672 289.7412 280.0285     288.0101     287.9712
## 5     276.3225 281.4956 290.1260 284.2948     280.2894     279.2735
## 6     276.1497 280.7134 284.6853 281.5111     281.1101     281.4089
##   South.Carolina South.Dakota Tennessee    Texas     Utah  Vermont Virginia
## 1       293.6616     275.3683  293.2874 296.6623 297.9245 286.0315 292.3678
## 2       292.7392     276.9592  288.2937 286.8796 289.0509 276.9201 288.4660
## 3       288.9966     292.2082  287.3370 294.2156 297.9158 280.5461 286.8486
## 4       293.6743     280.0482  291.2260 295.2065 284.5579 285.7118 291.0121
## 5       287.1819     282.9595  285.7185 293.3965 291.6520 276.0318 283.9284
## 6       288.8733     279.5673  285.3690 291.2612 285.1364 276.4167 285.4337
##   Washington West.Virginia Wisconsin  Wyoming  Florida
## 1   282.8672      292.5770  276.0997 286.7131 296.9210
## 2   281.6654      286.3808  274.4688 283.9113 296.8307
## 3   285.7566      286.2544  285.3811 296.8952 293.6598
## 4   278.3062      289.9524  280.9740 282.5517 296.7069
## 5   281.6213      283.3481  276.8769 288.0064 292.6155
## 6   279.4662      283.7928  275.4095 283.7294 294.1132
```

AUTO TEMPERATURE WEEK1

```
start_time <- Sys.time()

AUTO_TEMPERATURE_WEEK1_list <- mclapply(list_of_states, ES_TEMPERATURE, auto=TRUE, n_weeks_ahead=1,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 9.701151 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
AUTO_TEMPERATURE_WEEK1 <- bind_rows(AUTO_TEMPERATURE_WEEK1_list, .id = "State")
```

AUTO TEMPERATURE WEEK2

```
start_time <- Sys.time()

AUTO_TEMPERATURE_WEEK2_list <- mclapply(list_of_states, ES_TEMPERATURE, auto=TRUE, n_weeks_ahead=2,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 9.321846 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
AUTO_TEMPERATURE_WEEK2 <- bind_rows(AUTO_TEMPERATURE_WEEK2_list, .id = "State")
```

AUTO TEMPERATURE WEEK3

```
start_time <- Sys.time()

AUTO_TEMPERATURE_WEEK3_list <- mclapply(list_of_states, ES_TEMPERATURE, auto=TRUE, n_weeks_ahead=3,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 9.163023 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
AUTO_TEMPERATURE_WEEK3 <- bind_rows(AUTO_TEMPERATURE_WEEK3_list, .id = "State")
```

AUTO TEMPERATURE WEEK4

```
start_time <- Sys.time()

AUTO_TEMPERATURE_WEEK4_list <- mclapply(list_of_states, ES_TEMPERATURE, auto=TRUE, n_weeks_ahead=4,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 9.201534 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
AUTO_TEMPERATURE_WEEK4 <- bind_rows(AUTO_TEMPERATURE_WEEK4_list, .id = "State")
```

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

ES27 TEMPERATURE WEEK1

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

ES27_TEMPERATURE_WEEK1_list <- mclapply(list_of_states, ES_TEMPERATURE, ES27=TRUE, n_weeks_ahead=1,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 12.04466 mins
```

```r
# Combine the list of data frames into a single data frame with names as a column
ES27_TEMPERATURE_WEEK1 <- bind_rows(ES27_TEMPERATURE_WEEK1_list, .id = "State")
```

ES27 TEMPERATURE WEEK2

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

ES27_TEMPERATURE_WEEK2_list <- mclapply(list_of_states, ES_TEMPERATURE, ES27=TRUE, n_weeks_ahead=2,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 12.39249 mins
```

```r
# Combine the list of data frames into a single data frame with names as a column
ES27_TEMPERATURE_WEEK2 <- bind_rows(ES27_TEMPERATURE_WEEK2_list, .id = "State")
```

ES27 TEMPERATURE WEEK3

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

ES27_TEMPERATURE_WEEK3_list <- mclapply(list_of_states, ES_TEMPERATURE, ES27=TRUE, n_weeks_ahead=3,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 12.54213 mins
```

```r
# Combine the list of data frames into a single data frame with names as a column
ES27_TEMPERATURE_WEEK3 <- bind_rows(ES27_TEMPERATURE_WEEK3_list, .id = "State")
```

ES27 TEMPERATURE WEEK4

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

ES27_TEMPERATURE_WEEK4_list <- mclapply(list_of_states, ES_TEMPERATURE, ES27=TRUE, n_weeks_ahead=4,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 12.59198 mins
```

```r
# Combine the list of data frames into a single data frame with names as a column
ES27_TEMPERATURE_WEEK4 <- bind_rows(ES27_TEMPERATURE_WEEK4_list, .id = "State")
```

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

ES64 TEMPERATURE WEEK1

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

ES64_TEMPERATURE_WEEK1_list <- mclapply(list_of_states, ES_TEMPERATURE, ES64=TRUE, n_weeks_ahead=1,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 36.63639 mins
```

```r
# Combine the list of data frames into a single data frame with names as a column
ES64_TEMPERATURE_WEEK1 <- bind_rows(ES64_TEMPERATURE_WEEK1_list, .id = "State")
```

ES64 TEMPERATURE WEEK2

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

ES64_TEMPERATURE_WEEK2_list <- mclapply(list_of_states, ES_TEMPERATURE, ES64=TRUE, n_weeks_ahead=2,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 35.18824 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
ES64_TEMPERATURE_WEEK2 <- bind_rows(ES64_TEMPERATURE_WEEK2_list, .id = "State")
```

ES64 TEMPERATURE WEEK3

```
start_time <- Sys.time()

ES64_TEMPERATURE_WEEK3_list <- mclapply(list_of_states, ES_TEMPERATURE, ES64=TRUE, n_weeks_ahead=3,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 34.77658 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
ES64_TEMPERATURE_WEEK3 <- bind_rows(ES64_TEMPERATURE_WEEK3_list, .id = "State")
```

ES64 TEMPERATURE WEEK4

```
start_time <- Sys.time()

ES64_TEMPERATURE_WEEK4_list <- mclapply(list_of_states, ES_TEMPERATURE, ES64=TRUE, n_weeks_ahead=4,week
  setNames(names(list_of_states))

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

print(run_time)
```

```
## Time difference of 35.36914 mins
```

```
# Combine the list of data frames into a single data frame with names as a column
ES64_TEMPERATURE_WEEK4 <- bind_rows(ES64_TEMPERATURE_WEEK4_list, .id = "State")
```

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