TEAM: an R package for time series model identification

Methodology and Sampling Design Department
Information Technology and Communications Department
Statistics Spain (INE)

November 2024

INE
Instituto Nacional de Estadística

Context and motivation

Seasonal adjustment

Goal: Remove the seasonal (and calendar) component present in many time series to facilitate interpretation and comparison between successive periods.

Two main methods:

- X13: Based on fixed filters.
- **TRAMO-SEATS**: Based on ARIMA models. Adapts the filters for extraction of the seasonally adjusted series to the identified ARIMA model. It allows to obtain diagnostics on the seasonal adjustment process.

Revision policy

- A revision policy is needed since every time new data is available the regARIMA model and the filters should be reestimated.
- Compromise between being constantly revising the seasonally adjusted series and the best model at each period.

Revision policy in two phases:

1. Annual phase: once a year, the whole regARIMA model is reidentified.
2. Interannual phase: every time new data arrives (monthly or quarterly) the coefficients of the model are reestimated, but the model is not changed.

Issues with the annual identification phase

The annual phase of identifying regARIMA models for seasonal adjustment is associated with several significant practical challenges:

1. **Need to review a large number of series**: The promoting services handle hundreds of series that need to be seasonally adjusted.
2. **Lack of time and resources**: The review process is carried out in a short period and alongside other production tasks.
3. **Impact of COVID**: COVID has significantly complicated the identification of suitable models, increasing the time required for the process.

INE

Automatic identification: TRAMO

Due to the aforementioned issues, automated methods for identifying regARIMA models are necessary.

Until now, the best available identification tool was TRAMO:

- It identifies a single model...
- ... and sometimes this model is not acceptable (e.g., it fails hypothesis tests).
- TRAMO does not test all possible models.
- TRAMO does not consider indicators related to canonical decomposition or seasonal adjustment.

TEAM (Time-Series Exhaustive Automatic Modelling)

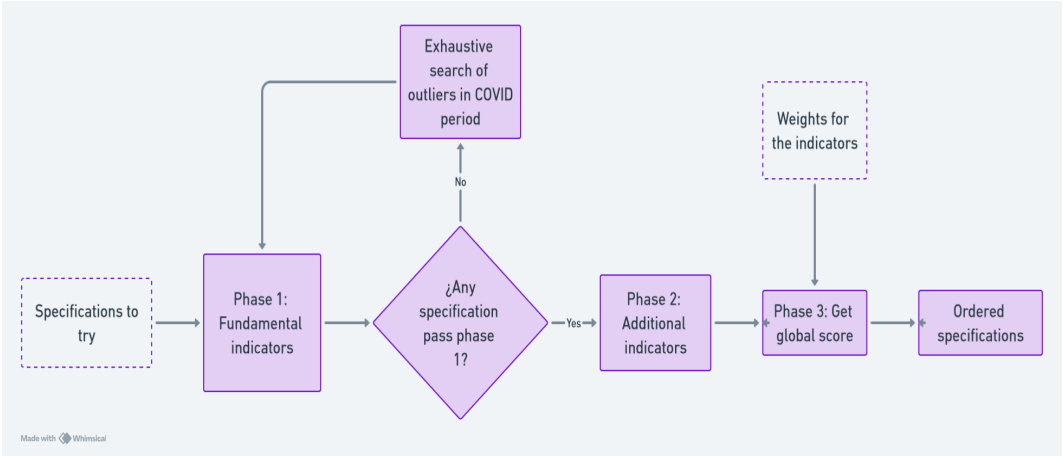
TEAM: Basic Principles

- **Objective**: Address TRAMO's shortcomings to provide a (semi-)automated identification tool with a robust methodology that ensures optimal model selection.

The basic principles behind the design of TEAM's methodology, which differentiate it from TRAMO, are:

- **Exhaustive model search**: All possible models (within user-defined combinations) are tested, and then ranked according to its quality with respect to seasonally adjustment.
- **Offering alternative models**: TEAM provides several models (5 by default) instead of a single one.
- **User adjustable**: Users can specify the models to be tested and prioritize certain quality dimensions over others.

TEAM: Methodology



TEAM: Implementation

- Implementation of TEAM in an **R package**.
- Uses the **JDemetra+ ecosystem** (both **rjdverse** and the Java code via **rJava**) for model estimation and obtaining quality indicators. Also contains some high-level functions for manipulating workspaces.
- **Modularity**: The estimation of models is separated from the ranking. Easy to replace the methodology developed for TRAMO-SEATS with other methodologies (e.g., for X11).
- **Parallelizability**: Each model can be processed independently. TEAM supports parallelism both at the local level (using multiple processor cores) and at the cluster level (using multiple computers).
- Execution times: In a workspace with 24 quarterly series from 1995 to 2023, it takes 23.85 minutes to test 1297 models per series (approximately 0.045 seconds per model).

Practical example

Example

```
1  library(team)
2
3  # Folder path
4  path <- "C:/Users/U852244/Documents/TEAMnew/CuentasSept2024"
5
6  source(file.path(path,"RankMethod.R")) # Function with the methodology
7  wspath <- file.path(path, "/wk_30024.xml") # JDemetra+ workspace
8
9  #### 1. Manipulation of workspaces ####
10
11 # Cleaning of outliers from 2020Q1
12 jws <- rjdemetra3::.jws_open(wspath)
13 sincedate <- as.Date("2020-01-01")
14 jSAItems <- team::wsjItems(jws)
15 specs <- team::tramospecs(jSAItems)
16 specs.noOu <- lapply(specs, team::remove.outliers, dat.begin=sincedate)
17
18 # Automatic outlier detection from 2020Q1
19 specs.noOu <- lapply(specs.noOu, function(x) team::apply.shift.tramo(
20   team::outlier.shift(T, 0, ALTSMask = "1100", span="From,2020-01-01"),x))
21 team::tramospecs(jSAItems) <- specs.noOu
22 team::wsjItems(jws) <- jSAItems
23 rjdemetra3::.jws_compute(jws)
24
```

Example

```
25
26 #### 2. Establish the specifications TEAM is going to try ####
27
28 # ARIMA orders
29 arord <- team::arima.shift(p=c(0,1,2), d=c(0,1,2), q =c(0,1,2), bp=c(0,1), bd=c(0,1), bq=c(0,1))
30 arord <- team::shift.remove(arord, "\\)\\(000\\)")
31 arord <- team::shift.remove(arord, "\\)\\(001\\)")
32
33 # Transformation
34 trf <- transf.shift(Func=c("NONE", "Log"))
35
36 # Trading days
37 trading_days <- team::td.shift(
38   TD= c("TD_NONE", "TDDays"),
39   LY= c("NONE", "LEAPYEAR"))
40
41 # Easter
42 easter <- team::easter.shift(EaOpt=c("NONE", "EAMON"), duration = 8)
43
44 ## All the combinations
45 model_specs <- neutral.shift() %U% (trf %X% trading_days %X% arord %X% easter)
46
47 result <- team::ws_shift.result(jws %:% 1:10,
48                                 shift=model_specs,
49                                 workers=NA,
50                                 weights = c(1,2,1,2,1)
51                                 )
52
```

Example

```
53 #### 3. Save result in a new workspace ####
54
55 rjdemetra3::save_workspace(result, file.path(path,paste0(wsname, "/wk_30024_result.xml")))
56 SetSA(file.path(path,paste0(wsname, "/wk_30024_result.xml")))
57 team::ws3.tov2(file.path(path,paste0(wsname, "/wk_30024_resultv2.xml")))
```

Results

	Serie	Score	Modification
1	Hoja1 PIB_demanda	0.8240204	(210)(011) TD=TD_NONE LY=NONE Func=Log EaOpt=EAMON duration=8
2	Hoja1 PIB_demanda	0.7994955	(102)(110) TD=TD_NONE LY=NONE Func=Log EaOpt=EAMON duration=8
3	Hoja1 PIB_demanda	0.7797443	(012)(010) TD=TDDays LY=NONE Func=Log EaOpt=EAMON duration=8
4	Hoja1 PIB_demanda	0.6578215	(210)(010) TD=TDDays LY=NONE Func=Log EaOpt=EAMON duration=8
5	Hoja1 PIB_demanda	0.6554632	(102)(010) TD=TD_NONE LY=NONE Func=NONE EaOpt=NONE
6	Hoja1 PIB_oferta	0.8240204	(210)(011) TD=TD_NONE LY=NONE Func=Log EaOpt=EAMON duration=8
7	Hoja1 PIB_oferta	0.7994955	(102)(110) TD=TD_NONE LY=NONE Func=Log EaOpt=EAMON duration=8
8	Hoja1 PIB_oferta	0.7797443	(012)(010) TD=TDDays LY=NONE Func=Log EaOpt=EAMON duration=8
9	Hoja1 PIB_oferta	0.6578215	(210)(010) TD=TDDays LY=NONE Func=Log EaOpt=EAMON duration=8
10	Hoja1 PIB_oferta	0.6554632	(102)(010) TD=TD_NONE LY=NONE Func=NONE EaOpt=NONE
11	Hoja1 PIB_promedio	0.8240204	(210)(011) TD=TD_NONE LY=NONE Func=Log EaOpt=EAMON duration=8
12	Hoja1 PIB_promedio	0.7994955	(102)(110) TD=TD_NONE LY=NONE Func=Log EaOpt=EAMON duration=8
13	Hoja1 PIB_promedio	0.7797443	(012)(010) TD=TDDays LY=NONE Func=Log EaOpt=EAMON duration=8

Future developments

Future of TEAM

- Finalize the development and implementation of robust strategies for outlier detection during the COVID period, particularly for monthly series.
- Dissemination of the open-source R package via the institutional GitHub of Statistics Spain.
- Better integration of TEAM with the Java code of JDemetra+, so it can be faster.

INE

Thank you for your attention!