X13-ARIMA-SEATS

Rohit Jangid

July 2024

Contents

1 Introduction

The Ministry of Statistics and Programme Implementation (MoSPI) is responsible for assembling and sharing key country indexes such as the Gross Domestic Product (GDP) and the Consumer Price Index (CPI). GDP represents the overall economic activity, while CPI measures inflation trends. Both indicators provide critical insights into a country's economic performance. Seasonal variations, due to factors like holidays and agriculture, often obscure the true trends. This necessitates **seasonal adjustment**, a statistical technique to remove these effects and reveal the underlying patterns in time series data.

Currently, MoSPI does not apply seasonal adjustments to its indices. This project's goal is to address the need for seasonal adjustment techniques using the widely adopted X-13-ARIMA-SEATS model, developed by the U.S. Census Bureau. Our objective is to provide clarity on the technical details of this model and equip MoSPI with the tools for correct implementation.

2 Problem Summary

Economic indicators such as GDP and CPI are often influenced by seasonal factors like holidays, agricultural cycles, and other periodic events. These factors introduce systematic patterns in the data, which, if not accounted for, can mislead decision-makers about the actual trends in the economy. For instance, a rise in consumer spending around holidays may inflate the perception of economic growth, which is temporary.

The **X-13-ARIMA-SEATS** model is a powerful statistical tool designed to remove such seasonal effects. It is based on ARIMA (AutoRegressive Integrated Moving Average) modeling combined with signal extraction techniques. This model enables users to separate the trend, seasonal, and irregular components from time series data.

Our project aims to:

- Provide a detailed technical explanation of the X-13-ARIMA-SEATS model for MoSPI.
- Test the appropriateness of the model on Indian economic data.
- Deliver an R package to apply the model on Indian data for seasonal adjustment.

We have made significant progress in understanding the ARIMA framework and its integration with signal extraction for seasonal adjustments. We are currently testing the model on sample datasets and preparing the technical document for submission to MoSPI (See [1], page 233).

3 X-13ARIMA-SEATS

The X-13ARIMA-SEATS is a comprehensive seasonal adjustment software developed by the U.S. Census Bureau. It is widely used for time series analysis to remove seasonal variations from data and identify underlying trends. The software integrates two powerful algorithms: the X-11 method and SEATS (Signal Extraction in ARIMA Time Series). These techniques allow users to adjust for seasonal effects and extract trend and irregular components.

3.1 SEATS and X-13

SEATS is a model-based approach that uses ARIMA modeling to decompose a time series into its trend, seasonal, and irregular components. It was originally developed by Agustin Maravall and Victor Gómez at the Bank of Spain. SEATS provides a statistically rigorous way to account for seasonal patterns and produce smoother trends (See X-13ARIMA-SEATS Manual, page 2).

X-13, on the other hand, builds on the earlier X-11 and X-12 methods, which use filter-based techniques for seasonal adjustment. The X-13ARIMA-SEATS software combines the strengths of both approaches, giving users flexibility in choosing between model-based and filter-based seasonal adjustments (See X-13ARIMA-SEATS Manual, page 2).

3.2 regARIMA and TRAMO

The regARIMA model used in X-13ARIMA-SEATS is based on TRAMO (Time Series Regression with ARIMA Noise, Missing Observations, and Outliers). TRAMO handles preprocessing tasks such as outlier detection, modeling calendar effects, and managing missing observations. It fits a regression model with ARIMA errors to clean the data before applying seasonal adjustment. The software is able to extend time series data through forecasting and backcasting, improving the accuracy of seasonal adjustments near the boundaries of the data (See X-13ARIMA-SEATS Manual, page 5).

3.3 Using X-13ARIMA-SEATS

The software can be executed using the command line by specifying the path to the input specification (spec) file. A generic command to run X-13ARIMA-SEATS is:

```
path\x13as path\filename
```

The spec file contains the necessary instructions for running the software, including details about the time series data, ARIMA models, and output preferences. It is a simple text file with a ‘.spc‘ extension.

3.4 Spec File and Key Specifications

A spec file consists of various specifications (specs) that control the flow of execution. Here is an example spec file for demonstration purpose.

```
series{
  title = "Consumer Food Price Index - All India Combined"
  start = 2013.01
  span = (2013.01, 2024.08)
  data = (
    105.4 106.4 106.5 107.5 109.1 112.4 115.2 117.3 119.0 121.1 123.9 118.7
    115.6 114.8 115.7 117.4 118.8 120.5 125.4 127.5 126.4 125.8 125.3 123.4
    122.7 122.7 122.8 123.4 124.5 127.1 128.1 130.3 131.3 132.4 132.9 131.3
    131.1 129.2 129.2 131.3 133.8 137.0 138.8 138.0 136.5 136.8 135.6 133.1
    131.9 131.8 131.8 132.1 132.4 134.1 138.3 140.1 138.2 139.4 141.5 139.7
    138.1 136.1 135.5 135.8 136.5 138.0 140.1 140.5 138.9 138.2 137.8 136.0
    135.0 135.1 135.9 137.3 139.0 141.1 143.4 144.7 146.0 149.1 151.6 155.3
    153.4 149.7 147.8 153.4 151.8 153.4 156.7 157.8 161.6 165.5 166.0 160.6
    156.4 155.5 155.0 156.4 159.4 161.3 162.9 162.7 162.7 166.9 169.1 167.1
    164.9 164.6 166.9 169.4 172.1 173.8 173.8 175.1 176.7 178.6 177.0 174.1
    174.8 174.4 174.9 175.9 177.2 181.7 193.8 192.5 188.4 190.4 192.4 190.7
    189.3 189.5 189.8 191.2 192.6 198.7 204.3 203.4
  )
}

transform{
  function = auto
}

automdl{maxorder = (3, ) }

outlier{types = (ls ao)}

estimate{
    save = residuals
}

regression {
  variables = (const, td)
  user = (diwali)
  start = 2013.01
  data = (
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0405 -0.0405  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.3405 -0.3405  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 -0.6595  0.6595  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.3405 -0.3405  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.3405 -0.3405  0.0
```

```
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 -0.3595  0.3595  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.3405 -0.3405  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 -0.6595  0.6595  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 -0.0595  0.0595  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.3405 -0.3405  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 -0.6595  0.6595  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.3405 -0.3405  0.0
   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.3405 -0.3405  0.0)
print = rmx
}

seats{}
```

3.4.1 SERIES

The SERIES spec defines the input time series data for analysis. It includes details such as the starting and ending dates, frequency, and data type. This spec is crucial as it tells the software what time series to use for seasonal adjustment and other analyses. The SERIES spec also allows users to set the precision of the data and handle missing values appropriately (See X-13ARIMA-SEATS Manual, page 181).

3.4.2 AUTOMODEL

The AUTOMODEL spec automates the selection of the best-fitting ARIMA model by comparing several candidate models and selecting the one with the lowest AIC (Akaike Information Criterion). This is particularly useful when the user is unsure of the correct model structure. The procedure is closely based on TRAMO, a method developed by Gómez and Maravall (2000) (See X-13ARIMA-SEATS Manual, page 50).

The method is detailed in Gómez and Maravall's paper, "Automatic Modeling Methods for Univariate Series" in *A Course in Time Series*, edited by D. Pena, G. C. Tiao, and R. S. Tsay, New York: J. Wiley and Sons, 2000.

3.4.3 Estimate

This spec controls the estimation method used for fitting the ARIMA model. Maximum likelihood estimation (MLE) is typically used, but users can modify various estimation settings such as the number of iterations and convergence criteria (See X-13ARIMA-SEATS Manual, page 100).

3.4.4 Outlier

The Outlier spec detects and adjusts for outliers in the data. Three types of outliers are commonly detected: Additive Outliers (AO), Level Shifts (LS), and Transitory Changes (TC). The software automatically identifies these outliers

and includes them as regressors in the model (See X-13ARIMA-SEATS Manual, page 133).

3.4.5 Regression

The Regression spec allows users to specify the regression variables used in the regARIMA model. These variables can include predefined effects, such as trading day and holiday adjustments, or user-defined regressors (See X-13ARIMA-SEATS Manual, page 144).

3.4.6 Transform

The Transform spec handles transformations of the data, such as logarithmic transformations, to stabilize the variance or make the series more stationary. This is particularly useful when dealing with time series that exhibit non-constant variance (See X-13ARIMA-SEATS Manual, page 212).

3.4.7 SEATS

The SEATS spec controls the use of the SEATS algorithm for seasonal adjustment. This includes options for model selection, output settings, and diagnostics for the decomposition of the time series into its components (See X-13ARIMA-SEATS Manual, page 169).

3.4.8 X-11

The X-11 spec is used when applying the X-11 seasonal adjustment method. This spec provides options for controlling seasonal filters, diagnostics, and the handling of trading day effects (See X-13ARIMA-SEATS Manual, page 223).

4 Program Workflow

The X-13ARIMA-SEATS program operates in several stages to transform raw time series data into its seasonal, trend, and irregular components. The main steps involve reading the raw data, applying the regARIMA model for transformations, handling outliers and missing values, and then using either SEATS or X-11 to extract the seasonal and other components.

4.1 Workflow Explanation

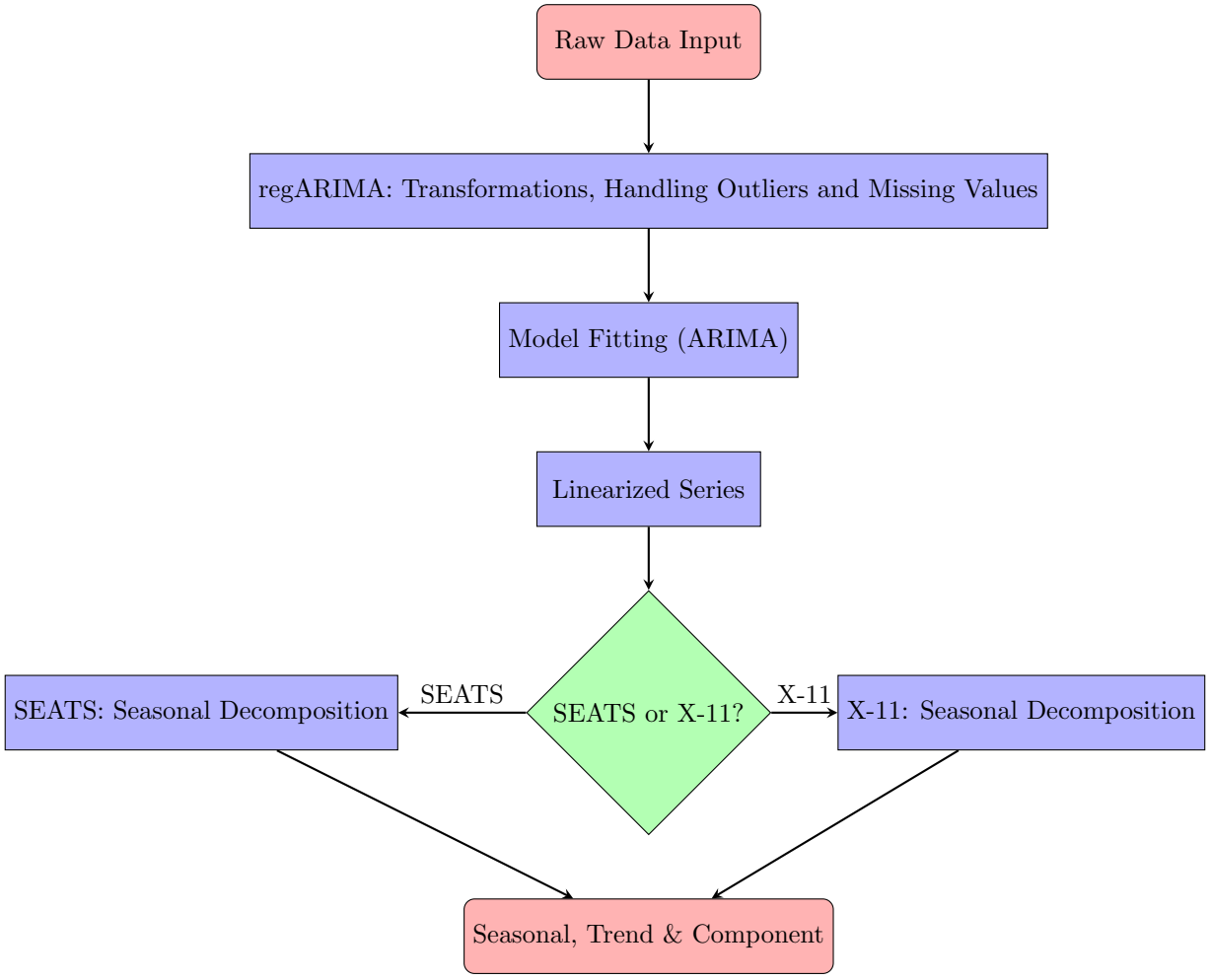
The general workflow of the X-13ARIMA-SEATS program can be broken down as follows:

- **Raw Data Input**: The process begins with inputting the raw time series data, which is defined in the ‘series‘ spec of the specification file.

- **regARIMA Model**: The regARIMA model is applied to the data. This involves:
  - **Transformations**: Automatically or manually defined transformations (e.g., logarithmic or differencing).
  - **Handling Outliers and Missing Values**: Identifies and adjusts outliers, and fills missing data where applicable.
  - **Model Fitting**: Fits an ARIMA model to the transformed data, accounting for these adjustments.

  The output from the regARIMA model is the linearized series, which is free of seasonal effects, outliers, and missing values.
- **Seasonal Adjustment**: This linearized series is passed through either the SEATS or X-11 algorithm to extract the seasonal, trend, and irregular components.
  - **SEATS**: A model-based approach for decomposing the series.
  - **X-11**: A filter-based approach for seasonal decomposition.
- **Final Output**: The program provides the seasonal component, trend component, and irregular component of the time series.

4.2 Flowchart of the Workflow



5 RegARIMA: Model Process and Workflow

The **RegARIMA** model in the X-13ARIMA-SEATS software is an advanced tool for handling regression effects, detecting outliers, and fitting ARIMA models to time series data. The process includes several key steps, each involving tests and decisions based on the time series characteristics.

5.1 Steps Involved in the RegARIMA Process

The RegARIMA model is executed in the following major steps:

5.1.1 1. Transformation Step

The first step in the RegARIMA process is to determine whether a transformation (such as a logarithmic transformation) is necessary to stabilize the variance of the time series. This decision is often based on statistical tests such as the Akaike Information Criterion (AIC) and the logarithmic likelihood of the model.

If the transformation is set to 'auto', the software evaluates whether a logarithmic transformation is beneficial. Otherwise, the user can specify a transformation manually. - **Example**: Logarithmic transformation is often chosen if the series exhibits multiplicative effects (See page 75, X-13ARIMA-SEATS Manual).

5.1.2 2. Initial Outlier Detection

In this step, the software scans the series for potential outliers using statistical tests. Outliers such as Additive Outliers (AO), Level Shifts (LS), and Transitory Changes (TC) are identified. These outliers are flagged and temporarily adjusted for in the data.

- **Example**: An AO might be detected due to a sudden, sharp spike in the data, and this is treated before proceeding with model identification (See page 102, X-13ARIMA-SEATS Manual).

5.1.3 3. Automatic Model Identification

After the initial outliers are handled by using appropriate regression variables, the software identifies the best-fitting ARIMA model using criteria such as the AIC, BIC, and Hannan-Quinn. The procedure is based on the TRAMO method (Gómez and Maravall, 2000), which selects a model by considering seasonal and non-seasonal components.

- **Model Structure**: The software tests several ARIMA configurations to identify the one that best captures the underlying data patterns. A typical model might take the form $(p, d, q)(P, D, Q)_s$ where p and P represent autoregressive terms, d and D represent differencing orders, and q and Q represent moving averages (See page 123, X-13ARIMA-SEATS Manual).

5.1.4 4. Final Outlier Detection

Once the ARIMA model is selected, the software performs a second outlier detection step. This step checks whether any additional outliers have emerged after the model was fitted. These outliers are also incorporated as regressors in the final model.

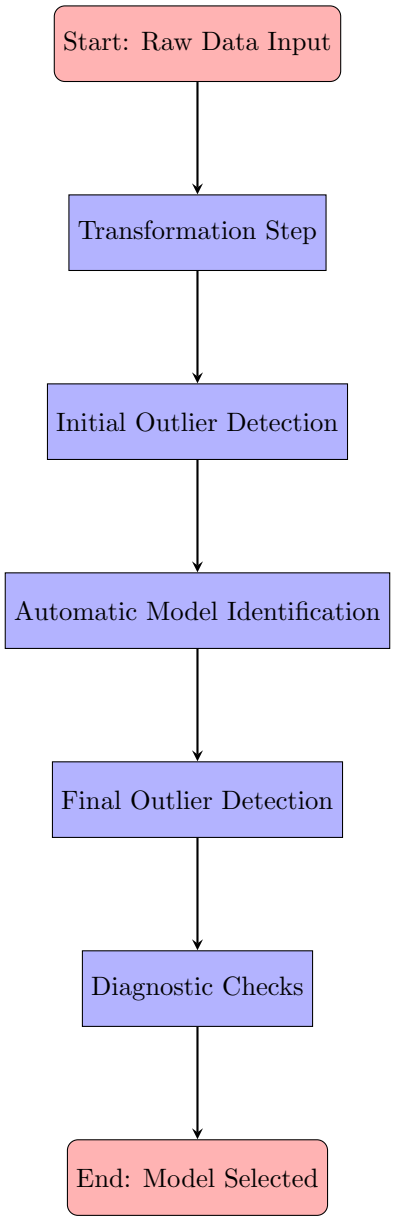
- **Example**: Outliers detected in this step are included in the regression equation to prevent them from distorting the seasonal adjustment (See page 134, X-13ARIMA-SEATS Manual).

5.1.5 5. Diagnostic Checks for the Model

Finally, the selected ARIMA model undergoes diagnostic checks to ensure that it adequately fits the data. The model's residuals are examined for normality and autocorrelation. If the model fails these diagnostic tests, adjustments are made.

- **Example**: The residual autocorrelation function (ACF) is checked to ensure that no significant patterns remain in the residuals, confirming that the model has successfully captured the data's underlying structure (See page 145, X-13ARIMA-SEATS Manual).

5.2 Flowchart of the RegARIMA Process



This flowchart illustrates the key decisions and tests made during the RegARIMA process. Each step ensures that the model correctly captures the structure of the time series while handling outliers and performing diagnostics.

References

[1] U.S. Census Bureau, "X-13ARIMA-SEATS Reference Manual," June 7, 2024. Available: `https://www.census.gov/data/software/x13as.html`.