Understanding X13-ARIMA-SEATS Models

Rohit Jangid (210872)

Indian Institute of Technology Kanpur

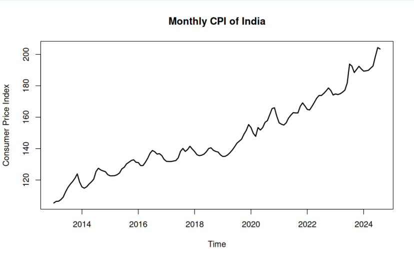
November 7, 2024

Mentor: Prof. Dootika Vats

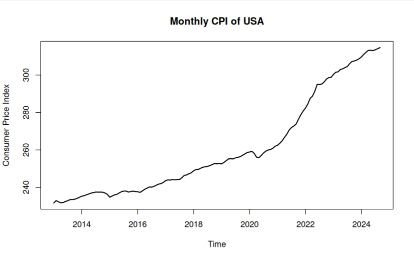
Acknowledgment:
Special thanks to the authors and resources referenced, including foundational work on X13-ARIMA-SEATS by the U.S. Census Bureau, Agustin Maravall, and Victor Gómez.

Motivation

Monthly Consumer Price Index: India vs USA



Indian Monthly CPI



USA Monthly CPI

Observing seasonal and trend differences in the Consumer Price Index (CPI) among these countries.

Consumer Price Index and Seasonal Adjustments

What is the Consumer Price Index (CPI)?

- Measures the average change in prices over time for a basket of goods and services.
- Key economic indicator for assessing inflation and guiding policy decisions.

Importance of Seasonal Adjustment

- Removes recurring fluctuations due to holidays, climate, and agricultural cycles.
- Reveals underlying economic trends, aiding in data interpretation over time.

X13-ARIMA-SEATS Software

- Developed by the U.S. Census Bureau for seasonal adjustments.
- Widely adopted in countries like the USA, Canada, Australia, and Spain.

India's Adoption Efforts

- MoSPI is exploring seasonal adjustments for CPI using X13-ARIMA-SEATS.
- Challenges due to India's unique festivals and economic cycles may require customization.

Project Goal and Expected Outcomes

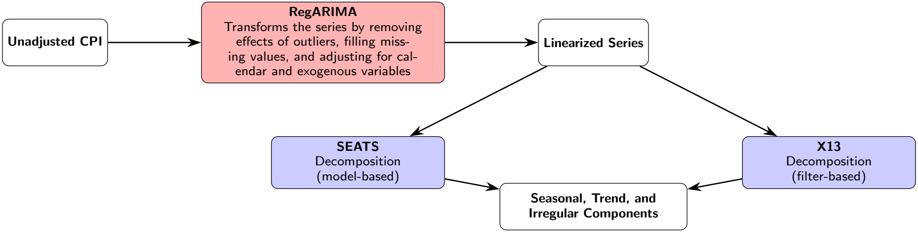
Objective of the Project

- Evaluate X13-ARIMA-SEATS applicability in the Indian context.
- Adapt the software to India's unique economic and cultural calendar.
- Advise MoSPI on implementing seasonal adjustment techniques for accurate CPI reporting.

Expected Outcomes

- Detailed evaluation report of X13-ARIMA-SEATS for Indian data.
- Customized R wrapper for X13-ARIMA-SEATS tailored to India.
- Guidance for MoSPI on using the software to improve economic indicator interpretation.

Process of Seasonal Extraction in X13-ARIMA-SEATS



Progress and Focus

- The current progress has covered the **RegARIMA** component, transforming the series by handling outliers, calendar effects, missing values, and exogenous regression variables.
- **SEATS** and **X13** decomposition methods, which provide seasonal, trend, and irregular components, will be the focus of future work.

RegARIMA Model

Overview

- Combines linear regression with ARIMA errors.
- Models both deterministic structures (e.g., calendar effects) and stochastic dependencies.

Mathematical Formulation

$$y_t = X_t \beta + z_t$$

where:

- y_t: Observed time series at time t.
- X_t: Vector of regression variables (e.g., calendar effects, external regressors).
- β: Vector of coefficients corresponding to the regression variables.
- z_t: Error term following an ARIMA(p, d, q) process.

Complete RegARIMA Model

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D(y_t - X_t\beta) = \theta(B)\Theta(B^s)a_t$$

- Models both deterministic and stochastic components in the time series.

Workflow of RegARIMA

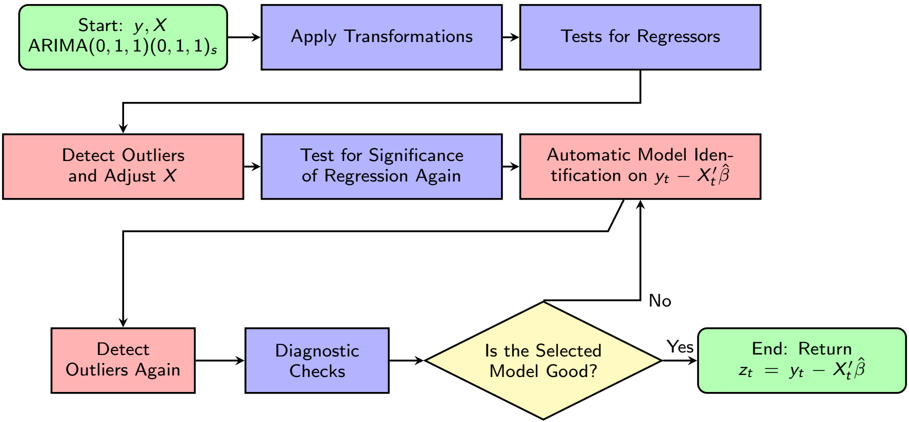


Figure: Flowchart Illustrating the RegARIMA Process

Transformation in RegARIMA

Function Transformation

Value	Transformation	Range for Y_t
none	Y_t	All values
log	$\log(Y_t)$	$Y_t > 0$
sqrt	$0.25 + 2(\sqrt{Y_t} - 1)$	$Y_t \geq 0$
inverse	$2 - \frac{1}{Y_t}$	$Y_t \neq 0$
logistic	$\log\left(\frac{Y_t}{1-Y_t}\right)$	$0 < Y_t < 1$

User-Specified Adjustments

$$f(x_i) = \frac{x_i}{C_i} \quad \text{or} \quad f(x_i) = x_i - C_i$$

where C_i are user-defined adjustment factors.

Box-Cox Power Transformation

$$y_t = \begin{cases} \log(Y_t) & \text{if } \lambda = 0 \\ \lambda^2 + \frac{Y_t^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \end{cases}$$

Length of Month (LoM) Adjustment

$$C_t = \frac{m_t - \bar{m}}{\bar{m}}$$

where:

- m_t: number of days in month t.
- $\bar{m} = 30.4375$: average month length.

Testing for Log Transformation and Jacobian Adjustment

Objective:

- Determine whether a log transformation is appropriate for the time series data.
- Ensure comparability of log-likelihoods between transformed and untransformed models.

Procedure for Testing Log Transformation:

1. **Model Fitting:** Default ARIMA$(0,1,1)(0,1,1)_s$ model on both untransformed and log-transformed series with all user-specified regressors.
2. **Calculate AICC:** Corrected Akaike Information Criterion (AICC) for both models.
3. **Compare AICC Values:** $\text{AICC}_{\text{nolog}} - \text{AICC}_{\text{log}} \leq \text{AICC}_{\text{diff}}$
    - If the condition is met, favor the log transformation.
    - Negative $\text{AICC}_{\text{diff}}$ values favor the log transformation.

Calculation of AICC: $\text{AICC}_N = -2L_N + 2n_p\left(\frac{N}{N-n_p-1}\right)$

Log-Likelihood Adjustment After Transformation:

$$L_N^* = \tilde{L}_N + \sum_{t=1}^{N} \log\left|\frac{df(Y_t)}{dY_t}\right|$$

- $\tilde{L}_N$: Log-likelihood for the transformed data.
- The summation term adjusts for the Jacobian of the transformation.
- Ensures comparability of log-likelihoods between transformed and untransformed models.

Regression Variables in RegARIMA

Purpose of Regression Variables

- Adjust for effects external to the time series' inherent properties.
- Model deterministic factors such as:
  - Trading day effects.
  - Holiday effects.
  - Outliers.
  - User-defined variables.
- Enhance model accuracy by accounting for known patterns and anomalies.

Types of Regression Variables

- **Trading Day Variables**
- **Holiday Effect Variables**
- **Length-of-Month/Quarter and Leap Year Variables**
- **Outlier Regressors**
- **User-Defined Regressors**

Trading Day Regressor: 6-Variable Model

Purpose

- Capture the individual effects of each weekday on economic activity.

Mathematical Formulation

$$T_t = (\text{Number of days of type } i \text{ month t}) - (\text{Number of Sundays month t}),$$

$$i = \text{Mon}, \text{Tue}, \ldots, \text{Sat}$$

Explanation

- **Contrasts Each Weekday with Sunday**: Allows for different coefficients for each weekday.
- Accounts for Specific Patterns: For instance, retail sales might be higher on Saturdays.

Example

- **In a Month**:
  - Mondays: 4
  - Sundays: 5
  - $T_{\text{Mon}} = 4 - 5 = -1$

Usage

- Suitable when different weekdays have distinct effects on the time series.

Trading Day Regressor: 1-Variable Model

Purpose

- Capture the effect of the varying number of weekdays and weekends in each month.

Trading Day Variable (`td1coef`)

- Represents the contrast between weekdays and weekends.
- Assumes all weekdays have a similar effect; Saturdays and Sundays have a different but consistent effect.

Mathematical Formulation

$$T_t = (\text{Number of weekdays}) - \frac{5}{2} \times (\text{Number of Saturdays and Sundays})$$

Explanation

- The factor $\frac{5}{2}$ balances the influence of weekends relative to weekdays.

Example

- **In a Month with 22 Weekdays and 8 Weekend Days**:

$$T_t = 22 - \frac{5}{2} \times 8 = 22 - 20 = 2$$

Usage

- Suitable when individual weekdays do not have significantly different

Holiday Effect Variables: Diwali Regressor

Purpose

- Capture the impact of Diwali on economic activities in India.
- Adjust for holidays that shift dates each year due to lunar calendars.
- Similar approach can be used for other movable holidays.

Creating the Diwali Regressor

1. **Define the Affected Period**:
   - Choose w days before and/or after Diwali reflecting its economic impact.
2. **Calculate Monthly Proportions**:
   - For each year, compute the proportion of the w-day window falling in each month.
3. **Construct the Regressor**:
   - Use these proportions as regressor values for corresponding months.
4. **Deseasonalize the Regressor**:
   - Subtract long-term monthly means to isolate the Diwali effect.
5. **Include in the Model**:
   - Add as a user-defined variable in the `regression` spec.

November

M	T	W	T	F	S	S
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	1
2	3	4	5	6	7	8

Outlier Detection

Outliers in X13-ARIMA-SEATS

- Many types of outliers can be specified.
- Some of the outlier types can be detected automatically.
- Users can give potential time points to test for the presence of other outlier types.

Outliers that can be detected automatically

- **Additive Outlier (AO)**:
  - An unexpected high or low value at a specific time point.
- **Level Shift (LS)**:
  - An abrupt change in the mean level of the series.
- **Temporary Change (TC)**:
  - A temporary shift in the mean that decays back to the previous level over time.

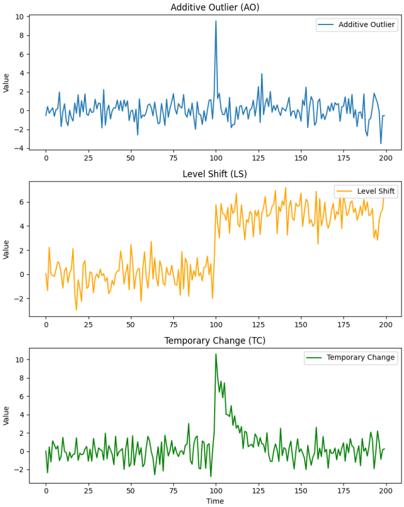


Figure: Illustration of Outlier Types

Incorporating Outliers Using Regressors

The regressors in the regression matrix

	$AO1953.02$	$LS1953.04$	$TC1953.03$
1953.01	0	-1	0
1953.02	1	-1	0
1953.03	0	-1	1
1953.04	0	0	0.5
1953.05	0	0	0.25
1953.06	0	0	0.125

- here $\alpha = 0.5$.
- When an outlier is detected, it can be adjusted by using the corresponding regressor in the model.

Additive Outlier at time t_0

$$AO_t^{(t_0)} = \begin{cases} 1, & \text{if } t = t_0 \\ 0, & \text{otherwise} \end{cases}$$

Level Shift at time t_0

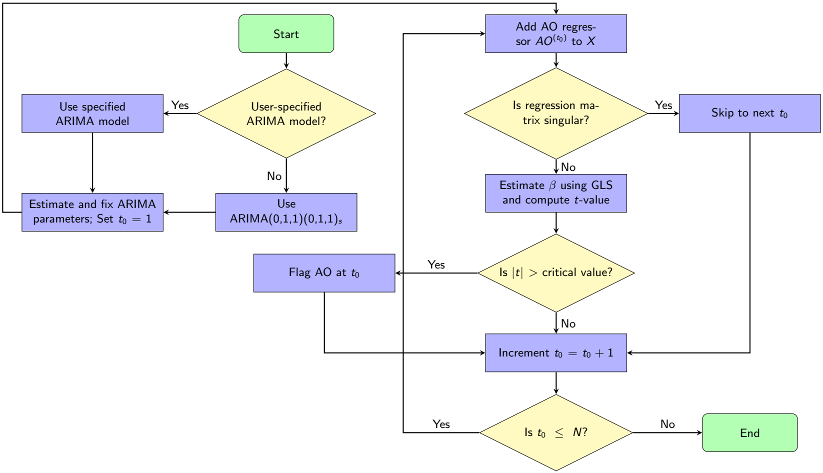
$$LS_t^{(t_0)} = \begin{cases} -1, & \text{if } t < t_0 \\ 0, & \text{if } t \geq t_0 \end{cases}$$

Temporary Change at time t_0

$$TC_t^{(t_0)} = \begin{cases} 0, & \text{if } t < t_0 \\ \alpha^{(t-t_0)}, & \text{if } t \geq t_0 \end{cases}$$

- α: Decay rate (default $\alpha = (0.7)^{1/s}$, s = seasonal period).

Outlier Flagging Process



Methods of Outlier Detection

AddOne Method

- Default method for outlier detection in X13-ARIMA-SEATS.
- Process:
  1. **Step 1**: Flag all potential outliers using current regression variables.
  2. **Step 2**: Add the most significant outlier to the model.
  3. **Step 3**: Repeat Steps 1-2 until no new outliers are detected.
  4. **Step 4**: Remove previously added outliers that have become insignificant.
  5. **Step 5**: Repeat Step 4 until all outliers are significant.

AddAll Method

- Alternative method for outlier detection.
- Differences from AddOne:
  - In the forward pass, all flagged outliers are added to the model at once.
  - In the backward pass, insignificant outliers are removed one at a time.

AddOne Method

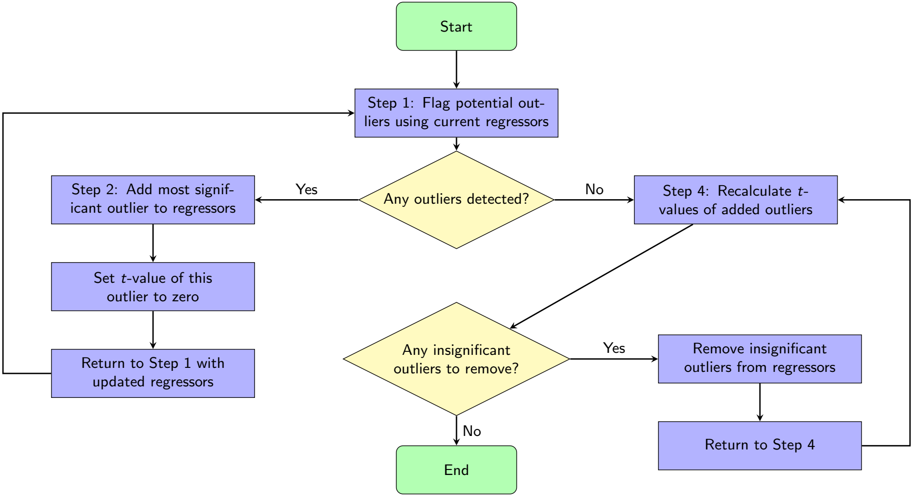


Figure: Flowchart of the AddOne Method for Outlier Detection

Automatic Model Selection in X13-ARIMA-SEATS

Outline of Automatic Model Selection Procedure

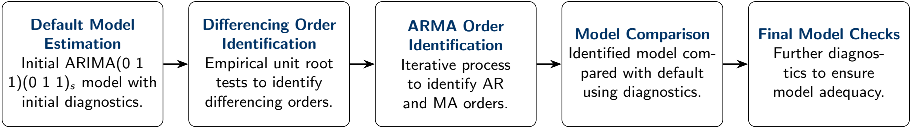


Figure: Flowchart of the Automatic Model Selection Procedure in X13-ARIMA-SEATS

Default Model Tasks

- Checks significance of trading day, Easter, and user-defined regressors using AICC criteria.
- Tests for inclusion of a constant term via t-statistic with critical value of 1.96.
- Conducts automatic outlier detection if specified.

Residual Diagnostics

- Ljung-Box Q statistic (lag 24 for monthly, lag 16 for quarterly series).
- Confidence coefficient $(1 - p\text{-value})$ of Ljung-Box Q statistic.
- t-value for mean of residuals and residual standard error estimate.

Model Identification Process

- Regression effects from the default model are removed.
- Adjusted *linearized series* is used for robust model identification.

Identification of Differencing Orders

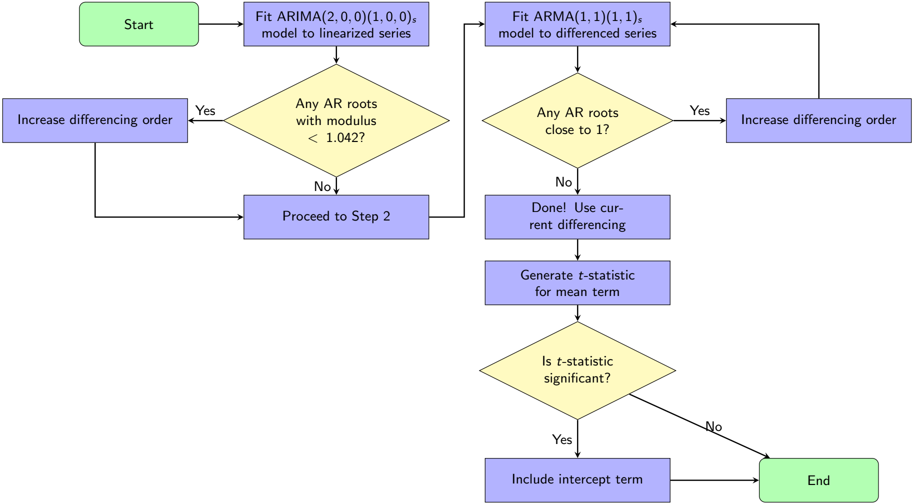


Figure: Flowchart for Identifying Appropriate Differencing Orders in ARIMA Modeling

Identification of ARMA Model Orders

- **Goal**: Select the best-fitting ARMA model by comparing candidate models based on Bayesian Information Criterion (BIC).
- **Criterion for Model Selection**: BIC2, a variant of BIC, is calculated for each model.

Three-Stage Procedure:

1. **Stage 1: Initial Seasonal Order Selection**
   - Consider ARIMA$(3, d, 0)(P, D, Q)_s$ models.
   - Select seasonal orders (P, Q) based on the lowest BIC2 value.
2. **Stage 2: Nonseasonal Order Selection**
   - Using selected (P, Q) from Stage 1, consider ARIMA$(p, d, q)(P, D, Q)_s$ models.
   - Select nonseasonal orders (p, q) based on the lowest BIC2 value.
3. **Stage 3: Refinement of Seasonal Orders**
   - Reconsider ARIMA$(p, d, q)(P, D, Q)_s$ with updated (p, q), refining (P, Q) to further minimize BIC2.

Model Selection and Final Checks:

- Track models with the five lowest BIC2 values, and select the most parsimonious and balanced model if alternatives are "close enough."

Identification of ARMA Model Orders

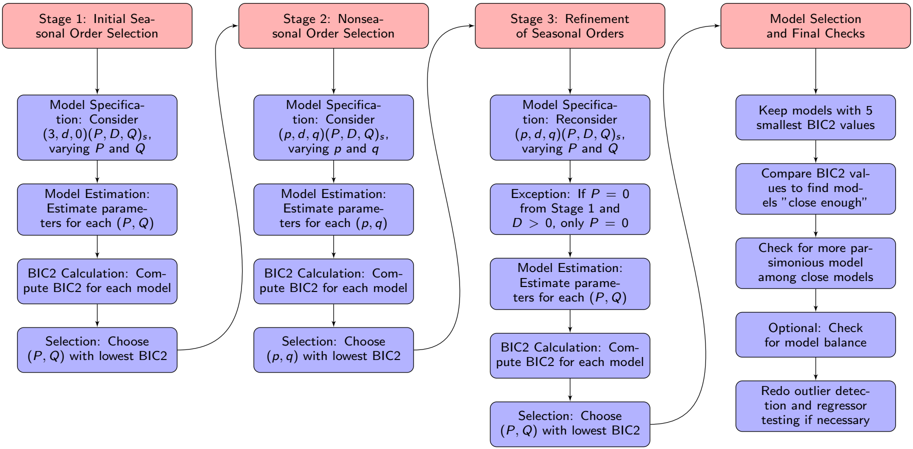


Figure: Flowchart for Identification of ARMA Model Orders

Final Selection of the ARIMA Model

- **Model Comparison**:
  - Compare automatic and default ARIMA models using residual diagnostics:
    - Ljung-Box Q statistics (Q_A, Q_D)
    - Residual standard errors (RSE_A, RSE_D)
    - Number of identified outliers
  - Prefer the default model if it meets specific criteria (better Q values, fewer outliers, lower residual error)
- **Model Acceptability and Adjustment**:
  - If residual autocorrelation is detected ($Q > 0.975$), reduce the critical value for outlier detection
  - Re-estimate the model and re-identify outliers using the adjusted critical value ($\text{CV}_r = (1 - \text{reducecv}) \times \text{CV}$)
- **Re-evaluation of Regressors**:
  - Test the significance of regressors (e.g., trading day effects, Easter effects, constant term) . Use t-statistics ($t > 1.96$) to determine significance
- **Final Model Checks**:
  - **Check for Unit Roots**:
    - In AR polynomials (modulus ≤ 1.05); adjust differencing order if necessary
    - In MA polynomials (sum of coefficients ≈ 1); adjust model accordingly
  - **Re-estimation and Outlier Detection**:
    - Re-estimate the model after adjustments
    - Perform outlier detection again if specified
  - **Simplify the Model**:
    - Remove insignificant ARMA parameters (t-statistic below threshold and small absolute value)
    - Ensure at least one ARMA parameter remains in the model

Conclusion

- **Project Achievements**:
  - Thoroughly understood and documented the functionalities and mathematical foundations of `X13-ARIMA-SEATS`
  - Focused on RegARIMA modeling and seasonal adjustment procedures
  - Assimilated dispersed information into a coherent and comprehensive guide
  - Implemented processes up to outlier detection in R using government CPI data
- **Findings**:
  - The program is very complicated with many moving parts and numerous decisions
  - Many aspects are based on empirical understanding rather than rigorous mathematics
- **Future Work**:
  - Implementing the model selection procedure
  - Understanding the SEATS algorithm and X-13 seasonal adjustment method

Thank You

Thank You for Your Attention!

Special Thanks to:

- **My Mentor:** Prof. Dootika Vats
- **Key Contributors and References:**
  - George E. P. Box and Gwilym M. Jenkins
  - Víctor Gómez and Agustín Maravall
  - Ruey S. Tsay
  - U.S. Census Bureau

Questions and Discussion Welcome