Exploring Interrelationships between Consumer Price Indexes (CPIs)

A Project Report Submitted to the
Department of Computer Science of
Ramakrishna Mission Vivekananda Educational and Research Institute, Belur,
in partial fulfilment of the requirements for the degree of
MSc in Computer Science.

Submitted by
SOURISH GHOSH
B2230036

Supervisor:
Dr. Sudipta Das
Department of Computer Science
Ramakrishna Mission Vivekananda Educational and Research Institute

Department of Computer Science
Ramakrishna Mission Educational and Research Institute
Belur Math, Howrah 711202, West Bengal, India
December 21, 2023

Exploring Interrelationships between Consumer Price Indexes (CPIs)

By

SOURISH GHOSH

Declaration by student:

“I hereby declare that the present dissertation is the outcome of my project work under the guidance of Dr. Sudipta Das and I have properly acknowledged the sources of materials used in my project report."

(Sourish Ghosh, ID No. B2230036)

A project report in the partial fulfilment of the requirements of the degree of MSc in Computer Science

Examined and approved on

by

Dr. Sudipta Das(supervisor)
Department of Computer Science
Ramakrishna Mission Vivekananda Educational and Research Institute

Countersigned by

Registrar
Ramakrishna Mission Vivekananda Educational and Research Institute

Department of Computer Science
Ramakrishna Mission Vivekananda Educational and Research Institute
Belur Math, Howrah 711202, West Bengal, India

Acknowledgement

The present project work is submitted in partial fulfilment of the requirements for the degree of Master of Science of Ramakrishna Mission Vivekananda Educational and Research Institute (RKMVERI). I express my deepest gratitude to my supervisor Dr. Sudipta Das of Ramakrishna Mission Vivekananda Educational and Research Institute (RKMVERI) for his inestimable support, encouragement, profound knowledge, largely helpful conversations and also for providing me a systematic way for the completion of my project work. His ability to work hard inspired me a lot. I am also extremely grateful to the Vice-Chancellor of this University for his encouragement and support throughout the course. Last but not the least, this work would not have been possible without support of my fellow classmates.

Belur
December 21, 2023

Sourish Ghosh
Department of Computer Science
Ramakrishna Mission Vivekananda Educational and Research Institute

Contents

List of Figures

Chapter 1

Introduction

This project focuses on exploring the interrelationships between Consumer Price Indexes (CPIs), a critical economic indicator reflecting changes in the average prices paid by consumers for goods and services. The study employs advanced statistical and time series analysis techniques to uncover patterns, correlations, and potential causal relationships between different CPI components. The investigation encompasses diverse aspects, including regional variations, sector-specific trends, and the impact of external factors on CPI dynamics. By delving into the interconnections of CPIs, this project aims to contribute valuable insights for policymakers, economists, and stakeholders, facilitating a deeper understanding of economic dynamics and informing strategies for inflation management and economic stability.

Chapter 2

Literature Survey

1. **A Prediction Approach for Stock Market Volatility Based on Time Series Data**: This research tries to analyze the time series data of the Indian stock market and build a statistical model that could efficiently predict the future stocks using ARIMA Model.
   Authors: Sheikh Mohammad Idrees, M. Afshar Alam, Parul Agarwal

2. **A Review of Missing Values Handling Methods on Time-Series Data**: In this paper, they will explain and describe several previous studies about missing values handling methods or approach on time series data. This paper also discuss some plausible option of methods to estimate missing values to be used by other researchers in this field of study.
   Authors: Irfan Pratama, Adhistya Erna Permanasari, Igi Ardiyanto, Rini Indrayani

Chapter 3

Proposed Algorithm

In this section, we provide short descriptions for the two used algorithms in the current work: Integration and Cointegration Test. To explain these we have to first introduce some terms related to it.

3.1 Time Series and Its Component

A time series is sequence indexed time, is denoted by x_t. Here, t is the time can be discrete or continuous.
A time series mainly have 4 components:
(i) Trend (T_t)
(ii) Seasonality (S_t)
(iii) Cyclicity (C_t)
(iv) Irregularity (ϵ_t)
Additive Model: $x_t = T_t + S_t + C_t + \epsilon_t$
Multiplicative Model: $x_t = T_t * S_t * C_t * \epsilon_t$

3.2 Stationarity

The time series $\{x_t\}$ with index t belongs to integers is said to be stationary (or, weakly stationary) if
(i) $E[x_t]^2 < \infty$
(ii) $E[x_t]$ is independent of t for all t
(iii) $cov(x_m, x_n)$ is independent of m, n
Where, all m and n belongs to set of integers, but it can be a function of (m-n).

3.3 Deterministic Trend

$$y_t = y_0 + a_0 t + \epsilon_t$$

In this above model, for the component a_0 there is a permanent effect of time $t > 0$. Since the value of a_0 cannot change with time and we can determine the value of $a_0 t$ at any fixed point of time, it is a deterministic trend. More precisely, it is a linear deterministic trend because max degree in $a_0 t$ is 1, considering $a_0 \neq 0$.

3.4 Stochastic Trend

$$y_t = y_0 + \sum_{i=1}^{t} \delta_i + \epsilon_t$$

In this above model, the coefficient of δ_i is one and it has a permanent effect of time $t > 0$. Since the value of δ_i changes for each point of time, the value of $\sum_{i=1}^{t} \delta_i$ cannot be determined, it is a stochastic trend.

3.5 Integration

If d is a non-negative integer, then $\{x_t\}$ is an integrated process, denoted as I(d), if

$$y_t = \nabla^d x_t = (1 - B)^d x_t$$

is stationary causal process. Where, B be the backshift operator.

3.6 Dickey-Fuller Test

H_0 : The time series has a unit root in AR polynomial.
H_a : The time series has no unit roots in AR polynomial.
Models Used:
(i) $x_t = x_{t-1} + \epsilon_t$ (Random Walk)
(ii) $x_t = a_0 + x_{t-1} + \epsilon_t$ (Random Walk with Drift)
(iii) $x_t = a_0 + a_1 t + x_{t-1} + \epsilon_t$ (Random Walk with Deterministic Trend)

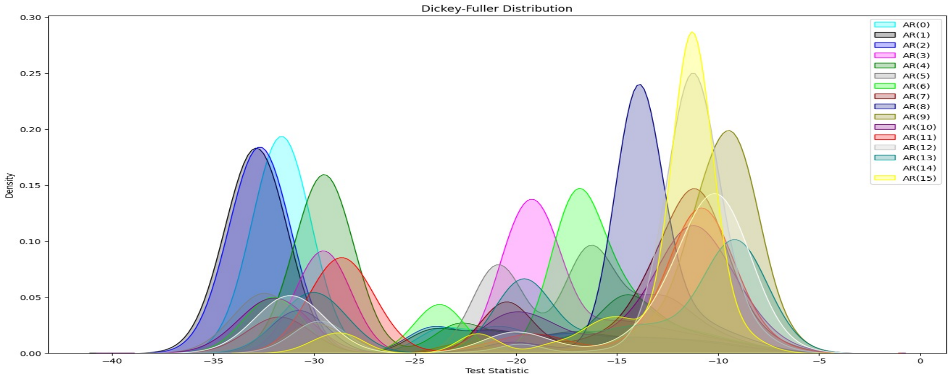

Figure 3.1: Generation of Dickey Fuller Statistics for AR(0) to AR(15)

3.7 Augmented Dickey-Fuller Test

H_0 : The time series has a unit root in AR polynomial.
H_a : The time series has no unit roots in AR polynomial.
Models Used:
(i) $\Theta(B)x_t = \epsilon_t$ (Random Walk)
(ii) $\Theta(B)x_t = a_0 + \epsilon_t$ (Random Walk with Drift)
(iii) $\Theta(B)x_t = a_0 + a_1 t + \epsilon_t$ (Random Walk with Deterministic Trend)
Where, $\Theta(B)$ be the polynomial of B of order p which has at least one unit root.

The main difference between DF and ADF test is in DF AR(1) model is used and in ADF AR(p) model is used. i.e. ADF is a generallization of DF test.

3.8 Cointegration

According to Engle and Granger (1987), the components of a vector $x_t = (x_{1t}, x_{2t}, ..., x_{nt})'$ are said to be cointegrated of order d, b, denoted as $x_t \sim CI(d, b)$ if All components of x_t are integrated of order d. $\exists \beta = (\beta_1, \beta_2, ..., \beta_n)'$ such that the linear combination $\beta' x_t = \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_n x_{nt}$ is integrated of order (d-b) where $b > 0$.
And $\beta = (\beta_1, \beta_2, ..., \beta_n)'$ is called cointegrating vector of $x_t = (x_{1t}, x_{2t}, ..., x_{nt})'$.

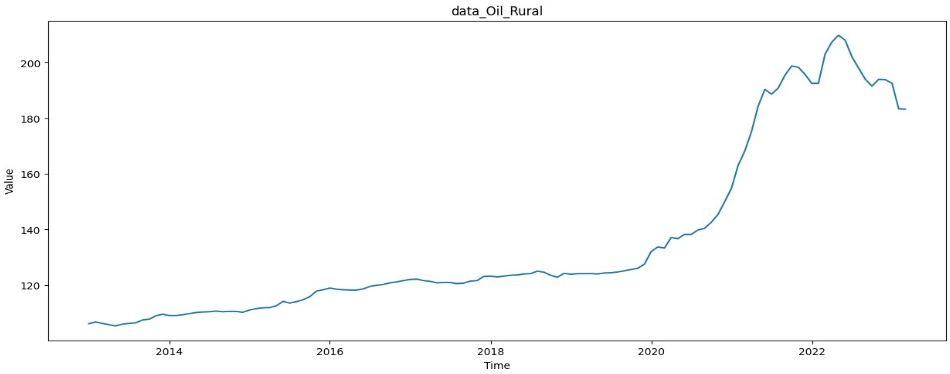

Figure 3.2: Time Series Plot of Oils and Fats CPI in Rural Sector

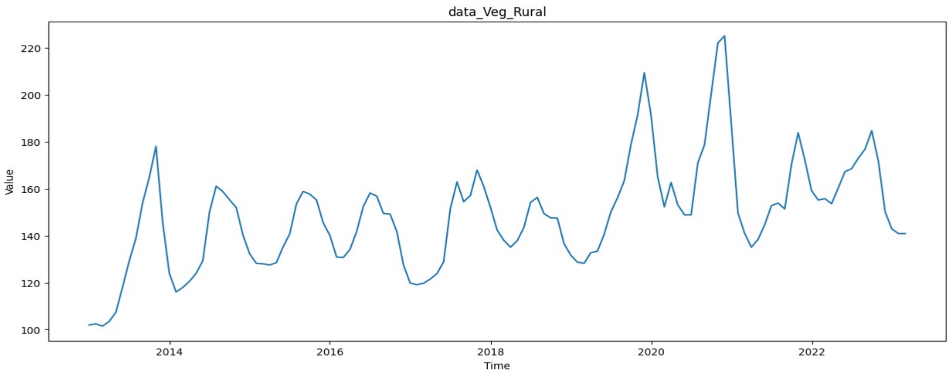

Figure 3.3: Time Series Plot of Vegetables CPI in Rural Sector

3.9 Time Series Plots

Clearly, there is trend in Oils and Fats CPI series in rural sector(Figure 3.2). Doing ADF test on this series we get that this series is integrated trend stationary.
Again, there are periodic ups and downs for vegetable CPI in every sector i.e. rural, urban, rural+urban(Figure 3.3, 3.4, 3.5).
Now, we have to check the periodicity of series to find any important seasonal component.

3.10 Autocorrelation Function(ACF) Plot

From the autocorrelation plot we can check the seasonality and the length of its period(if any).
Now, notice that for Fruit CPI in rural sector the ACF plot has a long tail type shape

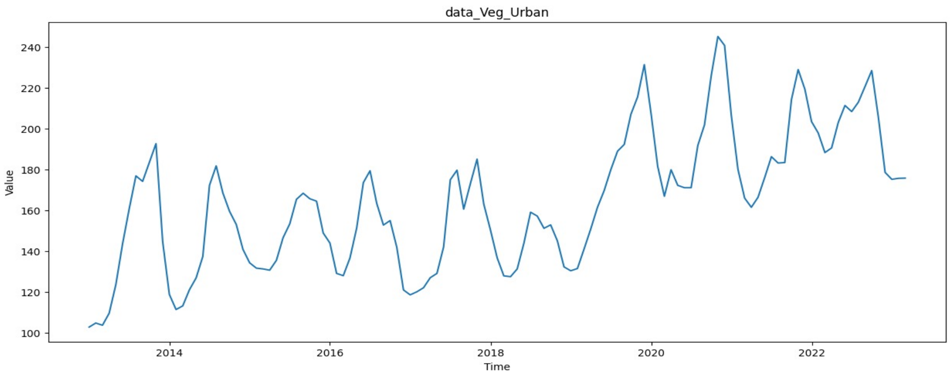

Figure 3.4: Time Series Plot of Vegetables CPI in Urban Sector

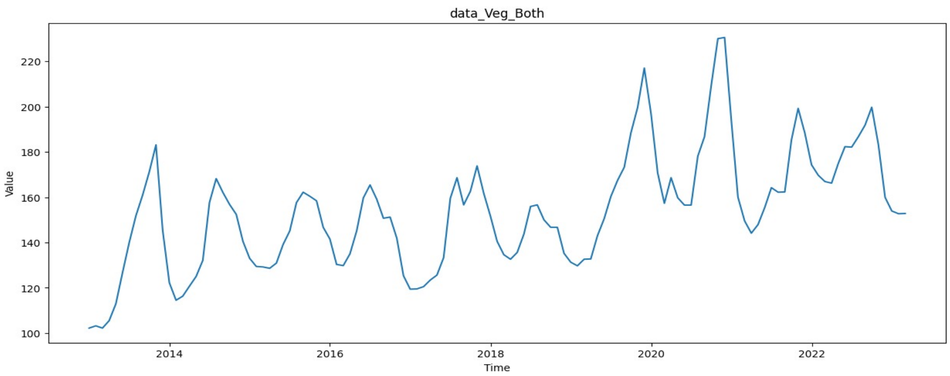

Figure 3.5: Time Series Plot of Vegetables CPI in Rural+Urban Sector

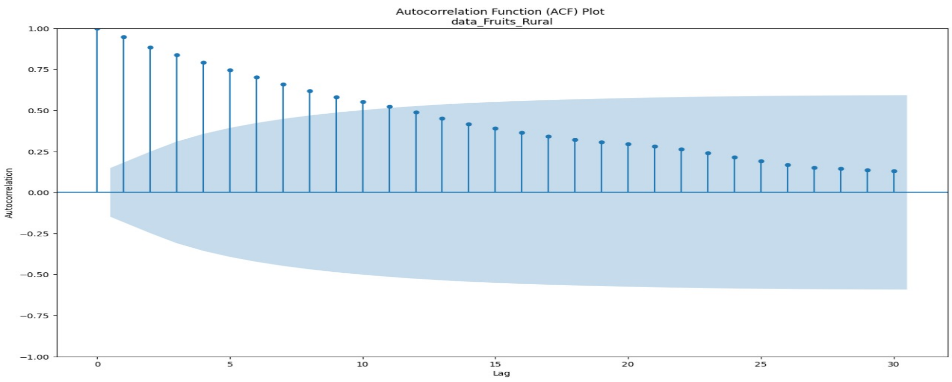

Figure 3.6: ACF Plot of Fruits CPI in Rural Sector

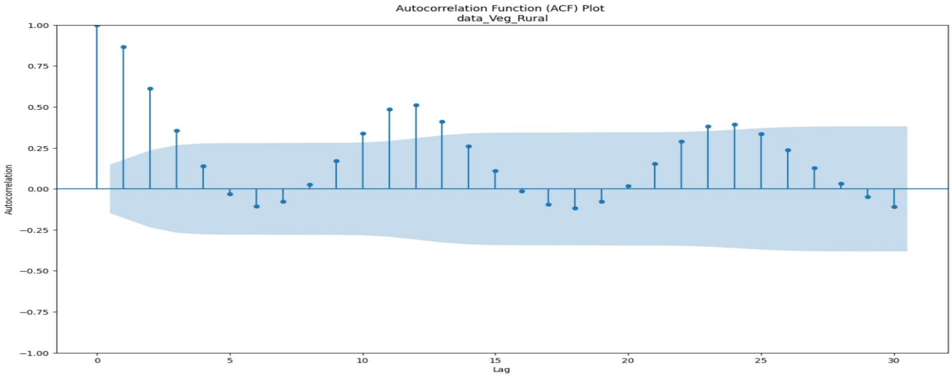

Figure 3.7: ACF Plot of Vegetables CPI in Rural Sector

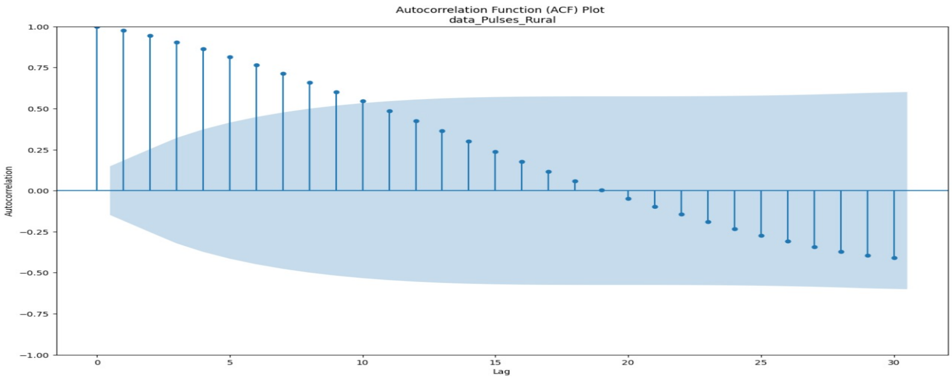

Figure 3.8: ACF Plot of Pulses and Products CPI in Rural Sector

that implies it has no periodicity or seasonality(Figure 3.6).
Also, for Vegetables CPI in rural sector the ACF has a periodic shape(Figure 3.7). From there if we calculate the difference between the two peaks, we get the period as 12. Since, it is a monthly data it is natural to get a period of 12 months which is the seasonality of that series.
Again, for Pulses and Products CPI in rural sector we are getting the periodic structure of ACF plot(Figure 3.8). But if we calculate the period it shows a period of more than 12 months. Since, the limit of period for seasonality is at most 12 months, it is not a part of seasonality.

Chapter 4

Results

From the statistical testing of these time series we get the following results. Here, each entry is a time series of a particular CPI in a particular sector.

4.1 Periodicity

4.1.1 Rural

1. Vegetables (Period 12)
2. Pulses and Products

4.1.2 Urban

1. Vegetables (Period 12)
2. Pulses and Products

4.1.3 Rural+Urban

1. Vegetables (Period 12)
2. Pulses and Products

4.2 Unit Roots

4.2.1 Rural

1. Oils and Fats
2. Footwear

3. Clothing and Footwear

4.2.2 Urban

1. Education
2. Oils and Fats
3. Fruits

4.2.3 Rural+Urban

1. Oils and Fats
2. Fruits
3. Housing
4. Health
5. Transport and Communications
6. Recreation and Amusement
7. Education

Chapter 5

Conclusion & Future Work

Since, the work is not completed yet we have to do cointegration tests on that series which are not stationary or integrated. So, there is no particular conclusions but some particular results.
When we get those series which are co-integrated then from there we get some relationship between them.
After getting the co-integrated series, if any series left then we have to check some other methods to check if they are related or not(Need more theoretical study).
Note that, the plots of other series are in the code.

Bibliography

5.1 Papers

1. A Review of Missing Values Handling Methods on Time-Series Data; Irfan Pratama, Adhistya Erna Permanasari, Igi Ardiyanto, Rini Indrayani
2. A Prediction Approach for Stock Market Volatility Based on Time Series Data; Sheikh Mohammad Idrees, M. Afshar Alam, Parul Agarwal

5.2 Books

1. Enders, Walter, 1948, Applied econometric time series Fourth edition
2. Tsay, Ruey S., 1951, Multivariate time series analysis: with R and financial applications
3. Brockwell Peter, Davis Richard, 2016, Introduction to Time Series and Forecasting Third Edition
4. Stock James, Watson Mark, 2015, Introduction to Econometrics Updated third edition
5. John D. Levendis, Time Series Econometrics

5.3 Data

1. All India Consumer Price Index Rural Urban from Jan 2013 upto April 2023