MATH 4425 Course Project Report

Xiyu REN Yang HE

May 24, 2024

1 Introduction

In this report, we analyze the inflation and unemployment rates in Hong Kong from January 1, 2003, to December 31, 2023. These two time series were selected due to their critical importance in understanding the economic health of a region. The inflation rate reflects changes in the cost of living and purchasing power, while the unemployment rate indicates the proportion of the labor force that is without work but seeking employment. Together, these indicators provide a comprehensive view of economic stability and growth.

The choice of these specific time series stems from several factors:

- **Economic Significance:** Inflation and unemployment are key macroeconomic indicators that policymakers and economists closely monitor. They influence decisions on monetary policy, fiscal policy, and labor market interventions.
- **Policy Implications:** Analyzing the relationship between inflation and unemployment can reveal insights into the economic conditions that influence both metrics. This understanding can help in formulating policies aimed at achieving sustainable economic growth and stability.
- **Data Availability:** Reliable and up-to-date data for both inflation and unemployment rates are readily available from reputable sources such as the Hong Kong Census and Statistics Department and the Hong Kong Labour Department. This availability ensures that the analysis is based on accurate and comprehensive data.
- **Relevance to Current Economic Conditions:** The period from 2023 to 2024 encompasses significant global economic events, including recovery from the COVID-19 pandemic and its impact on global supply chains, inflationary pressures, and labor markets. Analyzing this period provides insights into how these events have influenced Hong Kong's economy.

By examining these time series, this report aims to provide a detailed understanding of the trends and patterns in Hong Kong's inflation and unemployment rates, their correlation, and the potential implications for economic policy and planning.

2 Data Description

The data for this analysis was obtained from Trading Economics. The dataset includes the monthly inflation and unemployment rates for Hong Kong, covering the period from January 1, 2023, to December 31, 2024.

2.1 Inflation Rate Data

The inflation rate data represents the percentage change in the cost of living over time, measured on a monthly basis. This indicator is essential for understanding the purchasing power of consumers and the overall price stability in the economy. The dataset includes the following variables:

- **Country:** The country for which the data is recorded, in this case, Hong Kong.
- **Category:** The type of economic indicator, here it is the Inflation Rate.

- **DateTime:** The date of the observation, formatted as YYYY-MM-DD.
- **Close:** The inflation rate percentage for the given date.
- **Frequency:** The frequency of the data, which is Monthly.
- **HistoricalDataSymbol:** The symbol used for the historical data, specific to the dataset.
- **LastUpdate:** The date and time when the data was last updated.

2.2 Unemployment Rate Data

The unemployment rate data measures the percentage of the labor force that is unemployed and actively seeking employment. This indicator is a key measure of labor market health and economic stability. The dataset includes the following variables:

- **Country:** The country for which the data is recorded, in this case, Hong Kong.
- **Category:** The type of economic indicator, here it is the Unemployment Rate.
- **DateTime:** The date of the observation, formatted as YYYY-MM-DD.
- **Close:** The unemployment rate percentage for the given date.
- **Frequency:** The frequency of the data, which is Monthly.
- **HistoricalDataSymbol:** The symbol used for the historical data, specific to the dataset.
- **LastUpdate:** The date and time when the data was last updated.

2.3 Data Source and Reliability

The data was retrieved from Trading Economics, a reputable provider of economic statistics and indicators. Trading Economics compiles its data from official sources such as national statistical offices and central banks, ensuring a high level of accuracy and reliability. The availability of monthly data allows for detailed time series analysis and the identification of trends, seasonal patterns, and potential correlations between inflation and unemployment rates.

By utilizing this comprehensive and reliable dataset, we aim to conduct an in-depth analysis of the economic conditions in Hong Kong, focusing on the interplay between inflation and unemployment rates over the specified period.

3 Model Building

In this section, we outline the methodology used to build the time series models for analyzing the inflation and unemployment rates in Hong Kong. The primary objective is to identify patterns, trends, and potential correlations between these two economic indicators over the specified period.

3.1 Inflation Rate Modeling

We first focus on modeling the inflation rate data. The following steps outline the process:

3.1.1 Exploratory Data Analysis (EDA)

- **Visualization:** Plot the time series data in Figure 1 to examine trends, seasonal patterns, and anomalies. The observed data exhibits a clear stationary behavior, oscillating around its mean. Notably, an approximately periodic trend can be observed, particularly during the years 2008 to 2018. Furthermore, a significant anomaly is evident in the year 2020, attributed to the impact of the Covid-19 pandemic.
- **Summary Statistics:** Calculate key statistics such as mean, median, variance, and standard deviation, shown in Table 1.

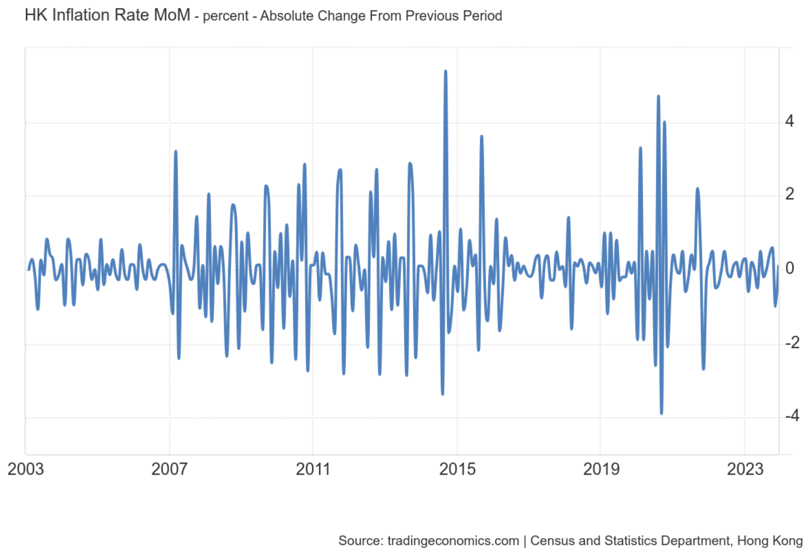


Figure 1: Hong Kong Monthly Inflation Rate Data from 01/01/2003 to 31/12/2023

Statistic	Value
Count	252
Mean	2.161905
Standard Deviation	2.009454
Minimum	-4.000000
25% Quantile	1.300000
Median (50% Quantile)	2.150000
75% Quantile	3.300000
Maximum	7.900000

Table 1: Statistical Descriptions of Hong Kong Monthly Inflation Rate Data

3.1.2 Plotting ACF and PACF

We plot the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the inflation data in Figure 2a and Figure 2b respectively.

From the ACF and PACF plots (Figure 2), we observe that the data appears to be stationary. The first two PACFs are significantly different from zero, and all ACFs gradually decay to zero, indicating no moving average (MA) terms. Therefore, we propose an ARIMA(2,0,0) model.

3.1.3 ARIMA(2,0,0) Model

Given the observations from the ACF and PACF plots, we fit an ARIMA(2,0,0) model to the inflation data. However, the 12th PACF is slightly outside the confidence level, suggesting potential seasonality.

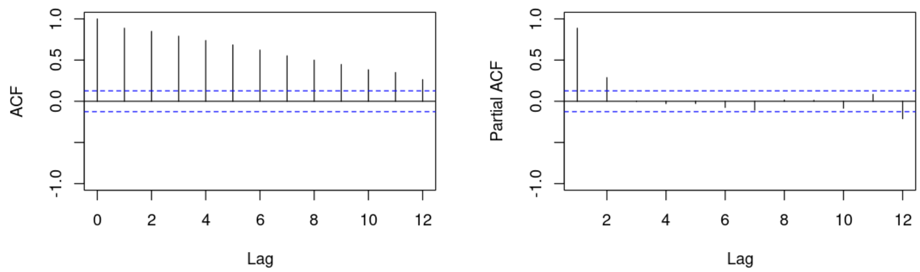
3.1.4 Seasonal Model

To account for possible seasonality, we consider the seasonal ARIMA model with a seasonal period of 12 (SARIMA). Several possible models are tested wtih teh results are shown below.

3.1.5 Diagnostic Testing and Model Selection

We perform the Ljung-Box test to evaluate the adequacy of the ARIMA(2,0,0) model. The p-value of the test is used to assess the model's goodness of fit.

The model with a higher p-value of the Ljung-Box test is favored due to its superior fit to the data. The models' performance is assessed using a diverse range of metrics, including the Akaike Information



(a) ACF of Inflation Rate Data (b) PACF of Inflation Rate Data

Figure 2: ACF and PACF of Inflation Rate Data

Criterion (AIC) and the Bayesian Information Criterion (BIC). Detailed results are presented in Table 2. We also conducted the tests on other trial of models, which are also presented in the table.

Models	p-value	AIC	BIC
ARIMA(2,0,0)	0.0000	646.1471	660.2649
$\text{SARIMA}(2,0,0)(1,0,0)_{12}$	0.4337	589.5148	607.1619
$\text{SARIMA}(3,0,3)(0,0,1)_{12}$	0.5838	564.1451	595.9100
$\text{SARIMA}(2,0,0)(1,0,1)_{12}$	0.8740	574.1896	595.3662
$\text{SARIMA}(2,0,0)(0,0,1)_{12}$	0.8968	572.8888	590.5359

Table 2: Results of p-value of Ljung-Box test, AIC and BIC for Inflation Rate Models

From the above results, we can see that the p-value of ARIMA(2,0,0) is smaller than the confidence level 0.05. Therefore we reject the null hypothesis and conclude that ARIMA(2,0,0) does not fit our model. Further, by comparing the p-value, AIC and BIC, we conclude that $\text{SARIMA}(2,0,0)(0,0,1)_{12}$ performs better than other candidate models.

$$\phi_p(B)(1-B^{12})y_t = \Theta_Q(B^{12})a_t$$

Where:

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2$ is the autoregressive polynomial. B is the backshift operator. $1 - B^{12}$ represents seasonal differencing with a lag of 12. y_t is the inflation rate time series. $\Theta(B^{12}) = 1 + \Theta_1 B^{12}$ is the moving average polynomial. a_t is the error term.

By building and analyzing these models, we aim to uncover significant insights into the dynamics of the inflation rate in Hong Kong and provide a robust basis for forecasting future trends.

3.2 Unemployment Rate Modeling

We focus on modeling the unemployment rate data using a similar approach as for the inflation rate with the additional usage of Vector Autoregression model.

3.2.1 Exploratory Data Analysis (EDA)

Perform exploratory data analysis on the unemployment rate data, including visualization and summary statistics.

- **Visualization:** Plot the time series data in Figure 3 to analyze trends, seasonal patterns, and anomalies. The unemployment rate is predominantly stable, consistently hovering around the 4% level. However, two notable anomalies are observed: a sharp increase in 2008 due to the global economic crisis and a prolonged period of elevated unemployment from 2020 to 2022 attributed to the impact of the Covid-19 pandemic. We will delve deeper into examining the stationarity of the unemployment rate later.

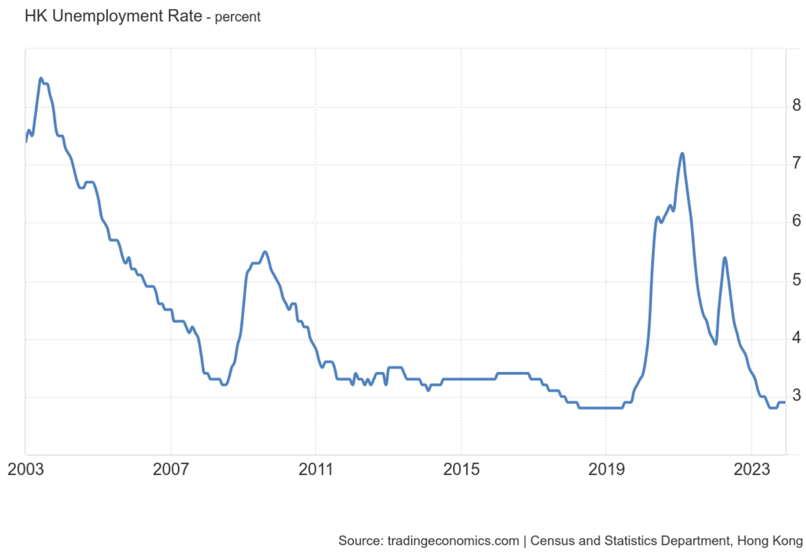


Figure 3: Hong Kong Unemployment Rate Data from 01/01/2003 to 31/12/2023

- **Summary Statistics:** Calculate key statistics such as mean, median, variance, and standard deviation shown in Table 3.

Statistic	**Value**
Count	252
Mean	4.272222
Standard Deviation	1.426340
Minimum	2.800000
25% Quantile	3.300000
Median (50% Quantile)	3.500000
75% Quantile	5.100000
Maximum	8.500000

Table 3: Statistical Descriptions of Close Data

- **Correlation Calculation:** Calculate the correlation coefficient between the unemployment rate and the inflation rate:

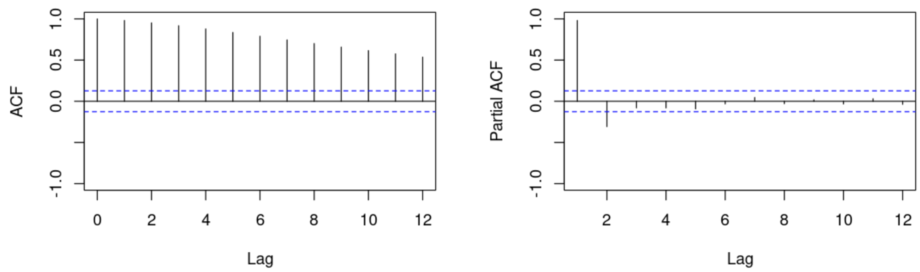
$$\rho_{\text{unemployment, inflation}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = -0.760844$$

where:

 - $\rho_{\text{unemployment, inflation}}$ is the Pearson correlation coefficient between unemployment rate (x) and inflation rate (y).
 - n is the number of observations.
 - x_i, y_i are individual observations of unemployment rate and inflation rate, respectively.
 - $\bar{x}$, $\bar{y}$ are the means of the unemployment rate and inflation rate, respectively.

3.2.2 Plotting ACF and PACF

Plot the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the unemployment rate data in Figure 4a and Figure 4b respectively.



(a) ACF of Unemployment Rate Data (b) PACF of Unemployment Rate Data

Figure 4: ACF and PACF of Inflation Rate Data

3.2.3 Stationary of Unemployment Rate

Test Statistic	−3.5055
p-value	0.0078
Number of Lags Used	12
Number of Observations Used	252
1% Critical Value	−3.4571
5% Critical Value	−2.8733
10% Critical Value	−2.5730
AIC of the Model	−272.2727

Table 4: Results of Augmented Dickey-Fuller Test for Stationarity

We perform an Augmented Dickey-Fuller (ADF) test to check the stationarity of the data. The null hypothesis of the ADF test is that the time series has a unit root, indicating that it is non-stationary. The ADF test statistic is a negative number. The more negative it is, the stronger the evidence against the null hypothesis, and the more likely the series is stationary. The p-value associated with the ADF statistic is used to determine the significance of the result. If the p-value is less than a chosen significance level (e.g., 0.05), then the null hypothesis is rejected, and the series is considered stationary. If the p-value is greater than the significance level, then the null hypothesis cannot be rejected, and the series is considered non-stationary. The ADF test compares the ADF statistic to critical values at certain confidence levels (e.g., 1%, 5%, 10%). If the ADF statistic is more negative than the critical value at a given confidence level, then the null hypothesis is rejected. If the p-value is less than the chosen significance level, you can reject the null hypothesis and conclude that the series is stationary. If the p-value is greater than the significance level, you fail to reject the null hypothesis, indicating that the series is non-stationary.

The results are presented in Table 4. From the table, we can see that the test statistic is less than The critical value at the 5% significance level, -2.8733. Therefore we reject the null hypothesis of a unit root at the 5% significance level. Moreover, A p-value less than 0.05 indicates that we can reject the null hypothesis at the 5% significance level.

Based on the test statistic and the p-value, we can conclude that the unemployment rate data is stationary. The null hypothesis of the presence of a unit root is rejected, suggesting that the time series does not have a unit root and is instead stationary. This means the properties of the series (mean, variance) do not change over time, making it suitable for time series modeling without further differencing.

3.2.4 ARIMA Model and Seasonal Model

Given the observations from the ACF and PACF plots, we fit an ARIMA(2,0,0) model to the unemployment rate data. We also consider seasonal modeling to check whether there is potential seasonality in unemployment rate.

3.2.5 Diagnostic Testing and Model Selection

Similar to inflation rate data, we perform the Ljung-Box test to evaluate the adequacy of the ARIMA(2,0,0) model. The p-value of the test is used to assess the model's goodness of fit. The results are shown in Table 5

Models	p-value	AIC	BIC
ARIMA(2,0,0)	0.8848	-271.5516	-247.1054
ARIMA(1,0,1)	0.0230	-265.6709	-258.6200
ARIMA(2,0,3)	0.5791	-266.0491	-241.3431
ARIMA(1,0,2)	0.4768	-265.3828	-247.7357
SARIMA$(2,0,0)(0,0,1)_{12}$	0.5311	-259.4450	-241.7979

Table 5: Results of p-value of Ljung-Box test, AIC and BIC for Unemployment Rate Models

Based on the findings, it can be inferred that the ARIMA(2,0,0) model provides a better fit to the data compared to other candidates. Conversely, the ARIMA(1,0,1) model is deemed inadequate in capturing the underlying data relationship, and as such, it is rejected at a 0.05 confidence level. Furthermore, seasonal patterns are not prominently observed in the unemployment rate data with a seasonality of 12.

3.3 Combining Inflation and Unemployment Rate Together

3.3.1 VAR Modeling

Building upon the analysis of the ACF and PACF plots (Figure 4), which suggested stationarity of the data and the absence of moving average (MA) terms, we further explore the interdependencies between unemployment and inflation. While the ARIMA(2,0,0) model provides insights into the dynamics of unemployment alone, we recognize the need to consider the relationship between unemployment and inflation as a holistic system. To capture this comprehensive perspective, we employ a Vector Autoregression (VAR) model. By incorporating both unemployment and inflation variables, the VAR model allows us to study the dynamic correlations and interdependence between these economic indicators over time.

3.3.2 Multivariate Portmanteau Test

We perform the multivariate portmanteau test on the vector series containing inflation rate and unemployment. Figure 5 shows that the p-values of Ljung-Box statistics are close to 0 for lag from 1 to 12. We thus conclude that there exist no auto- and cross-correlations in the vector series.

3.3.3 Selecting the Lag Order

We started by selecting potential values for the lag order (p). The lag order determines how many past values of the variables are included in the model. We considered various values for p to ensure we captured the appropriate temporal dynamics. As shown in Table 6, AIC, BIC and HQ all suggest that the lag order should be 2.

3.3.4 Fitting the VAR Model

After exploring different lag orders, we decided to fit the VAR model with $p = 2$. This means that the current values of unemployment and inflation are modeled based on their values two periods ago.

$$\text{VAR(2) model: } y_t = \vec{c} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + a_t$$

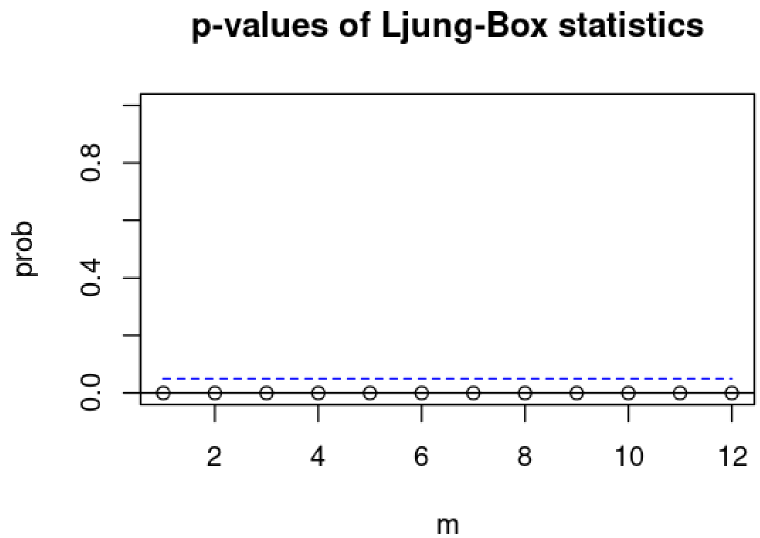


Figure 5: Multivariate Portmanteau Test for the Vector Series

p	AIC	BIC	HQ	p-vlaue
1	-3.6757	-3.6176	-3.6523	0.0000
2	-4.2401	-4.1241	-4.1933	0.0000
3	-4.2187	-4.0446	-4.1486	0.6234
4	-4.2104	-3.9783	-4.1169	0.2450
5	-4.2079	-3.9179	-4.0910	0.1552

Table 6: Results of AIC, BIC, HQ and p-value for different p

y_t is the unemployment rate time series and a_t represents the error term at time t.

$$y_t = \begin{bmatrix} 1.0256 \\ 0.1260 \end{bmatrix} + \begin{bmatrix} 0.5411 & -1.0800 \\ -0.0093 & 1.5900 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0.2907 & 0.9220 \\ 0.0005 & -0.6120 \end{bmatrix} y_{t-2} + a_t$$

Residual Covariance matrix:

$$\hat{\Sigma}_a = \begin{bmatrix} 0.6766 & -0.0044 \\ -0.0044 & 0.0191 \end{bmatrix}$$

[2,]

3.3.5 Residual Analysis

To assess the goodness of fit, we performed residual analysis on the VAR model. Residuals are the differences between the observed values and the values predicted by the model. We tested the residuals for stationarity and other statistical properties. Figure 6 shows that the test result is significant only after lag 4. We thus conclude that the model is adequate.

3.3.6 Model Refinement

To simplify the model and remove any unnecessary complexity, we applied the `refVAR` function with a threshold of 1.96. This function refines the VAR model by removing variables or lagged terms that do not contribute significantly to the model's explanatory power.

$$\text{VAR(2) model: } y_t = \vec{c} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + a_t$$

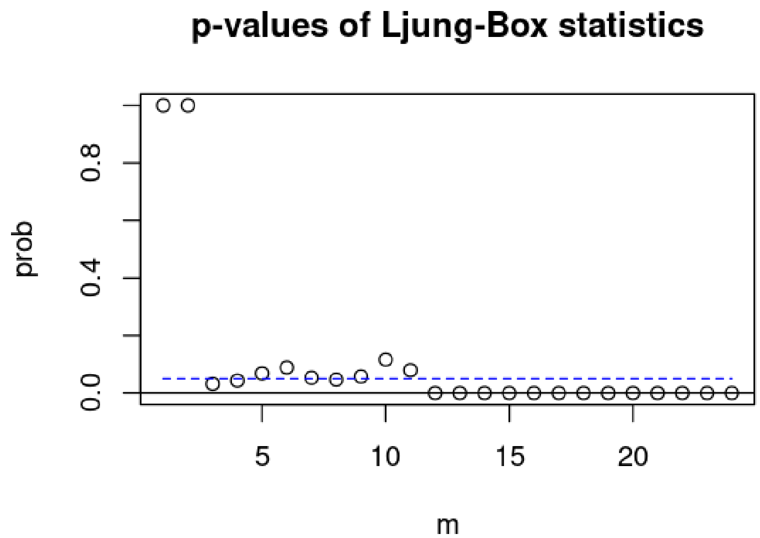


Figure 6: Multivariate Portmanteau Test for Residuals

$$y_t = \begin{bmatrix} 1.0256 \\ 0.0665 \end{bmatrix} + \begin{bmatrix} 0.5410 & -1.0800 \\ - & 1.5900 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0.2910 & 0.9220 \\ - & -0.6110 \end{bmatrix} y_{t-2} + a_t$$

$$\hat{\Sigma}_a = \begin{bmatrix} 0.6766 & -0.0044 \\ -0.0044 & 0.0193 \end{bmatrix}$$

This approach allows us to examine the relationship between unemployment and inflation, providing valuable insights into their interdependence in real-world economics. In the refined model, the two insignificant parameters show that the relationships between unemployment rate this month and inflation rate last month and the month before are not significant, while other relationships hold the same.

4 Conclusions

1. The analysis of the inflation rate reveals the presence of seasonal variations, particularly observed from 2008 to 2018, as depicted in Figure 1. Our modeling approach incorporating seasonality, the SARIMA model, outperformed the non-seasonal ARIMA model based on several evaluation metrics, including p-values, AIC, and BIC.

   The existence of seasonal variation in the inflation rate can be attributed to various factors, primarily stemming from fluctuations in prices of consumer goods and services. These fluctuations arise due to changes in weather conditions and the influence of holidays [1].

   For instance, during periods such as the Lunar New Year or other festive seasons, specific food items often experience price increases due to heightened demand or limited supply. Similarly, the prices of fresh vegetables tend to rise notably during rainy or typhoon seasons due to reduced supply.

   These price variations have a direct impact on the Consumer Price Index (CPI), which tracks the average price changes of a basket of goods and services over time. The original CPI, without adjustments for seasonal effects, reflects the actual prices paid by consumers.

2. Our analysis using the VAR model indicates an inverse relationship between the unemployment rate and the inflation rate. Consistent with the Phillips Curve theory, when the unemployment rate is low, upward pressure on wages and prices arises due to increased competition for workers, resulting in higher inflation. Conversely, when the unemployment rate is high, reduced pressure

on wages and prices leads to lower inflation. This trade-off suggests that policymakers face a dilemma of targeting either low unemployment or low inflation, as achieving both simultaneously may be challenging.

In conclusion, our findings highlight the existence of seasonal variations in the inflation rate and support the inverse relationship between the unemployment rate and inflation rate, in alignment with the Phillips Curve theory [2]. These insights contribute to a better understanding of the dynamics and interdependencies within the economy, providing valuable guidance for policymakers in their decision-making processes.

References

[1] Census and t. G. o. H. Statistics Department. Seasonally adjusted consumer price indices. 2008.

[2] A. W. Phillips. The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861–19571. *Economica*, 25:283–299, 1958.