W5453 – Advanced Time Series Analysis and Forecasting

# Project 1 : Study On The Relationship Between Japan Unemployment Rate And Real-GDP: VAR Model Application In R.


**Examiner:** Prof. Dr. Yuanhua Feng

**Assistant:** Mr. Dominik Schulz

**Submitted by :** Trusha Patel : Group number : 24

**Matriculation no:** 6872844

## Introduction

Forecasting the future helps to anticipate the changes within the marketplace. By having insight of current as well as future possibilities can help businesses to optimise their strategies and alter operations to change potential outcomes. Due to covid-19 pandamic, every country of world economy has been affected. In Japan, covid-19 situation was not as serios as in the other states like United states, India, France and many others. However, its economy has suffered significant damage, where the GDP decreased by 4.7% in 2020.  There was an increase in the unemployement rate for the first time after 11 years (Dyomina & Mazitova, 2021). It an interesting topic to research for many economic experts and viewed by them through different angles.

As our study is interested in analysis the GDP and unemployment of Japan our the period of time, we will focus on time series analysis. The time series is a collection of observations measured over time, it can be discrete or continous time units. This time series could be univerate or multivariate time series. If the series is based on single time-dependent variable then it's a univerate time series. Where as multivariate time series consists more than one-time dependent variable. The dependancy of each variable is not limited to its past values but also on other variables.  To forecast the behavior of time – dependent data, multivariate time-seires analysis is an effective statistical tool. This tool predicts future values based on the history of variations in the data (Chakraborty et al., 1992).

One of the most successful model for analysing Multivariate time series is Vector Autoregression model. This model is flexible and easy to use as well as proven to effective for describing the dynamic behavior of financial and economic time series. The Forecasts dervied from VAR models are relatively flexible as they can be restricted on the feasible future paths of certain variables in the model.

On the basis of existing literature, our paper studies the relationship between two macroeconomic variables in Japan based on VAR model. We would first provide detailed description of the VAR model and its mathematical application and later in the paper conduct statstical analysis on our example. The VAR model will be fitted to analyse the relationship between GDP and Unemployement rate of Japan.

## Detailed Description of VAR Model

The Vector autoregression (VAR) model extends the idea of univariate autoregression to $k$ time series regression, where the lagged values of all $k$ series appears as regressors.

A basic VAR model consists of a set of K endogeneous variables $y_t = (y_{1t}, \dots, y_{kt}, \dots, y_{Kt})$ for $k = 1, \dots, K$ . The VAR(p) model (Pfaff, 2008) is defined as follows :

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t ,$$

With $A_i$ are $(K\ x\ K)$ coefficient matrices for $i = 1, \dots, p$ and $u_t$ is a K-dimensional process with $E(u_t) = 0$ and time invariant positive definite covariance matrix $E(u_t u_t^T) = \Sigma_u$ (white noise).

One of the important characteristics of VAR(p) – process is its stability. This process generates stationary time series with time invariant means, variances and covariance structure, given sufficient starting values.

The stability condition for VAR(p) is as follows :

$$y_t \det\left(I_{Kp} - A_1 z - \cdots - A_p z^p\right) \neq 0 \ \ for\ |z| \leq 1.$$

If the outcome from the stability condition has a root for $z = 1$ , then either few or all variables of the process are integreated of order one i.e., $I(1)$. In practice , the empirical VAR(p)- process stability can be analyzed by considering the companion form and calculating the eigenvalues of the coefficient matrix.

A VAR(1) process can be written as :

$$\xi_t = A\xi_{t-1} + v_t,$$

With :

$$\xi_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \ A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \ v_t = \begin{bmatrix} u_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix},$$

Where the dimensions of the stacked vectors $\xi_t$ and $v_t$ is $(KP\ x\ 1)$ and the dimensions of the martix A is $(Kp\ x\ Kp)$. If the all eiegnvalues of $A_1$ are with modulus less than 1, then the VAR (p) – process is stable.

The coefficients of a VAR(p)- process for a sample of the endogenous variables $y_1, \dots y_t$ and sufficient presample values $y_{-p+1}, \dots, y_0$, can be estimated using the least squares efficiently by applied separately to each equations. OLS estimators of the VAR coefficients are consistent and jointly normal in large samples so that the usual inferential methods such as cofidence intervals and t-statistics can be used. The other possible procedure to estimate is the maximum likelihood, its theoretically better but compared to OLS, MLE is complex.

To estimate VAR models in R different R packages can be utilized. Broadly used packages are “MTS”(Tsay) & “vars”(Pfaff). The var packages from CRAN repository provides standard tools for estimation, diagnostic testing and prediction for VAR models. (Schmelzer, n.d.)

After the VAR model is estimated it is open for further analysis, where researchers are interested in diagnostic tests like testing for absence of autocorrelation, heterosceasticity or non-normality in the error process, causal inferences, forecasting or/and diagnosing model’s dynamic behavior.

(Eric, 2021) Boardly there are three types of VAR modes, the reduced form, the recursive form and structural VAR model. The reduced form VAR models each variables are considered to be a function of its own past values as well as past values of the other variables in the model. This models are simplest, but consists disadvantages such as concurrent variables are separate to from each other and the error terms will be correlated across equations. Due to which it becomes dificult to interpret the impact of individual variable on the structure.

The recursive VAR model consists of all the components of the reduced form. But it allows some variables to be functions of other concurrent variables. By imposing such a short run relationship between variables this model allows us to model structural shocks. The Structural VAR models comprises constraints that allows us to identify the causal relationships. These relationships can be further used to model and forecast impacts of individual stocks.

One of the important aspect of VAR model specification is Lag selection. In practical application, we choose a maximum number of lags, $p_{max}$, and then evalute the model performance including $p = 0,1, \dots p_{max}$. The most commonly used lag selections criteria are : Akaike information criterion (AIC), Bayesian information criterion (BIC) and Hannan- Quinn (HQ). The optimal model is the VAR(p) model which minimizes some lag selection criteria. Thus the best model is the model with lower BIC, AIC or HQ. A VAR (p) model by BIC/HQ is consistent. In case of finite sample the order by HQ can be higher, which by AIC is inconsistent and generally of higher order.

The crucial question to answer before forecasting the model is the usability of one time series to forecast another. Granger causality test is a statistical hypothesis test for determing wheather a variables is helpful for forecasting the behavior of another variable. This test only allows us to make inferences about forecasting capabilities and not stating the true causality.

In a VAR (p) model, the coefficients are denoted by $\phi_{i,jk}$ , where $i = 1, \dots, p, j, k = 1, 2$.

The two possible situations are as follows :

- If $\phi_{i,21} = 0$ for all i , but $\phi_{i,12} \neq 0$ for at least one i , then $y_{1t}$ causes $y_{2t}$ but $y_{2t}$ does not cause $y_{1t}$.
- If $\phi_{i,12} = 0$ for all i , but $\phi_{i,21} \neq 0$ for at least one i , then $y_{2t}$ causes $y_{2t}$ but $y_{1t}$ does not cause $y_{2t}$.

The basic idea is that if the prediction of one variable improved by incorporating the the second variable, then we can say that there is a causal influence on the first variable. The null hypothesis is that no explanatory power is added by jointly considering the lagged values as predictors. Thus suppose if $\phi_1$is statistically significant and $\phi_2$ is not, it can be said that changes in variable y causes changes in variable x or vice versa. But if both of them are statistically significant, there is a bivariate causal relationship among variables, if they are statistically insignificant, neither the changes in variable y nor changes in x have any effect on the other variables. (Hossain et al., 2015)

If the fitted model is adeqate, then it can be used for forecasting. For a VAR(p) model, the 1-step ahead forecast at the time origin $h$ is as follows (Hossain et al., 2015):

$$Y_h(1) = \varphi_0 + \sum_{i=1}^{p} \varphi_0 Y_{h+1-i}$$

The associated forecast errors is $e_h = a_{h+1}$ . The covariance matrix of the forecast error is $\sum$ . If the $Y_t$ is weakly stationary, then the forecast $Y_h(l)$ converges to its mean vector μ as the forecast horizon increases.

The next part of the paper will be regarding implementation of VAR model and forecasting on our example using the R programming.

## Implementation of VAR model on Japan Unemployement and Real-GDP dataset using R.

This study focus on the relaionship between Japan´s unemployment rate and Real-GDP. The data is selected during the period 1994 – 2022 as sample from Federal Reserve Economic Data (FRED). The unemployment rate is established on monthly frequency and stated in percentage (Organization for Economic Co-operation and Development, 1970). Whereas, economic growth for japan is measure by using the real – GDP data collected from FRED (JP. Cabinet Office, 1994). This data is noted quarterly and units are in billions of chained 2015 Yen

Firstly, the real-GDP data is transformed into logarithm, to make the data normally distributed so the statistical analysis results happen to more valid. Figure 1. Shows the log transformed real-GDP series over the period of time. The Figure 2. Shows the montly unepmloyment rate of all people between 15-64 years of Japan.

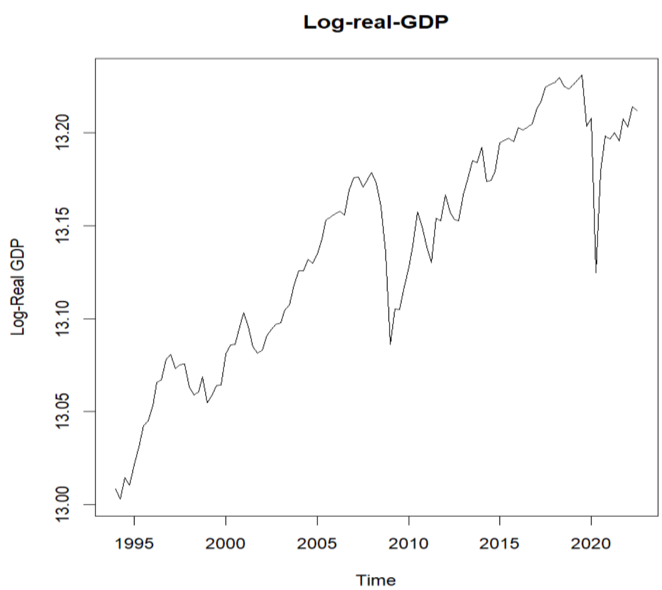


*Figure 1 : Log-real GDP*

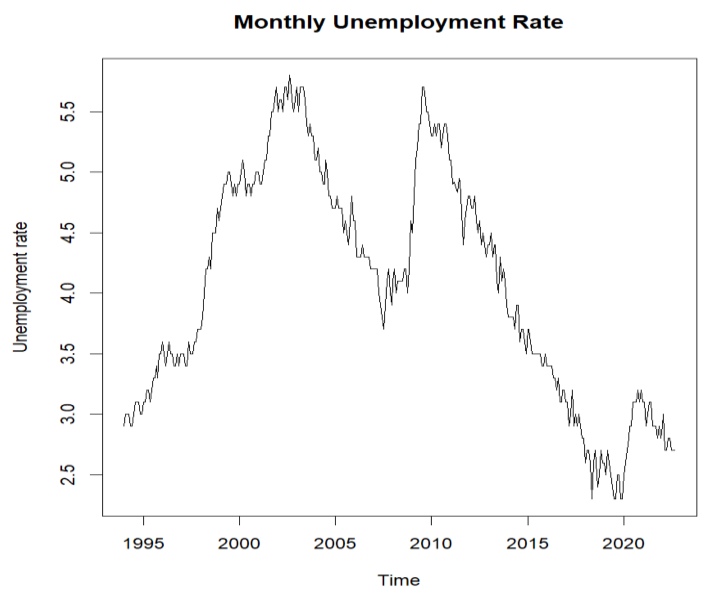


*Figure 2 : Unemployement Rate*

As we can see from the figure 1. the GDP has increased over the period of time, where as unemployment rate has shown fluctuation but ultimately decrease at the end of the period.

Before moving forward to selection and fitting the VAR model on our data, we are interested in viewing the correlation between both series. For that we run cor(cbind(GDP.D, UNEM.D)) code in R to get the cross correlation matrix.

The cross correlation matrix of two random vectors is a matrix constaining as elements the cross-correlations of all paris of elemnents of the random vectors. It is used to measure information between two time series. If the cross correlation value is closer to 1 then the set are more identical. We are interested in understanding the relationship between unemployment rate and real-GDP time series. The below shows the cross-correlation matrix of both time series.

| | Log Real-GDP | Unemployment rate |
|---|---|---|
| Log-Real-GDP | 1.0000 | -0.2267 |
| Unemployment | -0.2267 | 1.0000 |

*Table 1 : The cross correlation matrix of Unemployement and GDP.*

Thus from the table 1. We can intereprt that there is a negative correlation between both timeseries. Both of them moves in an opposite direction. If the Log-GDP increcases by some percentage then unemployement rate decreases by same percentage. Which logically makes senses as with the increase in the employements will lead to increase in the gross domestic product as the purchasing power of consumer increases.

**Vector Autoregession Model**

We are interested in fitting the VAR model to forecast our time series. For that we have installed the Multivariate Time series (MTS) package in R. The MTS package is used for analysing multivariate linear time series and estimating their volatitiy models. The package performs model specification, estimation, model checking and prediction for multivariate linear time series for various models. This package is from CRAN repository network.

Using the " m1=VAR(z,0, output=F) " We have fitted VAR model from 1 to 10. The results shows AR(p) matrix, standard error, residuals cov-matrix and selection criteria : AIC,BIC and HQ. The table 2 represents the outcomes of all the fitted VAR model selection criteria´s.

According to the table 2. findings, the BIC and HQ criteria´s highest value is in VAR(1) model and AIC highest value is VAR(10) model. Thus the best fitted model is selected based on BIC criteria is VAR (1) model. The below figure 3. represents the outcomes of all the three selection criteria´s for p = 0,1,…,10. The line in the figure 3 represents the best model selected by BIC and HQ, which is VAR(1) model.

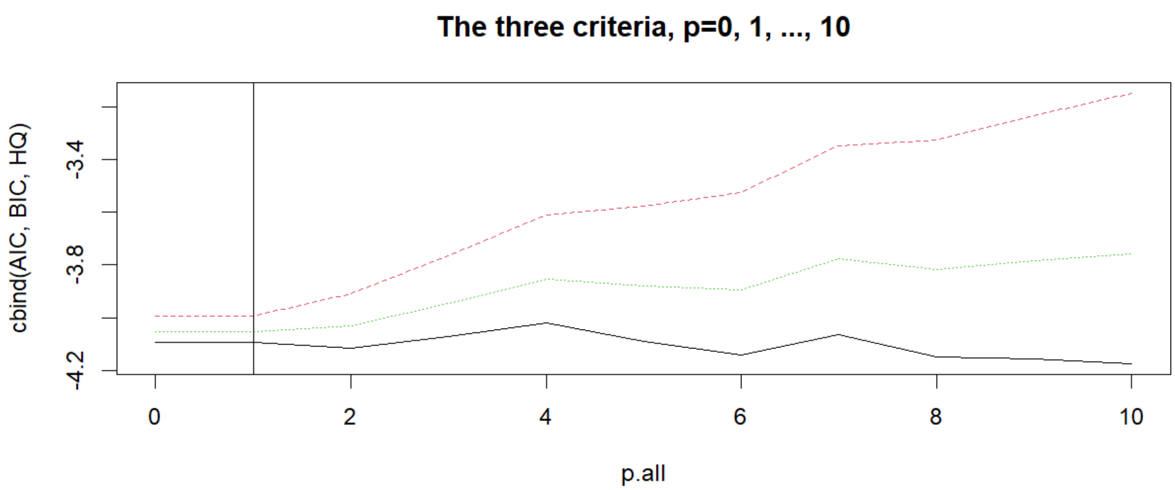


*Figure 3 : Three Selection Criteria*

| | VAR (1) | VAR (2) | VAR (3) | VAR (4) | VAR (5) |
|---|---|---|---|---|---|
| AIC | -4.0945 | -4.1146 | -4.0716 | -4.0189 | -4.0882 |
| BIC | -3.9922 | -3.9100 | -3.7647 | -3.6096 | -3.5766 |
| HQ | -4.0530 | -4.0317 | -3.9473 | -3.8531 | -3.8810 |

| | VAR (6) | VAR (7) | VAR (8) | VAR (9) | VAR (10) |
|---|---|---|---|---|---|
| AIC | -4.1412 | -4.0648 | -4.1469 | -4.1564 | -4.1723 |
| BIC | -3.5273 | -3.3485 | -3.3284 | -3.2355 | -3.1491 |
| HQ | -3.8925 | -3.7747 | -3.8154 | -3.7834 | -3.7579 |

*Table 2 : Three Selection Criteria*

Using the VAR( ) function of "vars" Package we have fitted the best VAR model. The "vars" package from CRAN repository is used for estimation, lag selection,diagnostic testing,forecasting, causality analysis, forecast error variance decomposition and impulse response functions of VAR models and estimation of SVAR and SVEC models. We have used it for estimation and forecasting of VAR(1) model. The below mention R code was used to fit the model.

model <- VAR(z, p = 1, type = "const", season = NULL, exog = NULL)

According to the summary of the model, the prediction is quite poor quality. As the value of adjusted R is quite low, we can say that fitted model not a good fit. This could be due to multiple reasons, one could be that Adjusted R-squared increases only when independent variable is significant and affects dependent variable.

**Granger- Causality Test**

As the fitted model was not a good fit. We were willing to observe the usefullness of both series to forecast another. The Granger causality test is statistical hypothesis test for determining whether one time series is useful in forecasting another. The below code is used to conduct the test in R.

causality(model, cause = NULL, vcov.=NULL, boot=FALSE, boot.runs=100)

- Granger causality H0: GDP.D do not Granger-cause UNEM.D
  p-value: 0.0008187
  As the P-value is less than 0.05 the hypothesis is statistically significant and its rejected. Thus, GDP do Granger-cause unemployment.

- H0: No instantaneous causality between: GDP.D and UNEM.D
  p-value: 0.02261
  Similar to the above interpretation, P-value is lower than 0.05, GDP and Unemployment has instantaneous causality.

**Forecasting**

To forecast our both time series we used the predict command function in R. The forecasting is based on 8 steps ahead. We are interested to forecast for the 95% confidence interval. The below code represents how we have forecasted the model in R.

predictions <- predict(model, n.ahead = 8, ci = 0.95)

Below figures represents the forecasted series. The Real-GDP seems to increase according to the forecasting and after a point becomes stable.

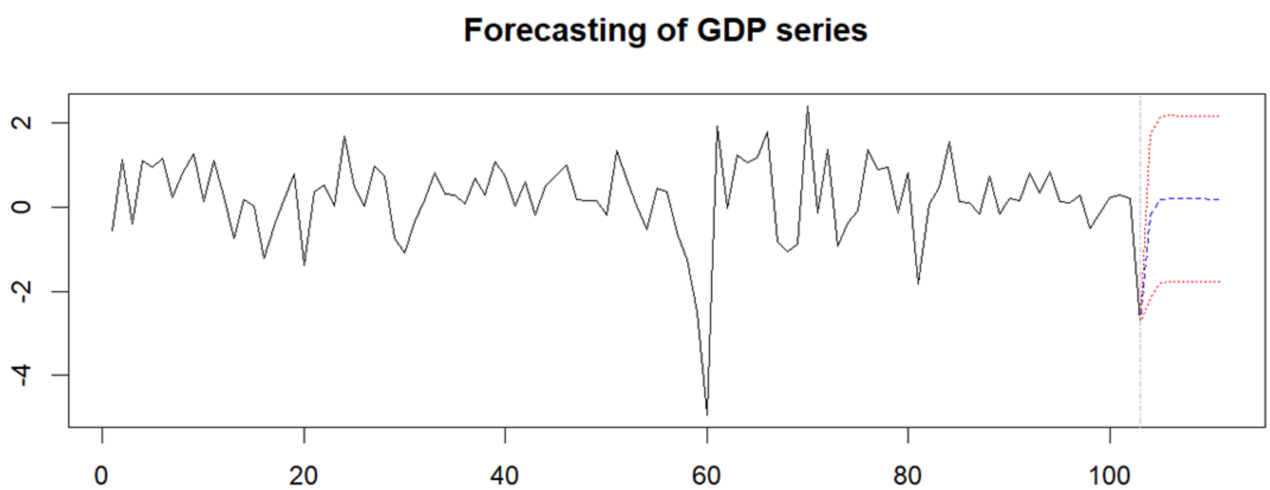


*Figure 4 : Forecasting of the Real-GDP series*

The figure 5. represents time series forecasting of unemployement rate. According to the graph, there might be increase in the unemployment rate a certain period and then it will start decreasing.

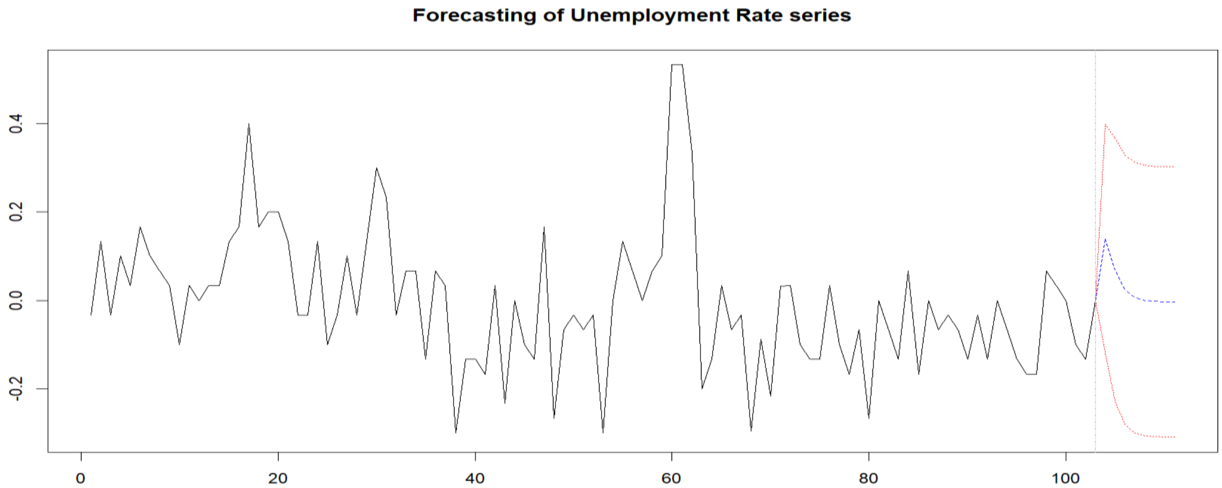


*Figure 5 : Forecasting of Unemployment rate series*

## Conclusion

The VAR model is a multivariate version of the AR model of univariate time series. In our study, we have used the VAR(1) model to for estimating and forecasting the relationship between unemployement rate and real-GDP in Japan for the period between from 1994 to 2022. It has been chosen as best model based on the selection criteria´s BIC and HQ.  We used different R packages like "vars" and "MTS". The model estimation results does not seem to be proven that statisfactory for analysis the relationship. After conducting the Granger causality tests, which involves in testing whether or not the lagged values of a given variable used in the VAR model helps in predicting other endogenous variables in the series. According to the results of this test, the null hypothesis was rejected which meant that GDP Granger causes Unemployment rate and vise-versa. Which is a real world situation as with decrease in unemployment rate there will relative increase in the GDP of the country. The forecasting of our model shows that GDP rate might increase in the coming period of time while there will be immediate increase then constant decrease.

This study could be extended by conducting the heteroscedasticty tests, impulse response analysis or unit root test on the dataset. Even there is possiblitity to include other time seires which could have impact on this series, like inflation, foreign trade or more. This study has a enormous chances of extention.

# References


- Chakraborty, K., Mehrotra, K., Mohan, C. K., & Ranka, S. (1992). Forecasting the behavior of multivariate time series using neural networks. *Neural Networks*, *5*(6), 961–970. https://doi.org/10.1016/S0893-6080(05)80092-9
- Dyomina, Y. V., & Mazitova, M. G. (2021). The COVID-19 pandemic and its impact on the Japanese economy. *Japanese Studies in Russia*, 57–75.
- Eric. (2021, April 15). Introduction to the Fundamentals of Vector Autoregressive Models. *Aptech*. https://www.aptech.com/blog/introduction-to-the-fundamentals-of-vector-autoregressive-models/
- Hossain, A., Kamruzzaman, M., & Ali, M. (2015). Vector Autoregressive (VAR) Modeling and Projection of DSE. *Chinese Business Review*, *14*, 273–289. https://doi.org/10.17265/1537-1506/2015.06.001
- JP. Cabinet Office. (1994, January 1). *Real Gross Domestic Product for Japan*. FRED, Federal Reserve Bank of St. Louis; FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/JPNRGDPEXP
- Organization for Economic Co-operation and Development. (1970, January 1). *Unemployment Rate: Aged 15-64: All Persons for Japan*. FRED, Federal Reserve Bank of St. Louis; FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/LRUN64TTJPM156S
- Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation Within *R* Package **vars**. *Journal of Statistical Software*, *27*(4). https://doi.org/10.18637/jss.v027.i04
- Schmelzer, C. H., Martin Arnold, Alexander Gerber, and Martin. (n.d.). *16.1 Vector Autoregressions | Introduction to Econometrics with R*. Retrieved 15 December 2022, from https://www.econometrics-with-r.org/