ISSN: 2456-1452
Maths 2021; 6(6): 31-37

www.mathsjournal.com
Received: 05-08-2021
Accepted: 13-10-2021

Akasam Srinivasulu
Research Scholar, Department of Statistics, S.V University, Tirupati, Andhra Pradesh, India

MV Narayana Murthy
Research Scholar, Department of Statistics, S.V University, Tirupati, Andhra Pradesh, India

M Bhupati Naidu
DDE, Professor in Department of Statistics, S.V University, Tirupati, Andhra Pradesh, India

Corresponding Author:
Akasam Srinivasulu
Research Scholar, Department of Statistics, S.V University, Tirupati, Andhra Pradesh, India

Testing granger causality and the assumptions of residuals in vector auto regressive model by using R

Akasam Srinivasulu, MV Narayana Murthy and M Bhupati Naidu

Abstract
In this paper, the objective is to study whether and how the unemployment situation affected the GDP in India. To achieve this, a vector autoregression (VAR) model is employed and data analysis is carried out. To predict the future impact of unemployment on the GDP, we used the granger casualty function. Here the researchers focus on the assumptions like Serial correlation, Heteroscedasticity and the Normality of residuals. The data of GDP and unemployment of India can be taken from the world bank by using the command WDI from 1991 to 2020.

Keywords: Unemploment, GDP, India, VAR, Granger causalty. forecasting

Introduction
The vector autoregression model is a multivariate ts model that relates current observations of a variable with past observations of itself and past observations of another variables in the system. VAR models fully differ from univariate autoregressive models because they allow feedback to occur between the variables in the model. A VAR model is made up of a system of equations that represents the relationships between multiple variables. In VAR model, first we know how many endogenous variables are included and how many autoregressive terms are included. If we have two endogenous variables and autoregressive terms, we say the model is a Bivariate VAR(2) model. VAR model is composed of n equations and includes p-lags of the variables. Lag selection is one of the important aspects of VAR model specification. In practical applications, we generally choose a maximum number of lags, pmax, and evaluate the performance of the model including p=0,1,…,pmax. The optimal model is then the model VAR(p) which minimizes some lag selection criteria. The most commonly used lag selection criteria are: Akaike (AIC), Schwarz-Bayesian (BIC), Hannan-Quinn (HQ).These methods are usually built into software and lag selection is almost completely automated .It is important to be deliberate about how many variables we include in our VAR model. Adding additional variables: Increases the number of coefficients to be estimated for each equation and each number of lags. Introduce additional estimation error.
Deciding the variables to include in a VAR model founded in theory, as much as possible. We use granger causality, to test the forecasting relevance of variables. The equation can be estimated using ols have the assumptions: The error term has a conditional mean of zero, the variables in the model are stationary, no perfect multicollinearity. Under these assumptions, the ols estimates will be consistent. This can be evaluated using t-statistics and p-values.

Review of Literature
A Graphical vector autoregressive Modelling approach displaying way to deal with the investigation of electronic dairy information by Beta Wild, Michalel Eichler. In this review the utilization of the graphical VAR approach for the investigation of electronic journal information prompts a more profound understanding into patient's dynamics and dependence structures. An expanding utilization of this demonstrating approach could prompt a superior comprehension of intricate mental and physiological instruments in various spaces of clinical consideration and exploration.

Application of Bayesian Vector Autoregressive Model in Regional Economic Forecast by Jinghao Ma, Yujie Shang, and Hongyan Zhang, in this study, the BVAR model created based on the VAR model not just only the consumption of the VAR model, yet in addition beats the issue of inordinate utilization of the VAR model's opportunity. The BVAR model is likewise more exact in foreseeing the defining moment of the provincial monetary development in 2020.

Unemployment and COVID-19 Impact in Greece: A Vector Autoregression (VAR) Data Analysis by Christos Katris, in this review, the analyst built and fitted Vector Autoregressive (VAR) models with the expect to investigate the effect of COVID-19 cases on Greece's overall joblessness and on two more delicate cases, i.e., Female and the Youth joblessness. Besides, the estimating capacity of the VAR model is viewed as restricted and other univariate approaches show up as best. Granger cause the general joblessness rates just for the EU27 nations. A shock in COVID-19 cases in Greece will have a lower sway in totally thought about sorts of joblessness. For all joblessness types the impact of COVID-19 cases is relied upon to be lower for Greece contrasted with the EU27 nations.

Data source, Variables & Methodology

The purpose of the study is if there is an effect of unemployment on GDP in India By using the VAR model. This study concentrates on Granger causality and also this model satisfying the assumptions of residuals (Heteroscedasticity, Autocorrelation, Normality) by using the Standard statistical tests.

Variables

In this study the variables are GDP and unemployment. The data (annual data) for these variables are taken from 1991 to 2020.

Methodology

The Vector Auto regression model is a statistical model which describes the evolution of multivariate linear time series with k endogenous variables. The evolution of these endogenous variables in the system is considered not only as function of their own history, but as a function of the lagged values of all endogenous variables. This model is a generalization of ARIMA models for univariate time series. This is the mostly used model for multivariate time series forecasting. A VAR(3) model is a model where each variable is linear combination of the last three periods (lags) of all the variables of the system.

The system of equations for a VAR (1) model with two time series (variables Y1 and Y2) is as follows:

$$Y_{1,t}=\alpha_1+\beta_{11,1}\, y_{1,t-1}+\beta_{12,1}\, y_{2,t-1}+\epsilon_{1,t}$$

$$Y_{2,t}=\alpha_2+\beta_{21,1} y_{1,t-1}+\beta_{22,1}\, y_{2,t-1}+\epsilon_{2,t}$$

Where, Y {1,t-1} and Y{2,t-1} are the first lag of time series Y1 and Y2 respectively.

The second order VAR (2) model for two variables would include up to two lags for each variable (Y1 and Y2).

$$Y_{1,t}=\alpha_1+\beta_{11,1}\, y_{1,t-1}+\beta_{12,1}\, y_{2,t-1}+ \beta_{11,2}\, y_{1,t-2}+\beta_{12,2}\, y_{2,t-2}+\epsilon_{1,t}$$

$$Y_{2,t}=\alpha_2+\beta_{21,1} y_{1,t-1}+\beta_{22,1}\, y_{2,t-1}+ \beta_{21,2} y_{1,t-2}+\beta_{22,2}\, y_{2,t-2}+\epsilon_{2,t}$$

VAR (2) model with three variables (Y1, Y2 and Y3) is

$$Y_{1,t}=\alpha_1+\beta_{11,1}\, y_{1,t-1}+\beta_{12,1}\, y_{2,t-1}+ \beta_{11,2}\, y_{1,t-2}+\beta_{12,2}\, y_{2,t-2}+ \beta_{13,1}\, y_{3,t-1}+\beta_{13,2}\, y_{3,t-2}+\epsilon_{1,t}$$

$$Y_{2,t}=\alpha_2+\beta_{21,1} y_{1,t-1}+\beta_{22,1}\, y_{2,t-1}+ \beta_{21,2} y_{1,t-2}+\beta_{22,2}\, y_{2,t-2}+ \beta_{23,1} y_{3,t-1}+\beta_{23,2}\, y_{3,t-2}+\epsilon_{2,t}$$

$$Y_{3,t}=\alpha_3+\beta_{31,1} y_{1,t-1}+\beta_{32,1}\, y_{2,t-1}+ \beta_{31,2} y_{1,t-2}+\beta_{32,2}\, y_{2,t-2}+ \beta_{33,1} y_{3,t-1}+\beta_{33,2}\, y_{3,t-2}+\epsilon_{3,t}$$

The VAR model is certainly useful for studying the impact of unemployment on GDP. One question is whether this approach is useful for forecasting. The answer is not straightforward, in the sense that accurate forecasting is a different task than studying the impact of a factor. Could a VAR model be used for both of these tasks effectively or not.

Testing Stationarity, Autocorrelation, Stability & Granger causality:

Stationary time series are detrended series, without periodic fluctuations. Stationarity is critical to development of a VAR model because in its absence, a model's statistics such as means and correlations will not accurately describe the time series signal. The augmented Dickey-Fuller test is used to test stationarity. The null hypothesis is that the time series is non-stationary and the alternative is that the series is stationary. Rejection of the null hypothesis indicates that the series does not need transformation to achieve stationarity.

Once a VAR model has been developed, the next step is to determine if the selected model provides an adequate description of the data. In familiar regression models, this is performed by examining the residuals, which are differences between the actual observations and model-fitted values. In time series models, autocorrelation of the residual values is used to determine the goodness of fit of the model. Autocorrelation of the residuals indicates that there is information that has not been accounted for in the model. Serial. test command is a standard tool for checking residual autocorrelation in VAR models. The null hypothesis is that there is no residual autocorrelation; the alternative is that residual autocorrelation exists.

Stability refers to checking whether the model is a good representation of how the time series evolved over the sampling window period. Technically, stability of a VAR system is evaluated using the roots of the characteristic polynomial of the coefficient matrix A.

VAR models depict the joint generation interaction of various variables over time, so they can be utilized for finding the connections between the factors. Granger causality is one kind of connection between time series. The fundamental thought of Granger causality can be expressed as though the forecast of one time series is improved by joining the information on a second time series, then, at that point, the last is said to impact the first. In particular, two autoregressive models are fitted to the initial time series with and without including the second time series and the improvement of the forecast is estimated by the proportion of the fluctuation of the error terms. The null hypothesis for GC is that no logical force is added by mutually thinking about the slacked upsides of y and x as indicators. The invalid theory that x doesn't cause y is dismissed if coefficients for the slacked upsides of x are huge; i.e., Granger called a variable x causal for a variable y if the lagged values of x are useful for further developing forecasts of (y at future occasions). The VAR structure is adaptable and gives a environment to carrying out this kind of examination.

Granger causality is used to investigate causality between two variables in a time series, it uses empirical data sets to find patterns of correlation. Causality is closely related to the idea of cause-and-effect, although it isn't exactly the same. A variable X is causal to variable Y if X is the cause of Y or Y is the cause of X. However, with Granger causality, we aren't testing a true cause-and-effect relationship.

Review of results

*GDP, unemployment data is time series data

*The optimal lag selection is selected by using the VAR select function, the opimal lag is 1

*In present study the researcher using two variables GDP, unemployment, so in this case two types of results.

In estimation results for GDP, const and unemployment at lag1 is significant but GDP at lag1 is not significant. Unemployment explains 23.79% of variability in GDP the remaining is explained by other factors.

In estimation results for unemployment, const and GDP at lag1 is least significant but GDP at lag1 is not significant. GDP explains 10.5% of variability in GDP the remaining is explained by other factors.

*The assumption of serial correlation is tested by using the function serial. Test command.

In this the null and alternative hypothesis is

H0: There is no serial correlation in residuals

H1: There is serial correlation in residuals

The value of P is greater than 0.05, here the value of p is greater than 0.05, so we fail to reject the nul hypothesis. Therefore there is no serial correlation presence in residuals.

* The assumption of Heteroscedasticity is tested by using the function arch. test command.

In this the null and alternative hypothesis is

H0: There is no heteroscedasticity present in residuals

H1: Heteroscedasticity present in residuals

The value of P is greater than 0.05, here the value of p is greater than 0.05, so we fail to reject the nul hypothesis. There fore there is no heteroscedasticity presence in residuals.

* The assumption of normality is tested by using the function Normality Test command.

*The stability of the model or if there is any structural breaks in the residuals can be tested by using the command stability

By seeing the graph of ols cusum chart of GDP & ols cusum chart of une we say that there is no points lie out side the limits .so there is no structural breaks in the residuals.

*Granger causality can be calculated for GDP and unemployment separately by using the command causality.

In Granger causality by GDP the null and alternative hypothesis is

H0: GDP do not granger causes to une

H1: GDP granger causes to une

Here the value of P is greater than 0.05, so we there is no evidence to reject the null hypothesis .

In this GDP do nor granger causes by unemployment

In Granger causality by une the null and alternative hypothesis is

H0:une do not granger causes to GDP

H1:une granger causes to GDP

Here the value of P is less than 0.05 ,so we say that Une granger causes by GDP

```
class(GDP)
## [1] "ts"
class(une)
## [1] "ts"
gu<-cbind(GDP,une)
lagselect<-VARselect(GDP)
lagselect$selection
```

```
## AIC(n) HQ(n) SC(n) FPE(n)
## 1 1 1 1
 model1<-VAR(gu,p=1,type="const",season=NULL,exog=NULL)
 summary(model1)
##
## VAR Estimation Results:
## =========================
## Endogenous variables: GDP, une
## Deterministic variables: const
## Sample size: 29
## Log Likelihood: -61.454
## Roots of the characteristic polynomial:
## 0.685 0.685
## Call:
## VAR(y = gu, p = 1, type = "const", exogen = NULL)
##
##
## Estimation results for equation GDP:
## ===================================
## GDP = GDP.l1 + une.l1 + const
##
## Estimate Std. Error t value Pr(>|t|)
## GDP.l1 0.3166 0.2801 1.130 0.2687
## une.l1 9.1355 3.7120 2.461 0.0208 *
## const -47.3762 20.7329 -2.285 0.0307 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 2.853 on 26 degrees of freedom
## Multiple R-Squared: 0.2379, Adjusted R-squared: 0.1793
## F-statistic: 4.058 on 2 and 26 DF, p-value: 0.02926
##
##
## Estimation results for equation une:
## ===================================
## une = GDP.l1 + une.l1 + const
##
## Estimate Std. Error t value Pr(>|t|)
## GDP.l1 -0.05214 0.03008 -1.733 0.0949 .
## une.l1 -0.02238 0.39859 -0.056 0.9557
## const 6.12643 2.22630 2.752 0.0107 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.3064 on 26 degrees of freedom
## Multiple R-Squared: 0.1058, Adjusted R-squared: 0.03698
## F-statistic: 1.538 on 2 and 26 DF, p-value: 0.2338
##
##
##
## Covariance matrix of residuals:
## GDP une
## GDP 8.1422 -0.68480
## une -0.6848 0.09388
##
## Correlation matrix of residuals:
## GDP une
## GDP 1.0000 -0.7832
## une -0.7832 1.0000
#Serial correlation

## Portmanteau Test (asymptotic)
##
```

```
## data: Residuals of VAR object model1
## Chi-squared = 27.869, df = 44, p-value = 0.9724
#Hetroscedasticity
##
## ARCH (multivariate)
##
## data: Residuals of VAR object model1
## Chi-squared = 51, df = 108, p-value = 1
#Normal distribution of the residuals
## $JB
##
## JB-Test (multivariate)
##
## data: Residuals of VAR object model1
## Chi-squared = 37.693, df = 4, p-value = 1.297e-07
##
##
## $Skewness
##
## Skewness only (multivariate)
##
## data: Residuals of VAR object model1
## Chi-squared = 13.371, df = 2, p-value = 0.001249
##
##
## $Kurtosis
##
## Kurtosis only (multivariate)
##
## data: Residuals of VAR object model1
## Chi-squared = 24.322, df = 2, p-value = 5.231e-06
#Testing for structurl breaks in the residuals
stability1<-stability(model1,type="OLS-CUSUM")
plot(stability1)
```

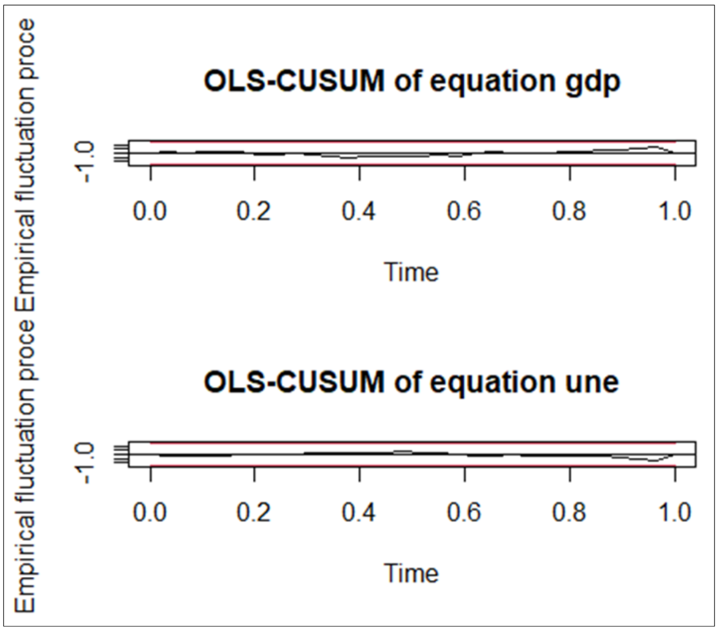

Fig 1: OLS-CUSUM

```
#Granger causality
grangerGDP<-causality(model1,cause="GDP")
grangerGDP
## $Granger
##
## Granger causality H0: GDP do not Granger-cause une
##
## data: VAR object model1
```

```
## F-Test = 3.0049, df1 = 1, df2 = 52, p-value = 0.08894
##
##
## $Instant
##
## H0: No instantaneous causality between: GDP and une
##
## data: VAR object model1
## Chi-squared = 11.026, df = 1, p-value = 0.0008983
grangerune<-causality(model1,cause="une")
grangerune
## $Granger
##
## Granger causality H0: une do not Granger-cause GDP
##
## data: VAR object model1
## F-Test = 6.057, df1 = 1, df2 = 52, p-value = 0.01721
##
##
## $Instant
##
## H0: No instantaneous causality between: une and GDP
##
## data: VAR object model1
## Chi-squared = 11.026, df = 1, p-value = 0.0008983
forecast<-predict(model1,n.ahead=4,ci=0.95)
forecast
## $GDP
## fcst lower upper CI
## [1,] 15.055627 9.462948 20.648306 5.592679
## [2,] 15.699248 8.678233 22.720263 7.021015
## [3,] 5.086460 -2.764871 12.937791 7.851331
## [4,] 1.661693 -6.299770 9.623155 7.961463
##
## $une
## fcst lower upper CI
## [1,] 6.382616 5.782073 6.983159 0.6005429
## [2,] 5.198604 4.535484 5.861724 0.6631198
## [3,] 5.191538 4.491296 5.891780 0.7002417
## [4,] 5.745053 5.022256 6.467849 0.7227966
```

Conclusion

In this paper VAR model is constructed for time series data GDP and unemployment in India. In this model there is no serial correlation presence in residuals, there is no Heteroscedasticity presence in the residuals. By using Granger causality test we say that univariate granger causality exits Thai is unemployment is granger causes to GDP in India.

References

1. Milani F. COVID-19 outbreak, social response, and early economic effects: A global VAR analysis of cross-country interdependencies. J. Popul. Econ 2021;34:223-252.
2. Su CW, Dai K, Ullah S, Andlib Z. COVID-19 pandemic and unemployment dynamics in European economies. Econ. Res. Ekon. Istraživanja 2021, 1-13.
3. Zimmermann KF, Karabulut G, Bilgin MH, Doker AC. Inter-country distancing, globalisation and the coronavirus pandemic. World Econ 2020;43:1484-1498.
4. Chen C, Liu L, Zhao N. Fear sentiment, uncertainty, and bitcoin price dynamics: The case of COVID-19. Emerg. Mark. Financ. Trade 2020;56:2298-2309.
5. Altig D, Baker S, Barrero JM, Bloom N, Bunn P, Chen S, *et al*. Economic uncertainty before and during the COVID-19 pandemic. J. Public Econ 2020;191:104274.
6. Primiceri GE, Tambalotti A. Macroeconomic Forecasting in the Time of COVID-19; Manuscript; Northwestern University: Evanston, IL, USA 2020.
7. Gharehgozli O, Nayebvali P, Gharehgozli A, Zamanian Z. Impact of COVID-19 on the Economic Output of the US Outbreak's Epicenter. Econ. Disasters Clim. Chang 2020;4:561-573.
8. Balcerzak AP, Żurek M. Foreign Direct Investment and Unemployment: VAR Analysis for Poland in the Years 1995–2009. Eur. Res. Stud 2011;14:3-14.
9. Linzert T. Sources of German Unemployment: Evidence from a Structural VAR Model. ZEW Discussion Paper No.01-41 2001.

10. Cuestas JC, Ordóñez J. Oil prices and unemployment in the UK before and after the crisis: A Bayesian VAR approach. A note. Phys. A Stat. Mech. Its Appl 2018;510:200-207.
11. Maidorn S. The effects of shocks on the Austrian unemployment rate—A structural VAR approach. Empir. Econ 2003;28:387-402.
12. Dolado JJ, Jimeno JF. The causes of Spanish unemployment: A structural VAR approach. Eur. Econ. Rev 1997;41:1281-1307.
13. Evans GW. Output and unemployment dynamics in the United States: 1950–1985. J. Appl. Econ. 1989, 4, 213–237.
14. Edlund, P.O.; Karlsson, S. Forecasting the Swedish unemployment rate VAR vs. transfer function modelling. Int. J. Forecast 1993;9:61-76.
15. Fenga L, Son-Turan S. Forecasting Youth Unemployment in the Aftermath of the COVID-19 Pandemic: The Italian Case; Research Square: Durham, NC, USA 2020.
16. Lütkepohl H. New Introduction to Multiple Time Series Analysis; Springer Science & Business Media: Berlin, Germany 2005.
17. Lütkepohl H. Impulse Response Function. In The New Palgrave Dictionary of Economics; Macmillan Publishers Ltd., Ed.; Palgrave Macmillan: London, UK 2018.
18. Hatemi-J A. Asymmetric generalized impulse responses with an application in finance. Econ. Model 2014;36:18-22.
19. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice, 2nd ed.; OTexts: Melbourne, Australia 2018.