|  |
| --- |
| **PROBLEM STATEMENT**  Consider the latest data on spot index and the futures price of Nikkei Stock Average 225 (NSA). Carry out VAR and Cointegration analysis. |

|  |  |  |
| --- | --- | --- |
| **VECTOR AUTOREGRESSIVE MODEL (VAR)**   * Vector Autoregression (VAR) model is an extension of univariate autoregression model to multivariate time series data. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables. * Consider a VAR model with K variables and order p:      * For a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable. * The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting.   **Data description**     * The above is the snippet of our dataset. * This is a monthly data on the Futures Price and Spot Index of Nikkei Stock Average 225 (NSA) collected from Investing.com. It contains the data for 120 months (10 years) starting from May 2012 to April 2022.   **Exploratory Data Analysis**     * The scatter plot shows there is positive correlation between Futures Price and the Spot Index      * From the above plot we can see that  1. Both the series are not mean-stationary 2. The series look Co-integrated  * **Checking for unit root Stationarity:**   We used Phillips Perron test to check unit root Stationarity of  variables  H0 : Variables are not Stationary  H1 : Variables are Stationary   1. For Futures Price : p-value = 0.1698 > 0.05 2. For Spot Index : p-value = 0.1402 > 0.05   In both the cases we do not reject H0  Hence the series are not stationary  According to the recent study, VAR model can be estimated with raw data in levels if the non-stationary data is cointegrated as theoretical work proves that estimation with such data will yield consistent parameter estimates. So we build our model on the original data which is not stationary. The following are the links we referred in order to be sure:  <https://www.tandfonline.com/doi/abs/10.1080/00036849200000119> &  <https://towardsdatascience.com/a-deep-dive-on-vector-autoregression-in-r-58767ebb3f06>  **MODEL BUILDING**   * **Finding the Optimum lags:**   We find the lags using VarSelect  AIC(n) HQ(n) SC(n) FPE(n)  1 1 1 1  All indicators account for 1 lag so we build model with one lag   * **Model Building using one lag:**   Equation for FuturesPrice and other output:    Equation for Spot\_Index and other output:     * Future Price at time t is only affected by Future Price values at time t-1, and not by Spot Index Values at time t-1 * Spot Index at time t is only affected by Future Price values at time t-1, and not by Spot Index Values at time t-1 * R2 values are greater than 0.95 indicating good fit. * All the roots are inside the Unit Circle * **Model Adequacy checking:**  1. **Serial Correlation**:   H0 : There is no Serial Correlation  H1 : There is Serial Correlation  p-value = 0.5865 > 0.05  In this test, we see that the residuals do not show signs of autocorrelation as we fail to reject the null hypothesis. So there is no serial correlation between the residuals   1. **Testing Breaks in Stability of the residuals:**     The stability test is a test for the presence of structural breaks    Based on the results of the test, there seems to be no structural breaks evident.  As such, our model passes this particular test  All the model Assumptions are satisfied  **Policy Simulations:**   1. **Granger’s causality testing**:   We checked the causality relationships between variables  Futures price Granger causes Spot Index but Spot Index does not  Granger-cause Futures price   1. **Impulse Response Function**:   We studied the impact of shock in one variable on the other variable   1. Spot Index as impulse and Futures price as response:     Shocks in Spot\_Index does not affect Futures prices, infact Only Shocks in Future Prices can afftect Future Prices   1. Futures price as impulse and Spot Index as response:     Shocks in Futures Price positively affects Spot\_Index   1. **Variance Decomposition**:   We traced the development of shocks in our system to explain the forecast error variances of both the variables   |  |  | | --- | --- | |  |  |     The forecast error of the Futures price is due to itself so shocks in Spot Index do not affect Futures price contemporaneously.  Whereas, Spot index is highly affected by shocks in Futures price.  **Forecasting**   * Forecasts were obtained using the predict() function in R, we forecasted 5 values as shown below:      * The fancharts for the forecasts are as follows:  1. **Futures Price:**      1. **Spot Index:**     **R codes**  # Load required packages to do VAR  library(tseries)  library(tidyverse)  library(urca)  library(vars)  library("mFilter")  library(forecast)  library(TSstudio)  data=read.csv(file.choose(),header=TRUE)  View(data)  data$Future.Price=rev(data$Future.Price)  data$Index=rev(data$Index)  data$Date=rev(data$Date)  View(data)  #A simple plot  library(ggplot2)  ggplot(data=data)+geom\_point(mapping=aes(x=Future.Price, y=Index))  #Positive correlation  ?ts  #Converting to time series  FuturesPrice=ts(data$Future.Price,start=c(2012,5),frequency=12)  Spot\_Index=ts(data$Index,start=c(2012,5),frequency=12)  #Plot the series  autoplot(cbind(FuturesPrice,Spot\_Index),ylab="FuturesPrice and Spot Index")  #To test Stationarity  ?pp.test  pp.test(FuturesPrice) #Not stationary  pp.test(Spot\_Index) #Not stationary  data.bv=cbind(FuturesPrice,Spot\_Index)  data.bv  #Finding the Optimum Lags  lagselect=VARselect(data.bv,lag.max=10, type="const")  lagselect$selection  ?VARselect  #All indicators account for 1 lag so we build model with 1 lag  #Building model  ?VAR  #VAR(1) model  Modeldata1=VAR(data.bv,p=1, type="const", season= NULL, exogen=NULL)  summary(Modeldata1)  #All rooots are inside the unit circle  #Dignosing the VAR Model  #Serial Correlation  ?serial.test  serial1=serial.test(Modeldata1,lags.pt=12,type="PT.asymptotic")  serial1  #p value>0.05 - So no Serial correlation  #Testing Breaks in Stability of the residuals  ?stability  Stability1=stability(Modeldata1,type="OLS-CUSUM")  plot(Stability1)  #There is stability since no curve goes above or below the red lines  # there seems to be no structural breaks evident  # Policy simulations  #VAR coefficients were not interpreted, we give preference to the  #applications  #Granger causuality  #2 vars so two casualties will be seen  ?causality  Granger\_FuturesPrice=causality(Modeldata1, cause="FuturesPrice")  Granger\_FuturesPrice$Granger  # FuturesPrice Granger-cause Spot\_Index  # Instantaneous causality between: FuturesPrice and Spot\_Index  Granger\_Spot\_Index=causality(Modeldata1,cause="Spot\_Index")  Granger\_Spot\_Index$Granger  # Spot\_Index do not Granger-cause FuturesPrice  # Instantaneous causality between: Spot\_Index and FuturesPrice  #Impulse Response Functions  #We see how a variable will behave n periods from now if I shock the other variables  #so we see how the stocks respond to stocks in itself and shocks in others  #Shock Spot Index and see Future Prices response and plot 20 periods ahead  Futures\_irf=irf(Modeldata1,impulse="Spot\_Index",response="FuturesPrice",n.ahead=20, boot=TRUE)  plot(Futures\_irf,ylab="Futures Price", main="Shock from Spot Index")  #Spot\_Index does not affect Future price, Future prices are affected by shocks in itself only  Index\_irf=irf(Modeldata1,impulse="FuturesPrice",response="Spot\_Index",n.ahead=20, boot=TRUE)  plot(Index\_irf,ylab="Spot Index", main="Shock from FuturePrices")  #Future Prices positively affects Spot\_Index  #Variance Decomposition:  #TO see how much these variables are influenced by shocks  FEVD1=fevd(Modeldata1,n.ahead=10)  plot(FEVD1)  #For FuturesPrice, Futures price affects itself if there is shock in it  #For Spot Index,shocks in Future price affects the Spot Index  #VAR Forecasting  ?predict  forecast=predict(Modeldata1,n.ahead=5,ci=0.95)  #Fanchart forecast for Futures Price  fanchart(forecast, names="FuturesPrice")  #Fanchart forecast for Spot Index  fanchart(forecast, names="Spot\_Index") |

**CO-INTEGRATION**

**Cointegration** is a [statistical](https://en.wikipedia.org/wiki/Statistical) property of a collection of [time series](https://en.wikipedia.org/wiki/Time_series) variables. First, all of the series must be integrated of order *d*. Next, if a [linear combination](https://en.wikipedia.org/wiki/Linear_combination) of this collection is integrated of order less than d, then the collection is said to be co-integrated.

Cointegration tests identify scenarios where two or more non-stationary time series are integrated together in a way that they cannot deviate from equilibrium in the long term. The tests are used to identify the degree of sensitivity of two variables to the same average price over a specified period of time.

More formally, cointegration is where two I(1) time series Xt and Yt can be described by the [stationary](https://www.statisticshowto.com/stationarity/)process  
Ut = Yt − π1 − π2Xt (π1 and π2 are constants)

**OBJECTIVE**

To check if the given two time series (Futures Price and Spot Index) are Co-Integrated or not

**ASSUMPTION**

The order of differencing required to make the time series stationary should be equal for both the time series.

|  |
| --- |
| **Exploratory Data Analysis**    Through this plot, we see that the two time series  “Futures price” and “Spot Index” appear to behave similarly over a long duration i.e. they appear to be Co-Integrated  Here, both the series (“Futures price” and “Spot Index”) appear to be non-Stationary     * There is strong Positive correlation between the two Series   **KEY OUTPUTS, PLOTS & INTERPRETATION**  Check for Stationarity using Augmented Dickey Fuller Test:  **Hypothesis:**  Ho:The data is not stationary  H1:The data is stationary  **For Futures Price**:  Original Series       * Since p-value > 0.05, the Futures Price series is not stationary hence we difference it to make it stationary.   1st Ordered Differenced Series       * Since p-value < 0.05, we conclude that the First ordered differenced series is stationary and the Future Price series is I(1)     **For Spot Index**:  Original Series       * Since p-value > 0.05, the Spot Index series is not stationary hence we difference it to make it stationary.   1st Order Differenced Series         * Since p-value < 0.05, we conclude that the First ordered differenced series is stationary and the Spot Index series is I(1) * Both the series are **I(1)**   Checking for correlation:  Correlation between Future Price and Spot Index is **0.9832**  **🡺** It can be seen that there is a strong positive correlation between “Futures price” and “Spot Index”.  To fit simple linear regression model to find hedge ratio:    **Hedge Ratio**:   * Hedge ratio comes out to be 0.96611   **Engle-Grangers Test**:  Check if the linear combination of two **I(1)** series is **I(0)**   * The linear combination is given as:   Futures Price – (Hedge Ratio\*Spot\_Index)   * The plot for the linear combination appears as follows:      * Testing for its Stationarity:      * The series is stationary since the p-value < 0.05 and hence the linear combination of **two I(1)** series is **I(0).** |
| **Conclusion**   * **Both Series are I(1) & their linear combination is I(0) series** * **Therefore the given two series are Co-Integrated**   **R Code**  library(tseries)  library(forecast)  data=read.csv(file.choose(),header=TRUE)  View(data)  data$Future.Price=rev(data$Future.Price)  data$Index=rev(data$Index)  data$Date=rev(data$Date)  View(data)  summary(data)  str(data)  #Converting to time series:  data$Future.Price=ts(data$Future.Price,start=c(2012,5),frequency=12)  data$Index=ts(data$Index,start=c(2012,5),frequency=12)  plot(data$Future.Price,xlab="Dates",  ylab="Futures Price & Index", type= "l",col="red",main="Cointegration")  lines(data$Index,col="blue")  legend("topleft",legend = c("Futures Price","Index"),  cex=0.8,inset=0.02,fill=c("red","blue"))  #------Check Stationarity Using Augmented Dickey Fuller Test  # Ho:The data is not stationary  # H1:The data is stationary  #From the plot both the series appear to be non-Stationary  # Checking if Futures price is I(1) series.  # We check stationarity of first order differenced series  adf.test(data$Future.Price, alternative="stationary",k=0) # original series: Not Stationary  adf.test(diff(data$Future.Price), alternative="stationary",k=0) # differenced series : Stationary  #Becomes stationary after one differencing  #Hence Futures.Price is an I(1) series.  # Plotting original & differenced series of Futures.Price  plot(data$Future.Price,type="l",col="blue",lwd=2)  plot(diff(data$Future.Price),type="l",col="blue",lwd=2)  # We check stationarity of first order differenced series (Index)  adf.test(data$Index, alternative="stationary",k=0) # original series: Not stationary  adf.test(diff(data$Index), alternative="stationary",k=0) # differenced series: Stationary  # First order differenced series is Stationary, hence the series is I(1)  # Plotting original & differenced series of Spot\_Index  plot(data$Index,type="l",col="blue",lwd=2)  plot(diff(data$Index),type="l",col="blue",lwd=2)  #----- Check correlation  # We check how these two series are related with correlation command  cor(data$Future.Price,data$Index)  library(ggplot2)  ggplot(data=data)+geom\_point(mapping=aes(x=Future.Price, y=Index))  #strong positive relation between the series is seen  # Fitting simple linear regression model to find hedge ratio  model<-lm(data$Future.Price~data$Index)  summary(model)  #Save and Print Regression Coefficients  Regression\_Coefficients<-model$coefficients  print(model$coefficients)  #Determine and Save Value of Beta (Hedge Ratio)  Hedge\_Ratio<-Regression\_Coefficients[2]  print(Hedge\_Ratio)  # calculating Spread i.e. linear combination of the two series  spread<- data$Future.Price-(Hedge\_Ratio\*data$Index)  plot(spread,type="l",col="blue",lwd=2)  head(spread)  tail(spread)  # Checking if Spread is I(0) series  # We check stationarity of the series  adf.test(spread,k=0)  #The series is stationary  # The spread is I(0)  # Both Series are I(1) & their linear combination is I(0) series  # Therefore the given two series are co-integrated |