VAR Model of Economic Time Series

Rahul .M

Thursday, 28 January 2021

1. Extracting and Manipulating Data from sources and creating a data frame
2. Checking the functionality of ACF and PACF
3. Checking the stationarity of the model
4. Finding the optimal lags for the model
5. Building VAR model
6. Performing VAR Diagnostics
7. Granger-Causality , Impulse response functions , Forecast Error variance Decomposition
8. Forecasting and Visualizing with fan charts

VAR Model of Economic Time Series
Rahul .M
Thursday, 28 January 2021

In the 1980s, Robert Litterman and Christopher Sims developed forecasting models based on vector auto regressions (VAR). The models use aggregate Economic variables that reflect in micro and macroeconomic series

There is an extensive literature on VAR modelling; see the citations in Pfaff (2008). The papers of Litterman and Sims in the references provide a good introduction to the mathematical framework for specifying vector auto regression models in a Bayesian framework. Sims, extending the model of Litterman, accommodates time-varying variances of the disturbance/innovation terms, and non-Gaussianity of these disturbances

The analysis in the following sections uses the R package to collect economic time series and fit vector-autoregressive model to a reduced set of these economic variables

In the case study I have mentioned many parameters that will have an impact on Residential Home Prices in USA

But for the scope of the case study I have considered few economic time series for forecasting the VAR model

Collecting the Economic Data

The Model is Built in R and the below packages are used in this model

- Fredr
- graphics
- quantmod
- tidyr
- dplyr
- ggplot2
- plotly
- vars
- tidyselect
- tseries
- forecast

Using the FREDR package that was built by **Sam Boyse** and with the help of an API key I can able to fetch the data from the FRED database

Economic time series used

- monthly supply homes : The months' supply is the ratio of houses for sale to houses sold
- unemployment rate : the number of unemployed as a percentage of the labour force
- federal funds interest rate : interest rate at which depository institutions trade federal funds
- recession : Smoothed recession probabilities for the United States
- housings starts : estimates of housing starts include units in structures being totally rebuilt on an existing foundation

VAR Model of Economic Time Series
Rahul .M
Thursday, 28 January 2021

The above Economic time series are indicated with a unique series number in FRED and using that unique series number I called from FRED database into my local R Environment.

Using the Graphics I can able to visualize the Real-time economic time series graph as per present date

Each series consists of 3 features

- Data
- Series_ id
- value

For our requirement we need only 2 variables so I discarded the series_id as it won't be helping at all for the model development

Then I changed the column names as per Economical terminology for better understanding

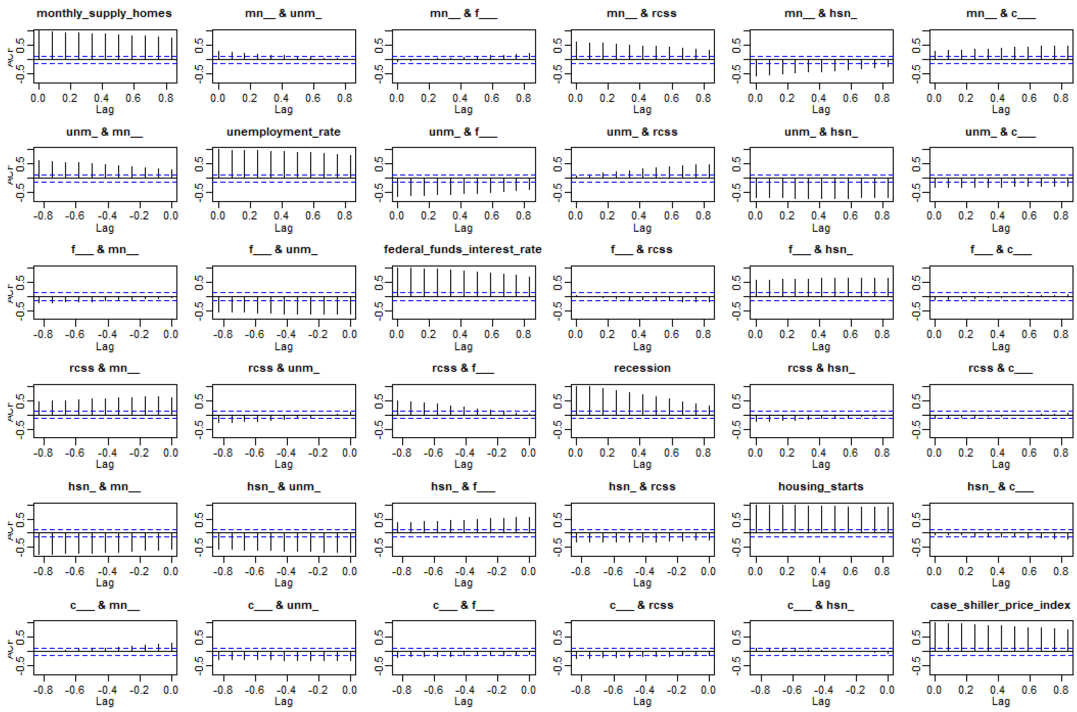
The next step is I merged all the series into a new data frame

Then converted the Economic series variables into time series variables and bind into a new time series data frame

Hence the new tidy data for the model input has been created

Plotting ACF and PACF graphs

The function ACF computes estimates of the auto covariance or autocorrelation function.



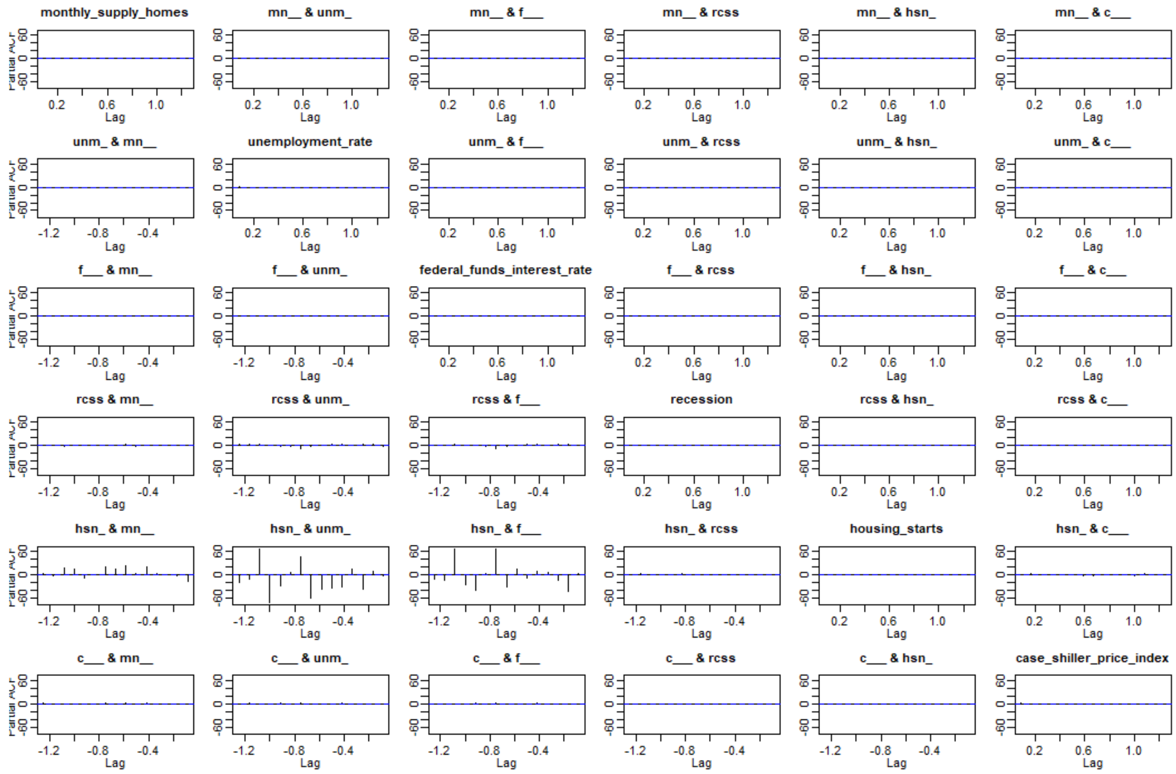
The ACF shows the nature of the observation at every lag. As it would be helpful in finding the predictions

From the above figure all the parameters (Diagonal elements) itself have a strong correlation with the observations at every lag and we can say that they are statistically significant

For the rest of the elements ACF says that there is negative correlation and not statistically significant

The parameters which are below the significance levels (Below the blue dotted lines) are not statistically significant for the model

Function PACF is the function used for the partial autocorrelations.



From the above graph we can say that, Assume that our dependent variable is Case-Shiller price index(y) say one of the independent variable is monthly supply of homes(x1) and The partial correlation between y and(x1) is the correlation between the variables determined taking into account how both y and (x1) are related with the other independent variables

So the rest of the independent variables are statistically insignificant according to PACF

VAR Model of Economic Time Series

Rahul .M

Thursday, 28 January 2021

Dickey-Fuller test:

The Augmented Dickey Fuller Test (ADF) is unit root test for stationarity. Unit roots can cause unpredictable results in your time series analysis

Augmented Dickey–Fuller test (ADF) tests the null hypothesis that a unit root is present in a time series sample i.e. Non- stationarity

The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity.

Assuming significance level alpha =0.05

Low p-values indicate stationarity. So when we run ADF on white noise, we expect to see a low p-value

High p-values indicate non- stationarity. We fail to reject the null hypothesis

S.no	Augmented Dickey –Fuller Test	P-value	Result
1	Monthly supply homes	0.6754	fail to reject the null hypothesis
2	Unemployment rate	0.4594	fail to reject the null hypothesis
3	Federal funds interest rate	0.06875	fail to reject the null hypothesis
4	recession	0.04462	reject the null hypothesis
5	Housing starts	0.8957	fail to reject the null hypothesis
6	Case shiller price index	0.713	fail to reject the null hypothesis

Now we need to find out the number of optimal lags that are used by the model and we can calculate it by using the following methods

- Akaike information criterion
- Hannan–Quinn information criterion
- Scwartz (SC) information
- Final Prediction Error

S.no	Method	Lag value
1	Akaike information criterion	9
2	Hannan–Quinn information criterion	3
3	Scwartz (SC) information	3
4	Final Prediction Error	4

According to HIC and SC criterion both got approximate result and tend to go with the optimal lag value as 3

Building a VAR model

As we already created the time series data frame and using the VAR function with the optimal lag value as 3 based on HIC and SC

VAR Model of Economic Time Series

Rahul .M

Thursday, 28 January 2021

As it is Non- stationarity the type I used in VAR is trend and there is no seasonality and Exogenous variables in the series

As we can see the VAR Estimation results are not significant as we don't have significance lags in the model

But what's important is the all the roots are inside the unit circle. So we can say that the model is stable

Roots of the characteristic polynomial
0.9925
0.9912
0.9912
0.9414
0.9414
0.8888
0.8312
0.8312
0.6029
0.6029
0.5574
0.5574
0.4801
0.4801
0.3933
0.3933
0.24
0.04486

VAR Diagnostics

We are running couple of tests to test the model or to diagnose the model

Serial correlation

- Serial correlation is the relationship between a given variable and a lagged version of itself over various time intervals.
- It measures the relationship between a variable's current values given its past values.

S.no	Test Type	P-value	Result
1	Serial Correlation	$2.887e^{-5}$	Failed to reject null hypothesis

Here for the serial test for the Economic series is tests for autocorrelation, and finds that the null of autocorrelation cannot be rejected because the p-value $2.887e^{-5}$ is less than the significance level of 0.05
There is some Autocorrelation present in the model

Null Hypothesis: Autocorrelation
~~Alternate Hypothesis: No Autocorrelation~~

Heteroscedasticity test

- From Ancient Greek hetero "different" and skedasis "dispersion".
- If the variability of the random disturbance is different across elements of the vector.
- Here, variability could be quantified by the variance or any other measure of statistical dispersion. Thus heteroscedasticity is the absence of homoscedasticity.
- In time series takes the form of arch effects .Essentially those are periods of Volatility
  So we are trying to test a bit of volatility here

S.no	Test Type	P-value	Result
1	Heteroscedasticity test	1	Reject null hypothesis

~~Null Hypothesis: model suffers from Heteroscedasticity [P- value <0.05]~~
Alternate Hypothesis: no effect of Heteroscedasticity [P- value >0.05]
Here the model won't suffer from Heteroscedasticity

Normality test

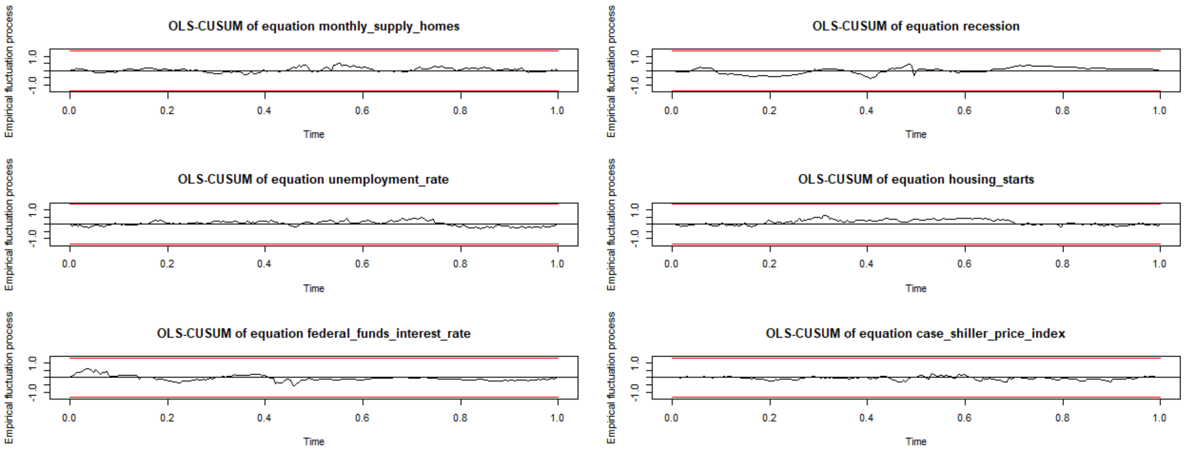
- Normality tests are used to determine if a data set is well-modelled by a normal distribution
- And to compute how likely it is for a random variable underlying the data set to be normally distributed
- We want our residuals to be normally distributed and in R it runs 3 types of tests for normality mentioned below
- JB-Test (multivariate)
- Skewness only (multivariate)
- Kurtosis only (multivariate)

S.no	Test Type	P-value	Result
1	JB-Test	$2.2e^{-16}$	Reject null hypothesis
2	Skewness only	$2.2e^{-16}$	Reject null hypothesis
3	Kurtosis	$2.2e^{-16}$	Reject null hypothesis

Here for this model the tests for the normality are insignificant as all of the values are less than 0.05

Stability test

- Tests for structural change in linear regression models from the generalized fluctuation test framework as well as from the F test (Chow test) framework.
- This tests the stability of residuals
- Notice there are no points in the graph which exceeds the two red lines upper and lower confidence interval
- Hence we can say that the model is stable



Structural Stability graph

Granger causality

- The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another
- A time series X is said to Granger-cause Y if it can be shown, usually through a series of t-tests and F-tests on lagged values of X (and with lagged values of Y also included), that those X values provide statistically significant information about future values of Y
- When compared to the **OLS** we can test from unidirectional but using Granger we can test unidirectional , bidirectional and granger cause itself
- We want to see the association between variables more of a causality than a correlation
- We can do 2 methods (Granger and Instant)

---

1. Monthly supply of homes

msh_granger   <- causality(model, cause = 'monthly_supply_homes')

Null Hypothesis: monthly_supply_homes do not Granger-cause
Unemployment rate, federal funds interest rate recession housing starts
Case Shiller price index
Reject Null Hypothesis if P value is less than 0.05
P – Value: 0.001804

---

2. Monthly supply of homes

uem_granger   <- causality(model , cause = 'unemployment_rate')

Null Hypothesis: Unemployment rate do not Granger-cause monthly supply homes, federal funds interest rate, recession, Housing starts, case Shiller price index

Reject Null Hypothesis if P value is less than 0.05
P – Value: 0.03932

---

3. ffir_granger   <- causality(model , cause = 'federal_funds_interest_rate')

   H0: unemployment rate do not Granger-cause, monthly supply homes federal funds interest rate, recession, housing starts, case Shiller price index

Reject Null Hypothesis if P value is less than 0.05
P – Value: 0.05728

---

4. rec_granger   <- causality(model , cause = 'recession')

   H0: recession do not Granger-cause monthly supply homes, federal funds interest rate, recession, housing starts, case Shiller price index , unemployment

Reject Null Hypothesis if P value is less than 0.05
P – Value: 9.904e^{-05}

---

5. hs_granger <- causality(model , cause = 'housing_starts')

   H0: housing starts do not Granger-cause monthly supply homes, federal funds interest rate, recession, case Shiller price index, unemployment rate

   Reject Null Hypothesis if P value is less than 0.05
   P – Value: 0.5487

   ---

6. cpi_granger <- causality(model , cause = 'case_shiller_price_index')

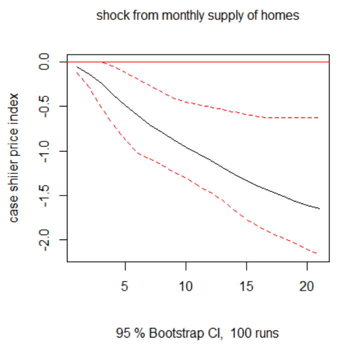
   Granger causality H0: case Shiller price index do not Granger-cause, cause monthly supply homes, federal funds interest rate, recession, unemployment rate, housing starts
   Reject Null Hypothesis if P value is less than 0.05
   P – Value: 0.0004064

Impulse Response functions

- Their main purpose is to describe the evolution of a model's variables in reaction to a shock in one or more variables.
- This feature allows to trace the transmission of a single shock within an otherwise noisy system of equations and, thus, makes them very useful tools in the assessment of economic policies
- For example say I shocked the monthly supply homes and I want to see the response of the Home Price index

Impulse in Monthly supply of homes

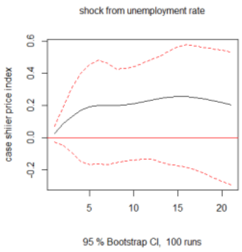
```
msh_irf   <- irf(model , impulse= 'monthly_supply_homes' ,
            response = 'case_shiller_price_index' ,
            n.ahead = 20 ,
            boot = TRUE)
```



Here in the graph the shock is caused due to monthly supply of homes and it negatively affects the Case Shiller price index

Impulse in Monthly supply homes

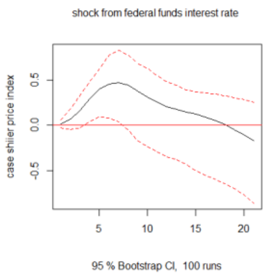
```
uer_irf <- irf(model , impulse= 'unemployment_rate' ,
         response = 'case_shiller_price_index' ,
         n.ahead = 20 ,
         boot = TRUE)
```



Shock from unemployment will have an positive affect on case shiller price index over the time periods

Impulse in Federal funds interest rate

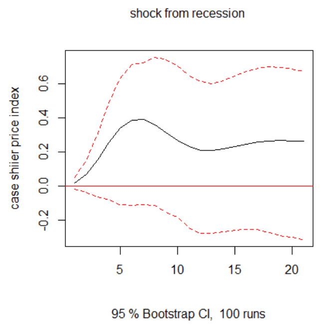
```
ffir_irf <- irf(model , impulse= 'federal_funds_interest_rate' ,
        response = 'case_shiller_price_index' ,
        n.ahead = 20 ,
        boot = TRUE)
```



Initially over the time period the case shiller price index was increased and as the time period are increasing there is a decline in the price index

Impulse in Recession

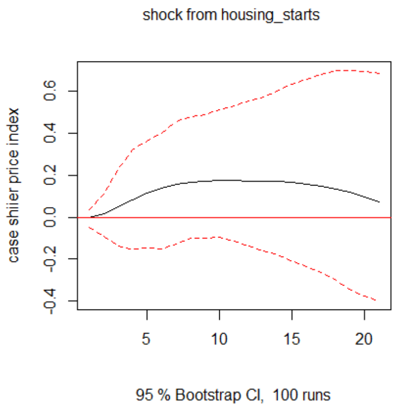
```
rec_irf <- irf(model , impulse= 'recession' ,
        response = 'case_shiller_price_index' ,
        n.ahead = 20 ,
        boot = TRUE)
```



Initially over the time period the case shiller price index was increased and as the time period is increasing there is a decline in the price index and there is a lot of room for errors as there is a large gap for the confidence interval

Impulse in Housing Starts

```
hs_irf <- irf(model , impulse= 'housing_starts' ,
        response = 'case_shiller_price_index' ,
        n.ahead = 20 ,
        boot = TRUE)
```

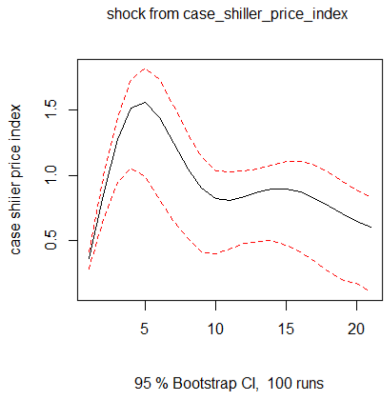


The response of CPI from Housing Starts are exponentially adjusted and in the long run there is a decline in the CPI and at in beginning there is a strong Increase in CPI

Impulse in Case Shiller price index

```
cpi_irf <- irf(model , impulse= 'case_shiller_price_index' ,
         response = 'case_shiller_price_index' ,
         n.ahead = 20 ,
```

boot = TRUE)



The Shock from CPI is positively affecting the response of CPI
It is clearly visible from the graph

Variance Decomposition

- In econometrics and other applications of multivariate time series analysis, a variance decomposition or forecast error variance decomposition (FEVD) is used to aid in the interpretation of a vector auto regression (VAR) model once it has been fitted
- The variance decomposition indicates the amount of information each variable contributes to the other variables in the auto regression. It determines how much of the forecast error variance of each of the variables can be explained by exogenous shocks to the other variables
- How much these variables are influenced by the shocks
- From the graph all of the series are influenced by their own shocks
- At initial period the change in the parameters are with respect to their own shocks and at longer periods the shock from other series are considerable at low amounts
- Graphs are in GitHub

VAR forecasting

- Using predict function in R we can forecast the vale Economic time series variables for the next ten periods on the time frame with a confidence interval of 95% and I used the fan charts to visualize the graph
- Graphs are in GitHub
- Overall the Case Shiller Price Index was increased at initial period and it decreased at longer periods