



Vector Autoregression

White Paper



December, 2012
www.mu-sigma.com

Vector Autoregression

Prabakar Rajasekaran

December 13, 2012

1 Introduction

Vector autoregression (VAR) is an econometric model used to capture the evolution and the interdependencies between multiple time series models. The VAR approach was made popular by Sims (1980). This approach has many advantages over the traditional structural modeling approach. VAR treats all variables in the model as endogenous. It considers the lagged terms of all endogenous variables as exogenous, where each of the dependent variable at time t depends on different combinations of all independent variables at time $t - n$ with error term. Since there is no contemporaneous exogenous variable, each equation can be estimated using a simple regression procedure.

The four variable VAR model has been used to illustrate the model procedures and to build impulse response function, where the variables being used are stocks trading in NYSE like *FAS*, *USO*, *UUP* and *VIX*.

2 Stationarity Test

Before building a VAR model, it is necessary to check the stationarity of the data, for which we use augmented Dickey-Fuller (ADF) test. A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space. Consequently, parameters such as the mean and variance, if they exist, also do not change over time or position. Stationarity is used as a tool in time series analysis, where the raw data are often transformed to become stationary. The ADF test results for the variables are given below:

```
> # Loading the packages vars and tseries
> library(vars)
> library(tseries)
> # ADF Test for FAS
> adf.test(FAS)
```

Augmented Dickey-Fuller Test

```
data: FAS
Dickey-Fuller = -3.8027, Lag order = 10, p-value = 0.01886
alternative hypothesis: stationary
```

```
> # ADF Test for USO
> adf.test(USO)
```

Augmented Dickey-Fuller Test

```
data: USO
Dickey-Fuller = -2.4703, Lag order = 10, p-value = 0.3792
alternative hypothesis: stationary
```

```
> # ADF Test for UUP
> adf.test(UUP)
```

Augmented Dickey-Fuller Test

```
data: UUP
Dickey-Fuller = -2.285, Lag order = 10, p-value = 0.4577
alternative hypothesis: stationary
```

```
> # ADF Test for VIX
> adf.test(VIX)
```

Augmented Dickey-Fuller Test

```
data: VIX
Dickey-Fuller = -2.9432, Lag order = 10, p-value = 0.1791
alternative hypothesis: stationary
```

The results show the probability value (p-value) for which the null hypothesis is nonstationary. For most variables, the p-values are greater than 0.05, which means we accept the null hypothesis and reject the alternative hypothesis of stationary. The stationarity can also be viewed by the time series plot of these variables shown in *Figure 1*. Since all variables are non stationary, to make it stationary the first difference of the variables can be performed.

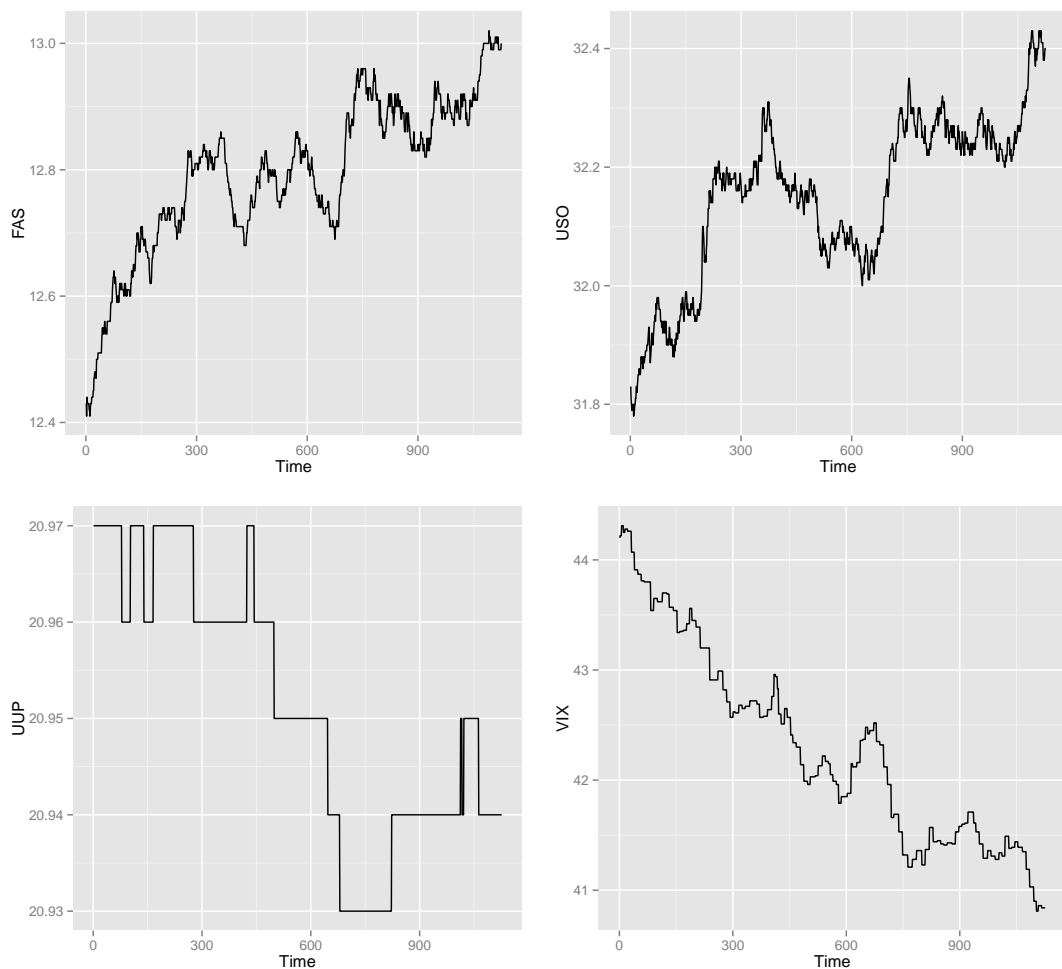


Figure 1: Time series plots of the data

3 VAR Model

Before going into the standard VAR model, let's consider the following simultaneous equation model:

$$BX_t = A(L)X_{t-n} + Ce_t \quad (1)$$

where X_t is the vector of endogenous variable, e_t is the vector of white noise error term, B is the coefficients of the contemporaneous variables matrix and $A(L)$ is the coefficients of lagged variables matrix.

Now, the standard VAR model can be represented from equation (1) by:

$$X_t = B^{-1}A(L)X_{t-n} + B^{-1}Ce_t \quad (2)$$

The above equation can be transformed to the generalized form below:

$$X_t = D(L)X_{t-n} + u_t \quad (3)$$

where, $D(L) = B^{-1}A(L)$ and $u_t = B^{-1}Ce_t$

Equation (3) is called the standard VAR framework, where u_t is the reduced form error from the standard VAR framework. The reduced form error (u_t) is represented as a function of the uncorrelated shocks (e_t) and e_t is serially uncorrelated with the shocks of other variables.

4 Modelling

The variables are made stationary by taking the first difference. Before building the VAR model, it is necessary to calculate the optimal lag used in the model, which is estimated using information criteria like AIC or BIC. The information criteria results are given below:

```
> # Calculating the optimal lag
> ic <- VARselect(data[1:95,], lag.max=6)
> ic

$selection
AIC(n)  HQ(n)  SC(n)  FPE(n)
      1      1      1      1

$criteria
          1          2          3          4          5
AIC(n) -3.920532e+01 -3.900770e+01 -3.889271e+01 -3.864766e+01 -3.898587e+01
HQ(n)   -3.897991e+01 -3.860195e+01 -3.830663e+01 -3.788124e+01 -3.803913e+01
SC(n)   -3.864608e+01 -3.800106e+01 -3.743867e+01 -3.674623e+01 -3.663705e+01
FPE(n)  9.409166e-18  1.149153e-17  1.296488e-17  1.674128e-17  1.214488e-17
          6
AIC(n) -3.883303e+01
HQ(n)   -3.770596e+01
SC(n)   -3.603681e+01
FPE(n)  1.452060e-17
```

From the above results, the optimal lag to be chosen is 1 and it is selected based on the minimum value of the information criteria results, which is shown in *Figure 2*, where the AIC and HQ values are minimum at lag 1. In R, the *VAR* function gives only the coefficients of the equations as the output. If we need to know the detailed result of the

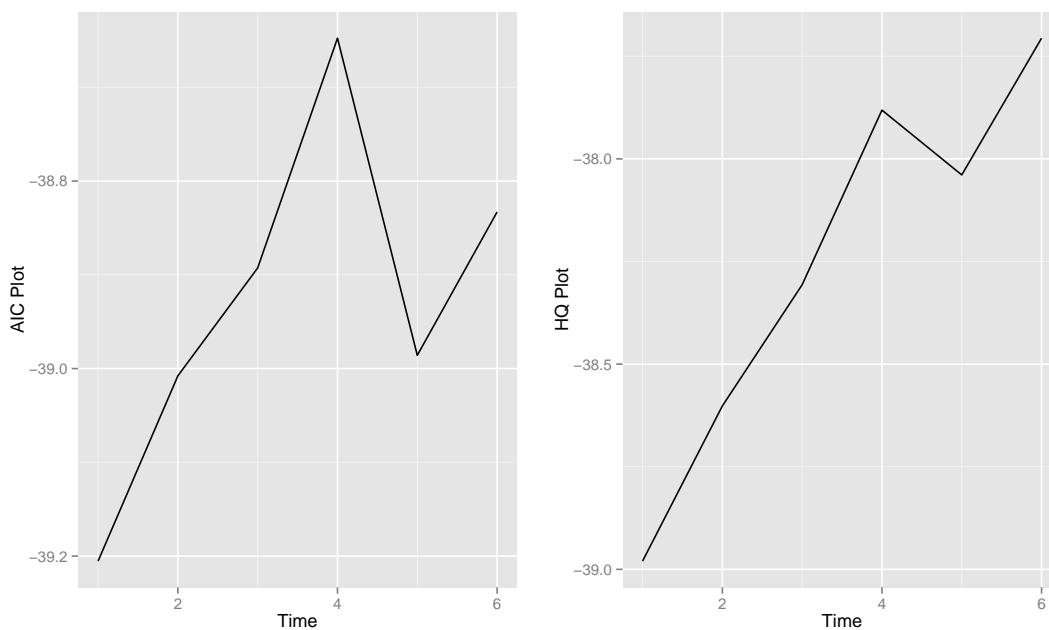


Figure 2: Plot of the Information Criteria results

VAR estimation, we can use the *summary* command.

```
> # Building the VAR model
> var<-VAR(data[1:95,],p=1,type="const")
> summary(var$varresult$FAS)
```

Call:

```
lm(formula = y ~ -1 + ., data = datamat)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.0294829	-0.0051496	-0.0006495	0.0047322	0.0249602

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
FAS.11	0.72200	0.06694	10.786	< 2e-16 ***
USO.11	0.26029	0.06914	3.765	0.000299 ***
UUP.11	-0.08397	0.32481	-0.259	0.796595
VIX.11	-0.01886	0.00929	-2.030	0.045302 *
const	-2.22531	6.18307	-0.360	0.719772

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.008352 on 89 degrees of freedom

Multiple R-squared: 0.9855, Adjusted R-squared: 0.9849

F-statistic: 1513 on 4 and 89 DF, p-value: < 2.2e-16

The VAR results would show four equations (for convenience the result of FAS equation alone is shown above), where the first equation shows the value of FAS at time t as the dependent variable regressed with respect to the lag of all the other variables and the second equation would show USO as the dependent variable regressed with respect to the lag of all the variables and so on.

5 Forecasting

The forecasting generally works by keeping the calculated coefficients from the above model fixed and adding the new present values into the equation to get the n step ahead forecasted value.

Since in the above example, we used the contemporaneous matrix (B matrix) as an identity matrix (ie for a FAS as dependent variable equation, FAS at time t is not influenced by any other variable at time t). So the dependent variables are equated only on the lagged independent variables.

From equation (1)

$$B X_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \text{FAS}_t \\ \text{USO}_t \\ \text{UUP}_t \\ \text{VIX}_t \end{bmatrix}$$

So the forecast for each equation happens separately, the results shows five period ahead forecast for FAS alone.

```
> # Forecasting based on the VAR model
> predicted<-predict(var,n.ahead = 5, ci =0.95)
> predicted$fcst$FAS
```

	fcst	lower	upper	CI
[1,]	12.61130	12.59493	12.62767	0.01637018
[2,]	12.60538	12.58327	12.62750	0.02211308
[3,]	12.60119	12.57450	12.62787	0.02668408
[4,]	12.59807	12.56727	12.62887	0.03079727
[5,]	12.59564	12.56102	12.63025	0.03461640

where *var* is the name given to our VAR model , *n.ahead* is the period ahead predictions required and *ci* is the confidence interval and *fcst* is the forecasted value of FAS.

The predicted vs actual value is shown in *Figure 3*.

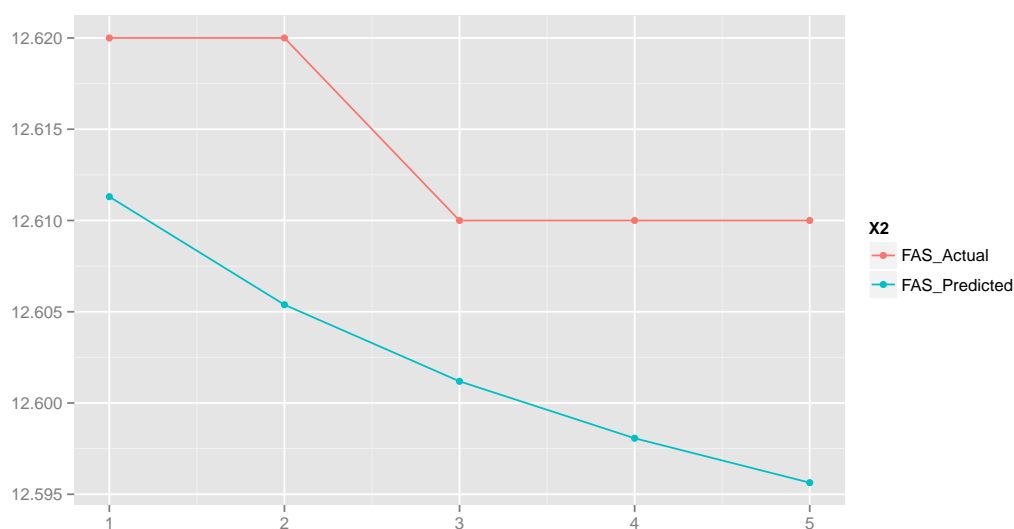


Figure 3: FAS Actual vs Predicted Graph

6 Impulse Response Function (IRF)

Impulse response trace out the responsiveness of the dependent variables in the VAR model to shocks to each of the other variables. So, for each variable from each equation separately, a unit shock applied to the error term for the particular equation affects upon the VAR system over time. Thus, if there are g variables in a system, a total of g^2 impulse responses could be generated. This is achieved in practice by expressing the VAR model as a VMA – that is, the vector autoregressive model is written as a vector moving average (in the same way as done for univariate autoregressive models). Provided that the system is stable, the shock should gradually die away.

The IRF can be explained through an example given below:

consider the equation

$$Y_t = AY_{t-1} + u_t \quad (4)$$

where

$$A = \begin{bmatrix} 0.5 & 0.3 \\ 0 & 0.2 \end{bmatrix}$$

Consider the effect of shock at time $t = 0$ which effects the other variable at time $t = 1$,

$$Y_0 = \begin{bmatrix} u_{10} \\ u_{20} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$Y_1 = A * Y_0 = \begin{bmatrix} 0.5 & 0.3 \\ 0 & 0.2 \end{bmatrix} * \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$$

$$Y_2 = A * Y_1 = \begin{bmatrix} 0.5 & 0.3 \\ 0 & 0.2 \end{bmatrix} * \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0 \end{bmatrix}$$

Although it is fairly easy to see what the effects of shocks to the variables will be in such a simple VAR, the same principles can be applied in the context of VARs containing more equations or more lags, where it is much more difficult to see the interactions between the equations.

The impulse response graph for above model is shown in *Figure 4*, where the graph shows how a unit shock at time 0 for FAS affects the future values of all other variables. The graph also shows that the shocks get diminished as time progress.

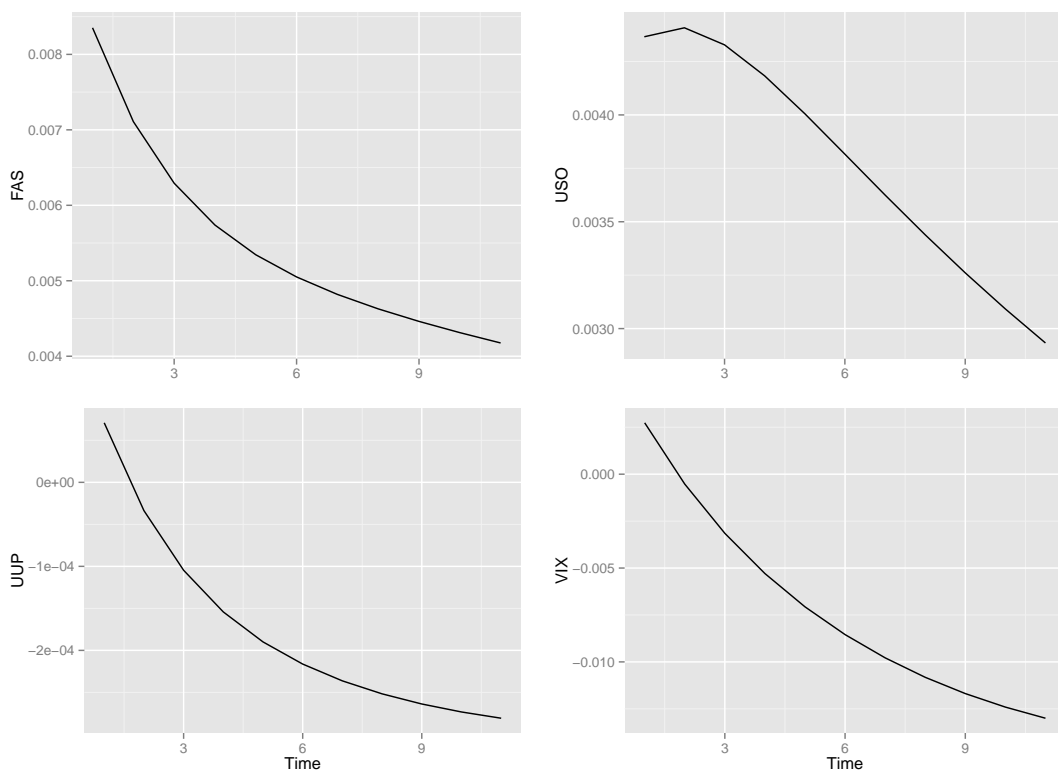


Figure 4: Impulse Response Graph

7 Summary and Future work

The vector autoregression (VAR) model is a structure of equations, where the check for stationarity is performed before building a model, if the variables are stationary then the next process is to find the optimal lag using the information criteria's like AIC or BIC. The optimal lag is used to build the VAR model. The VAR model can be used for forecasting and building the impulse response function model.

Another interesting approach would be to use an extension of the VAR model known as structural VAR. In this method, we introduce independent variables at time t in addition to the auto regressive terms. For example, in the above VAR model, FAS at time t is influenced by the lagged values of all the other variable. In the structural VAR method, we can have other variables like USO or UUP at time t in the FAS equation as well.

8 References

- Chris Brooks, (2008). "Introductory Econometrics for Finance", Second Edition, Cambridge University Press .
- Hamilton, James D. (1994). "Time Series Analysis", Princeton University Press. (p. 293)
- Frank Smets Gert Peersman, 2001. "The monetary transmission mechanism in the Euro area: more evidence from VAR analysis" (MTN conference paper), Working Paper Series 091, European Central Bank

Mu Sigma is a leading provider of decision sciences and analytics solutions, helping companies institutionalize data-driven decision making. We work with market-leading companies across multiple verticals, solving high impact business problems in the areas of Marketing, Supply Chain and Risk analytics. For these clients we have built an integrated decision support ecosystem of people, processes, methodologies & proprietary IP and technology assets that serve as a platform for cross-pollination and innovation. Mu Sigma has driven disruptive innovation in the analytics industry by integrating the disciplines of business, math, and technology in a sustainable model. With over 75 Fortune 500 clients and over 2000 decision science professionals we are one of the largest pure-play decision sciences and analytics companies.

Learn more at <http://www.mu-sigma.com/contact.html> us for further information:

Mu Sigma Inc., 3400 Dundee Rd, Suite 160, Northbrook, IL – 60062

www.mu-sigma.com

© Copyright 2012 - 2013 Mu Sigma Inc.

No part of this document may be reproduced or transmitted in any form or by any means electronic or mechanical, for any purpose without the express written permission of Mu Sigma Inc. Information in this document is subject to change without prior notice.