TADA

Vector Autoregressive (VAR) Model

Mutivariate Time Series Forecasting

Hannah Do | 12.10.2022

Contents

1. Introduction

Terminology

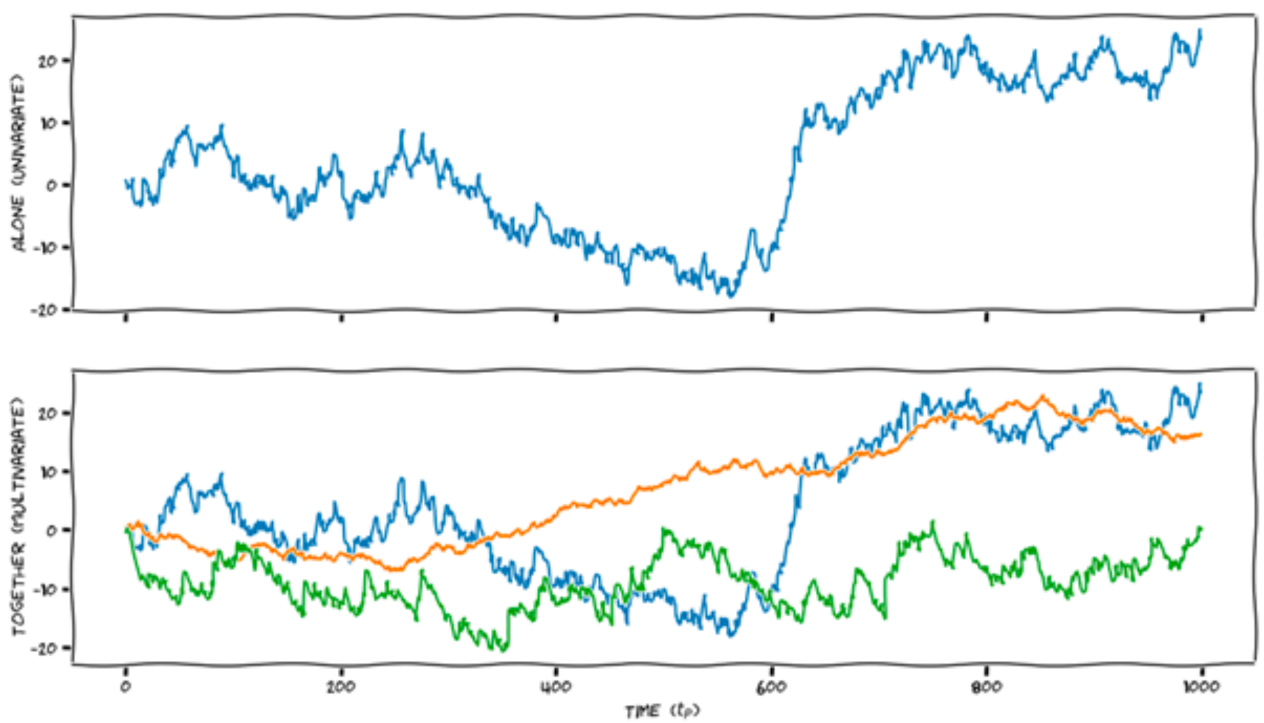
- **Univariable** : One independent variable
- **Univariate** : One dependent (target) variable (y = 1)
- **Multivariable** : Multiple independent variables
- **Multivariate** : Multiple dependent (target) variables (y > 1)

Ex) 4 independent variables & 1 dependent variable :

→ Multivariable regression analysis

→ Univariate regression analysis

1. Introduction



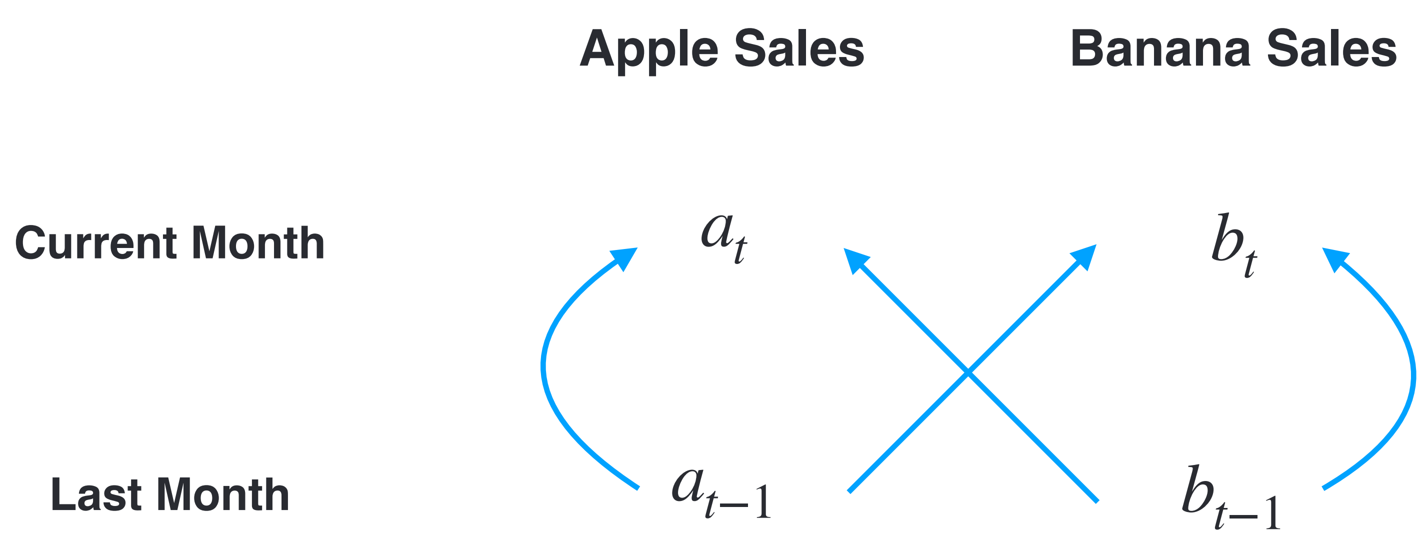
Prophet

VAR

VAR : Vector AutoRegressive model

- Introduced by Christopher Sims (1980) as a macroeconometric framework.
- Statistical model used to capture the relationship between multiple quantities as they change over time.

2. Basic Concept



VAR(1)

VAR 1 Model : Model for one time period

$$a_t = C_{11}a_{t-1} + C_{12}b_{t-1} + \epsilon_{a_1 t}$$

$$b_t = C_{21}a_{t-1} + C_{22}b_{t-1} + \epsilon_{b_1 t}$$

2. Basic Concept

ritvikmath : *Vector Auto Regression: Time Series Talk*

VAR(1)

VAR 1 Model : Model for one time period

$$a_t = C_{11}a_{t-1} + C_{12}b_{t-1} + \epsilon_{a_1 t}$$

$$b_t = C_{21}a_{t-1} + C_{22}b_{t-1} + \epsilon_{b_1 t}$$

fruits vector - this month

$$f_t = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} a_{t-1} \\ b_{t-1} \end{bmatrix} + \epsilon_t$$

epsilon vector (errors)

constants vector (coefficients) C

f_{t-1} fruits vector - last month

$$f_t = Cf_{t-1} + \epsilon_t$$

VAR(x)

VAR x Model : Model for x time periods

$$a_t = C_{11}a_{t-1} + C_{12}b_{t-1} + \ldots + C_{1x}a_{t-x} + C_{1x}b_{t-x} + \epsilon_{a_x t}$$

$$b_t = C_{21}a_{t-1} + C_{22}b_{t-1} + \ldots + C_{2x}a_{t-x} + C_{2x}b_{t-x} + \epsilon_{b_x t}$$

2. Basic Concept

What makes up a VAR model?

- Number of endogenous variables
- Number of autoregressive terms

A VAR model is composed of n-equations (representing n endogenous variables) and includes x-lags of the variables (representing x autoregressive terms).

Below model can be regarded as a Bivariate VAR(3) model.

VAR(3)

VAR 3 Model : Model for 3 time periods

$$a_t = C_{11}a_{t-1} + C_{12}b_{t-1} + C_{13}a_{t-2} + C_{14}b_{t-2} + C_{15}a_{t-3} + C_{16}b_{t-3} + \epsilon_{a_3t}$$

$$b_t = C_{21}a_{t-1} + C_{22}b_{t-1} + C_{23}a_{t-2} + C_{24}b_{t-2} + C_{25}a_{t-3} + C_{26}b_{t-3} + \epsilon_{b_3t}$$

3. Lag selection

1. Determining the number of lags in a VAR model

- A maximum number of lags p_{max} is normally chosen for model performance evaluation.
  - log likelihood (l)
  - number of parameters (k)
  - number of samples used for fitting (n)
  - **Akaike's Information Criteria**
  - $AIC = 2k - 2l$
    - Low AIC is preferred -> higher log likelihood with **less parameters**
  - **Bayesian Information Criteria**
  - $BIC = \ln(n)k - 2l$
    - Low BIC is preferred -> higher log likelihood with less parameters and **less samples used in fitting**

VAR Lag Order Selection Criteria
Endogenous variables: GDP
Exogenous variables: C
Date: 02/09/18 Time: 16:07
Sample: 1970Q1 1991Q4
Included observations: 80

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-622.1049	NA	341036.8	15.57762	15.60740	15.58956
1	-400.5328	432.0655	1373.961	10.06332	10.12287	10.08720
2	-394.4911	11.63021*	1211.284*	9.937278*	10.02660*	9.973092*
3	-394.4378	0.101412	1240.352	9.960944	10.08005	10.00870
4	-394.4375	0.000484	1271.845	9.985938	10.13481	10.04563
5	-394.4124	0.046409	1303.379	10.01031	10.18896	10.08194
6	-394.3900	0.040880	1335.849	10.03475	10.24318	10.11831
7	-394.3862	0.006886	1369.839	10.05965	10.29786	10.15516
8	-394.2344	0.269514	1399.595	10.08086	10.34884	10.18830

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

CrunchEconometrix : Lag Structure for gdp (EViews)

3. Lag selection

2. Choosing the right variables

Granger-causality statistics

Granger-causality statistics test whether one variable is statistically significant when predicting another variable.

Part of F-statistics that test if the coefficients of all lags of a variable are jointly equal to zero in the equation for another variable.

As the p-value of the F-statistic decreases, evidence that a variable is relevant for predict another variable increases.

Ex) On a Granger-causality test of X on Y, if the p-value is less than 0.05 -> X does help predict Y

If p-value is greater than 0.05 -> X does not help predict Y

4. Inference & Evaluation

VAR model equation can be estimated using OLS given following circumstances :

1. The Error Term has a conditional mean of zero
    - No matter which value we choose for X, the error term u must not show any systematic pattern and must have a mean of 0.
2. The variables in the model are stationary
    - Statistical properties of a process generating a time series do not change over time
    - No seasonality, no trend
3. Large Outliers are Unlikely
    - OLS suffers from sensitivity to outliers

4. Inference & Evaluation

Forecasting

Forecasts are generated for VAR models using an iterative forecasting algorithm:

1. Estimate the VAR model using OLS for each equation.
2. Compute the one-period-ahead forecast for all variables.
3. Compute the two-period-ahead forecasts, using the one-period-ahead forecast.
4. Iterate until the h-step ahead forecasts are computed.

Evaluation

1. Granger causality tests :
   - Testing whether one variable is useful in forecasting another (Wald Test)
2. Impulse response analysis :
   - The response of one variable to a sudden but temporary change in another variable
3. Forecast error variance decomposition :
   - The proportion of the forecast variance of each variable is attributed to the effects of the other variables.

5. Applications

Example question	Field	Description
How are vital signs in cardiorespiratory patients dynamically related?	Medicine	A VAR system is used to model the past and current relationships between heart rate, respiratory rate, blood pressure and SpO2.
How do risks of COVID-19 infections interact across age groups?	Epidemiology	Count data of past infections across different age groups was used to model the relationships between infection rates across those age groups.
Is there a bi-directional relationship between personal income and personal consumption spending?	Economics	A two-equation VAR system is used to model the relationship between income and consumption over time.
How can we model the gene expression networks?	Biology	The relationships across large networks of genes are modeled using a sparse structural VAR model.
What is driving inflation more -- monetary policy shocks or external shocks?	Macroeconomics	A structural VAR model is used to compute variance decomposition and impulse response functions following monetary shocks and external system shocks.

6. References

1. Ritvikmath (Oct 4, 2019) Vector Auto Regression : Time Series Talk
    - hhttps://www.youtube.com/watch?v=UQQHSbeIaB0&ab_channel=ritvikmath
2. Eric (Apr 15, 2021) Introduction to the Fundamentals of Vector Autoregressive Models
    - https://www.aptech.com/blog/introduction-to-the-fundamentals-of-vector-autoregressive-models/
3. Ritvikmath (Oct 7, 2020) Time Series Model Selection (AIC & BIC) : Time Series Talk
    - https://www.youtube.com/watch?v=McEN54l3EPU&ab_channel=ritvikmath
4. Hanck C., Arnold M., Gerber A., Schmelzer, M. (Oct 6, 2021) 4.4 The Least Squares Assumptions , Introduction to Econometrics with R
    - https://www.econometrics-with-r.org/4-4-tlsa.html