VAR model

Lucas Dupont and Mathieu Marauri

Contents

1 Introduction

In this document we will use the following notations:

- X_t will stand for a time serie. If a second time serie is needed we will use Y_t.
- B stands for the lag or back-shift operator. $B(X_t) = X_{t-1}$.

The necessary steps to perform a VAR model will be presented in this document.

Several hypotheses on the series must be verified to compute a VAR model:

- they must be stationary. The definition is given in Annex A: the weak stationarity.
- they must not be co-integrated

2 The stationarity

Every time series incorporated in the model must be stationary. Before testing the series we must know if there is a trend or a season. Graphically we can detect a trend or even a season. Plotting the serie is therefore a good start.

2.1 Test for a linear trend

The non-parametric Mann-Kendall trend test is used to detect a linear trend. The null hypothesis is no linear trend whereas the alternative hypothesis is a linear trend in the serie. The function used to perform this test is the 'MannKendall' function from the 'Kendall' package.

This test is described in Annex B: Mann-Kendall trend test.

If a trend is detected we have to de-trend the serie. To do so we can differentiate the time serie: $(1 - B)X_t$.

2.2 How to detect a seasonal pattern

A good way to see if there is a trend, besides plotting the serie, is to graph the power spectrum (Annex C: power spectrum). If no pics can be observed then we can consider that there is no seasonal pattern. Otherwise prominent pikes indicates a season at the given period (12 for months for instance).

In R the function 'periodogram' from the 'TSA' package can be used.

To remove the seasonal pattern we an differentiate the serie using $(1 - B^s)$ where s stands for the period. This operator applied to a monthly seasonal serie X_t removes the season.

2.3 The staionarity itself

After having removed the trend and the seasonal pattern we have to check if the serie is stationary. To do so we use a unit root test: the augmented Dickey Fuller test. The null hypothesis is that the serie is non-stationary. The function we use is the 'adf.test' from the 'tseries' packages.

The test is described in Annex D: the augmented Dickey Fuller test.

In order to deal with stationary issues we have to differentiate the serie.

At the end the stationarity of a time serie is easily obtained by differentiating the time serie. This operation takes care of the trend, the seasonal pattern and the stationarity issue.

3 The co-integration

Two time series, X_t and Y_t, are said to be co-integrated if:

- X_t and Y_t are integrated of order one,
- there exist μ and β such that we have: $Y_t = \mu + \beta * X_t + u_t$ (1) where u_t are the residuals, and u_t are stationary.

The process to detect co-integration is the following, as defined by Engle-Granger:

Step 1: get the order of integration of our series. If they are of order one we can proceed to the next step. If one of them is stationary then they are not co-integrated.

Step 2: compute the regression defined above (1). Keep the residuals terms for the next step.

Step 3: perform an augmented Dickey-Fuller unit root test o the residuals. If they are stationary then there is co-integration. This means that the null hypothesis of a unit root is rejected. Else the series do not co-integrate.

The 'egcm' function from the 'egcm' package performs the Engle-Granger approach.

At this point we can safely perform a VAR model. To do so we use the 'VAR' function from the 'vars' package.

Nevertheless if the series co-integrate then a VAR model is not appropriate and a VAR-ECM model should be used. This model is not detailed here.

4 The order of the model

When computing a VAR model one has to decide the optimal number of lag terms. The first thing to consider is the number of observations and the number of variables (2 in our case). In order for the parameters to be estimated it is necessary to have:

$$n(np+1) \leq nT \Leftrightarrow p \leq \frac{T-1}{n}$$

where n is the number of variables, T the number of periods and p the number of lags.

Once the number of lags is decided we can compare the obtained models to find the best one. To do so there exist several criteria: the FPE, the AIC, the BIC and the HQ. The model with the minimum value for one of these criteria is supposed to be the best. A diagnostic is still necessary to validate the model (see Diagnostic of a VAR model).

4.1 Final Predictor Error

The final predictor error criterion is given by:

$$FPE(p) = \left(\frac{T+np+1}{T-nm-1}\right)^n det(\Sigma_p)$$

Where Σ_p is the maximum likelihood estimator of the variance-covriance matrix of the errors.

The first part of the criterion increases with p whereas $det(\Sigma_p)$ decreases with p.

4.2 The Aikaike's Information Criterion

The Aikaike's information criterion is given by:

$$AIC(p) = log\left(det\left(\Sigma_p\right)\right) + \frac{2pn^2}{T}$$

Once again the two different terms have an opposite behaviour when p increases.

4.3 The Bayesian Information Criterion

5 Diagnostic of a VAR model

A VAR model has been computed, we have the following equations, for two time series:

$$\left\{\begin{array}{c} X_t = \alpha_1 + \sum_{i=1}^{k} \beta_{11,i} X_{t-i} + \sum_{i=1}^{k} \beta_{12,i} Y_{t-i} + \epsilon_{1,t} \\ Y_t = \alpha_1 + \sum_{i=1}^{k} \beta_{21,i} Y_{t-i} + \sum_{i=1}^{k} \beta_{22,i} X_{t-i} + \epsilon_{2,t} \end{array}\right. \tag{1}$$

Where $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})$ must verify:

$$E(\epsilon_t \epsilon_{t-j}) = 0,\ \forall j\ in\ \mathbb{Z} \setminus \{0\}$$

In order to test:

$$\left\{\begin{array}{c} (H_0) : \epsilon_t\ are\ not\ autocorrelated \\ (H_1) : \epsilon_t\ are\ autocorrelated \end{array}\right. \tag{2}$$

We use the multivariate Ljung-Box test. It is further explained in Annex E: the Ljung-Box test.

We must not reject the null hypothesis.

If the null hypothesis is rejected it means that there is some information left in the error terms. Therefore the order of the VAR model must be augmented.

6 Conclusion

The process described in this document requires several hypotheses to be tested. The different tests used are described in the annexes. If some hypotheses are not checked then some other models can be computed. They are not explained here.

7 Annexes

In this section the different methods are more precisely explained.

7.1 Annex A: the weak stationarity

The stationary that need to be verified is the weak stationary. A serie X_t is stationary if:

- $E(X_t^2) < \infty, \forall t$,
- $E(X_t) = M$,
- $\forall t,\ \forall h,\ cov(X_t, X_{t+h}) = \gamma(h)$

7.2 Annex B: Mann-Kendall trend test

The Mann-Kendall (MK) test statistically assesses if there is a monotonic trend in the time serie. This trend may or may not be linear. It is a non-parametric test.

The hypotheses are:

$$\begin{cases} (H_0) : no\ monotonic\ trend \\ (H_1) : monotonic\ trend \end{cases} \tag{3}$$

We note $x_1, x_2, ..., x_T$ the observations.

We compute $sign(x_j - x_k)\ \forall\ j > k$ where

$$sign(x_j - x_k) = \begin{cases} 1 \ \ if\ x_j - x_k > 0 \\ 0 \ \ if\ x_j - x_k = 0 \\ -1 \ \ if\ x_j - x_k < 0 \end{cases} \tag{4}$$

Then we compute $S = \sum_{k=1}^{T-1} \sum_{j=k+1}^{T} sign(x_j - x_k)$

For the statistic we need to have the variance of S which is given by:

$$Var(S) = \frac{1}{18}[T(T-1)(2T+5) - \sum_{p=1}^{g} t_p(t_p - 1)(2t_p + 5)]$$

with:

- g = the number of tied groups,
- t_p the number of observations in the p^{th} tied group

Finally we have the Mann-Kendall test statistic Z_{MK}:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{Var(S)}},\ if\ S > 0 \\ 0,\ if\ S = 0 \\ \frac{S+1}{\sqrt{Var(S)}},\ if\ S < 0 \end{cases} \tag{5}$$

Under H_0 Z_{MK} follows a normal distribution. Therefore we reject the null hypothesis of no monotonic trend if we have:

$$|Z_{MK}| \geq Z_{1-\alpha/2}$$

Where $Z_{1-\alpha/2}$ is the quantile of order $1 - \alpha/2$ of a normal distribution.

7.3 Annex C: power spectrum

In order to detect a seasonal pattern one can use the graph of the power spectrum of the serie. The periodogram graphs a measure of the relative importance of possible frequency values that might explain the oscillation pattern of the observed data.

We note $x_1, x_2, ..., x_T$ the observations and $\omega_j = \frac{j}{T}\ for\ j = 1, 2, ..., T/2$. ω_j are the possible frequencies.

The serie x_t is represented as follow:

$$x_t = \sum_{j=1}^{T/2} \left[\beta_1 \left(\frac{j}{T} \right) \cos(2\pi\omega_j t) + \beta_2 \left(\frac{j}{T} \right) \sin(2\pi\omega_j t) \right].$$

We can see $\beta_1 \left(\frac{j}{T}\right)$ and $\beta_2 \left(\frac{j}{T}\right)$ as regression parameters. Since j goes from 1 to T/2 there are T parameters. To estimate the parameters we use the Fast Fourier Transform. Once the parameters have been estimated we define:

$$P\left(\frac{j}{T}\right) = \hat{\beta}_1^2 \left(\frac{j}{T}\right) + \hat{\beta}_2^2 \left(\frac{j}{T}\right)$$

This is the value of the periodogram at frequency j/T.

7.4 Annex D: the augmented Dickey Fuller test.

The augmented Dickey-Fuller tests:

$$\begin{cases} (H_0) : \rho = 1 \\ (H_1) : \rho \neq 1 \end{cases} \tag{6}$$

in the following equation:

$$X_t = \alpha + \beta t + \rho X_{t-1} + \sum_{i=1}^{k} \theta_i \Delta X_{t-i} + \epsilon_t$$

The parameter k is an argument of the 'adf.test' function.

The test statistic is:

$$\Phi = \frac{(SCR_0 - SCR_1)/2}{SCR_1/(n-3)}$$

Where SCR_0 stands for the sum of squared residuals under the null hypothesis and SCR_1 under the alternative hypothesis.

This test statistic is to be compared with the quantile of order α of the Dickey-Fuller distribution.

If the null hypothesis is rejected, that is to say if $\rho \neq 1$ then we can say that there is no unit root in the serie and therefore that the serie is stationary.

7.5 Annex E: the Ljung-Box test

This test evaluates the auto-correlation of the error terms $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})$.

We note:

$$C_{0j}(h) = T^{-1} \sum_{t=j+1}^{T} \epsilon_t \epsilon'_{t-j}$$

The test statistic is:

$$Q_{LB}(s) = T(T+2) \sum_{j=1}^{s} \frac{tr(C_{0j} C_{00}^{-1} C'_{0j} C_{00}^{-1})}{T-j}$$

Where s stands for the order of auto-correlation.

Under (H_0) $Q_{LB}(s)$ follows a χ^2 distribution with $k^2(s-L)$ degrees of freedom. L is the the order of the VAR model, k the number of series used in the model (2 in our case). The null hypothesis is rejected if $Q_{LB}(s) > \chi^2_{1-\alpha}(k^2(s-L))$ where $\chi^2_{1-\alpha}(k^2(s-L))$ is the quantile of order $1-\alpha$ of a χ^2 distribution with $k^2(s-L)$ degrees of freedom.

In R the function 'LjungBox' in the 'portes' package can be used.