Vector autoregressive models

Kevin Kotzé

Contents

Introduction

- VAR models are widely used in time series research:
  - Examine the dynamic relationships that exist between variables
  - Important forecasting tools that are used by economic & policy-making institutions

- Most of the concepts in this lecture are multivariate extensions of the tools and concepts that apply to autoregressive models

- This lecture introduces some of the key ideas and methods used in VAR analysis, where we discuss:
  - stability properties and moving average representation
  - issues related to specification, estimation and forecasting

- Granger causality

Notation

- To describe the use of multivariate techniques, we need to introduce new notation:
    - Small letters denote a $(K \times 1)$ vector of random variables, where

$$\boldsymbol{y}_t = \left(y_{1,t}, \ldots, y_{K,t}\right)'$$

- The VAR model of order p can then be written as,

$$\boldsymbol{y}_t = A_1 \boldsymbol{y}_{t-1} + \ldots + A_p \boldsymbol{y}_{t-p} + C D_t + \boldsymbol{u}_t$$

- where A_j is a $(K \times K)$ coefficient matrix, for $j = \{1, \ldots, p\}$
- C is the coefficient matrix for deterministic regressors
- D_t is the matrix for deterministic regressors
- $\boldsymbol{u}_t$ is a $(K \times 1)$ dimension vector of error terms

Notation

- The vector of error terms are assumed to be white noise

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{u}_t\right] &= 0 \\
\mathbb{E}\left[\boldsymbol{u}_t \boldsymbol{u}_t'\right] &= \Sigma_{\boldsymbol{u}} \text{ which is positive definite}
\end{aligned}
$$

- This VAR is termed a reduced-form representation, which differs to the structural VAR (SVAR) that is discussed later

- Model relates the k'th variable in the vector $\boldsymbol{y}_t$ to past values of itself and all other variables in the system

Basic VAR model

- For simplicity, assume $K = 2$, and $p = 1$,

$$\boldsymbol{y}_t = A_1 \boldsymbol{y}_{t-1} + \boldsymbol{u}_t$$

- where y_t, μ, A_1, and $\boldsymbol{u}_t$ are given as,

$$\boldsymbol{y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}, A_1 = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \text{ and } \boldsymbol{u}_t = \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

- For example, assume the elements of A_1 are given as,

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 1 & 0.2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

- where after some matrix manipulations,

$$\begin{aligned} y_{1,t} &= 0.5 y_{1,t-1} + u_{1,t} \\ y_{2,t} &= 1 y_{1,t-1} + 0.2 y_{2,t-1} + u_{2,t} \end{aligned}$$

Basic VAR model

- The above model suggests:
  - $y_{2,t}$ depends on past values of itself and past values of $y_{1,t}$
  - $y_{1,t}$ only depends on past values of itself
- The variables that are to be included will typically depend on the purpose of the study
- Usually include variables that may have various dynamic interactions or a perceived causal relationship

The companion form

- Useful to express the $VAR(p)$ as a $VAR(1)$ in the companion form,

$$Z_t = \Gamma_0 + \Gamma_1 Z_{t-1} + \Upsilon_t$$

- where we have,

$$Z_t = \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p+1} \end{bmatrix}, \quad \Gamma_0 = \begin{bmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \Upsilon_t = \begin{bmatrix} \boldsymbol{u}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The companion form

- So that the matrix notation is

$$\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-p+1} \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \boldsymbol{y}_{t-3} \\ \vdots \\ \boldsymbol{y}_{t-p} \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- where the vectors Z_t, Γ_0 and Υ_t are $Kp \times 1$
- A_j for $j = 1, \ldots, p$ is $K \times K$, and
- Γ_1 is $Kp \times Kp$
- In this case Γ_1 is called the companion-form matrix

Stability of VAR model

- The VAR is covariance-stationary when the effect of the shocks, $\boldsymbol{u}_t$, dissipate
- This occurs when the eigenvalues of the companion form matrix are all less than one in absolute value
- The eigenvalues of the matrix Γ_1 are represented by λ in the expression,

$$|\Gamma_1 - \lambda I| = 0$$

- To derive the eigenvalues in our bivariate $VAR(1)$ example,

$$\det\left[\begin{bmatrix} 0.5 & 0 \\ 1 & 0.2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right] = \det\left[\begin{bmatrix} 0.5-\lambda & 0 \\ 1 & 0.2-\lambda \end{bmatrix}\right]$$
$$(0.5-\lambda)(0.2-\lambda) = 0$$

- Hence,

$$\lambda_1 = 0.5, \quad \lambda_2 = 0.2$$

Stability of VAR model

- Certain researchers consider the values of the *characteristic roots*, which may be defined as z in the expression

$$\left| I - \Gamma_1 z \right|$$

- where the interpretation is reversed, as a stable stochastic process has characteristic roots that lie outside the unit circle
- The interested reader may wish to consult Hamilton (1994)

Simulating stable VAR processes

- We can simulate the above bivariate $VAR(1)$ with $y_{k,0} = 0$, $\mu_k = 1$ for $k = [1, 2]$ and

$$\boldsymbol{u}_t \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \right)$$

- Note that the processes fluctuate around a constant mean & their variability does not appear to change with time

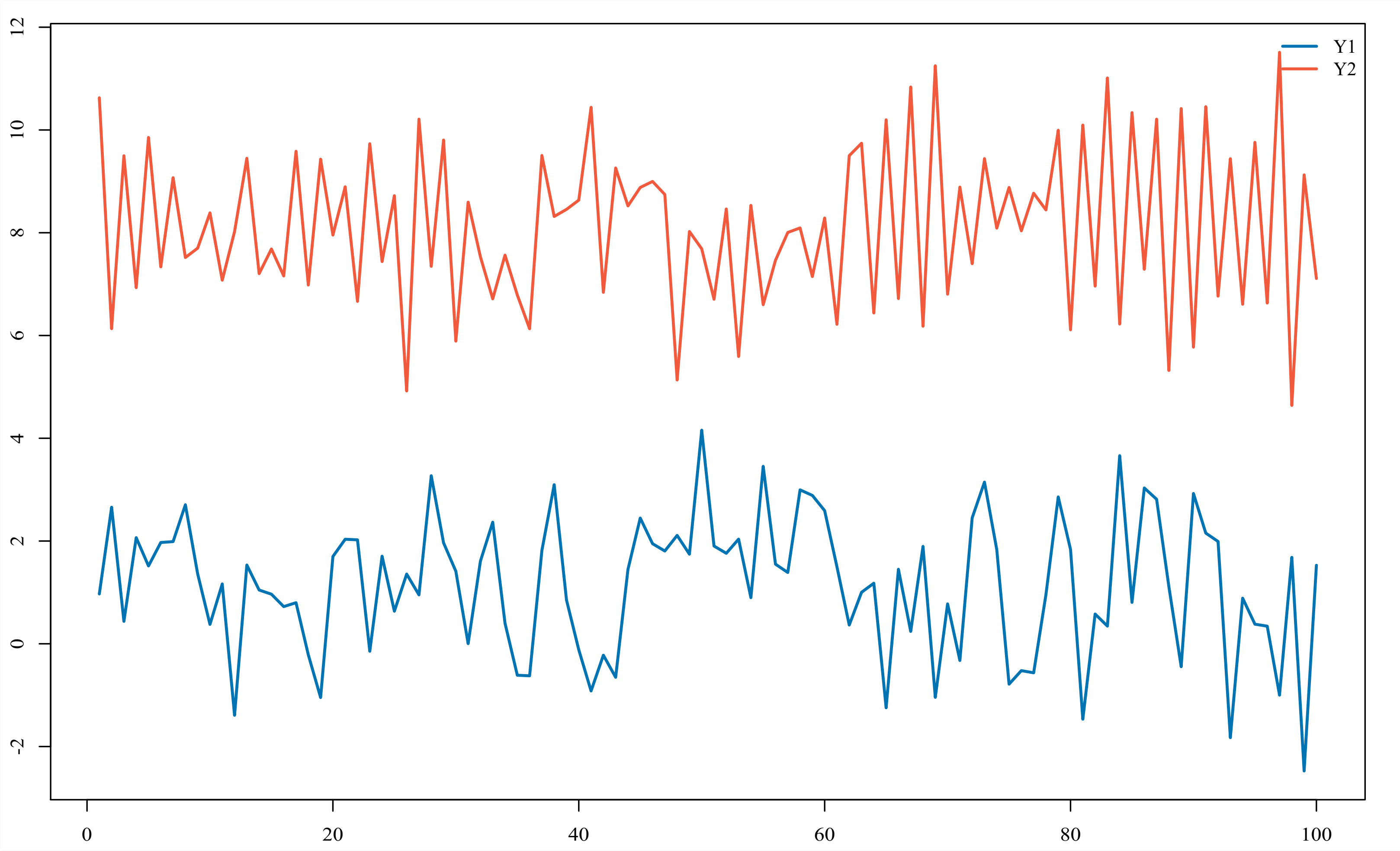

Figure : Simulated VAR processes

Wold representation

- Just as the stable $AR(p)$ model has a MA representation, the stable $VAR(p)$ has a VMA representation - termed the Wold decomposition
- Theorem states that every covariance-stationary time series can be written as the sum of two uncorrelated processes:
  - deterministic component, κ_t, (which could be the mean)
  - infinite moving average representation of $\sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$
- Hence,

$$y_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + \kappa_t$$

- where we assume $\theta_0 = 1$
- $\sum_{j=0}^{\infty} |\theta_j| < \infty$
- ε_t is white noise

Wold representation

- This would involve fitting an infinite number of parameters $(\theta_1, \theta_2, \theta_3, \ldots)$ to the data
- With a finite number of observations, this is not possible
- One can approximate $\theta(L)$ by using models that have a finite number of parameters
- Since we can write a $VAR(p)$ as a $VAR(1)$ model using the companion form, consider the example,

$$\boldsymbol{y}_t = \mu + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{u}_t$$

- Using the lag operator,

$$(\boldsymbol{I} - \boldsymbol{A}_1 L)\boldsymbol{y}_t = \mu + \boldsymbol{u}_t$$

Wold representation

- Using the expression, $(I - A_1 L) = A(L)$ we can write,

$$A(L)\boldsymbol{y}_t = \mu + \boldsymbol{u}_t$$

- Multiplying with $A(L)^{-1}$ we get the VMA representation,

$$\begin{aligned}\boldsymbol{y}_t &= A(L)^{-1}\mu + A(L)^{-1}\boldsymbol{u}_t \\ &= B(L)\mu + B(L)\boldsymbol{u}_t \\ &= \varphi + \sum_{j=0}^{\infty} B_j \boldsymbol{u}_{t-j}\end{aligned}$$

Wold representation

- Where we have used the geometric rule

$$A(L)^{-1} = (I - A_1 L)^{-1} = \sum_{j=0}^{p} A_1^j L^j \equiv B(L) = \sum_{j=0}^{\infty} B_j L^j$$

- with $B_0 = I$ and $\varphi = \left(\sum_{j=0}^{\infty} B_j \right) \mu$

Finding the MA coefficients

- The MA coefficients, B_j, are derived from the relationship $I = B(L)A(L)$
- Since, $A(L)^{-1}A(L) = I$ and $A(L)^{-1} = B(L)$. Therefore,

$$
\begin{aligned}
I &= B(L)A(L) \\
I &= (B_0 + B_1 L + B_2 L^2 + \ldots)(I - A_1 L - A_2 L^2 - \ldots - A_p L^p) \\
&= [B_0 + B_1 L + B_2 L^2 + \ldots] \\
&\quad -[B_0 A_1 L + B_1 A_1 L^2 + B_2 A_1 L^3 + \ldots] \\
&\quad -[B_0 A_2 L^2 + B_1 A_2 L^3 + B_2 A_2 L^4 + \ldots] - \ldots \\
&= B_0 + (B_1 - B_0 A_1)L + (B_2 - B_1 A_1 - B_0 A_2)L^2 + \ldots \\
&\quad + \left(B_p - \sum_{j=1}^{i} B_{p-j} A_j\right) L^p + \ldots
\end{aligned}
$$

Finding the MA coefficients

- Solving for the relevant lags (noting that $A_1 = 0$ for $j > p$), we get,

$$\begin{aligned}
B_0 &= I \\
B_1 &= B_0 A_1 \\
B_2 &= B_1 A_1 + B_0 A_2 \\
\vdots &\quad \vdots \\
B_i &= \sum_{j=1}^{i} B_{i-j} A_j \qquad \text{for } i = 1, 2, \ldots
\end{aligned}$$

- Hence, the B_j parameters can be computed recursively

Mean, variance & autocovariance

- The first two moments of the VAR can be derived from the MA representation

$$\boldsymbol{y}_t = \varphi + \sum_{j=0}^{\infty} B_j \boldsymbol{u}_{t-j}$$

- where $\varphi = \left(\sum_{j=0}^{\infty} B_j\right)\mu = (I - A_1)^{-1}\mu$

- Since the error terms are assumed to be Gaussian white noise the expected mean value is,

$$\mathbb{E}[\boldsymbol{y}_t] = \varphi = (I - A_1)^{-1}\mu$$

- This mean may be termed the steady-state of the system

Mean, variance & autocovariance

- The covariance and autocovariances, denoted Ψ, may then be derived with the aid of Yule-Walker equations, where we write the process in the mean-adjusted form

$$\boldsymbol{y}_t - \varphi = A_1(\boldsymbol{y}_{t-1} - \varphi) + \boldsymbol{u}_t$$

- Postmultiplying by $(\boldsymbol{y}_{t-s} - \varphi)'$ and taking expectation gives,

$$\begin{aligned}\mathbb{E}\left[(\boldsymbol{y}_t - \varphi)\left(\boldsymbol{y}_{t-s} - \varphi\right)'\right] = A_1 \mathbb{E}\left[(\boldsymbol{y}_{t-1} - \varphi)\left(\boldsymbol{y}_{t-s} - \varphi\right)'\right] \\ + \mathbb{E}\left[\boldsymbol{u}_t(\boldsymbol{y}_{t-s} - \varphi)'\right]\end{aligned}$$

Mean, variance & autocovariance

- Thus for $\boldsymbol{s} = 0$,

$$\boldsymbol{\Psi}_{\boldsymbol{s}} = A_1 \boldsymbol{\Psi}_{-1} + \Sigma_{\boldsymbol{u}} = A_1 \boldsymbol{\Psi}_1' + \Sigma_{\boldsymbol{u}}$$

- where after the second equality sign, we used the fact that $\boldsymbol{\Psi}_{-1} = \boldsymbol{\Psi}_1'$
- Hence, for $\boldsymbol{s} > 0$, we have

$$\boldsymbol{\Psi}_{\boldsymbol{s}} = A_1 \boldsymbol{\Psi}_{\boldsymbol{s}-1}$$

- when A_1 and $\Sigma_{\boldsymbol{u}}$ are known, we can compute the autocovariance for $\boldsymbol{s} = \{0, \ldots, S\}$ using the above two expressions

Mean, variance & autocovariance

- Hence, for $s = 1$, we have $\Psi_1 = A_1 \Psi_0$
- Substituting into the expression for Ψ_0 and noting that $[A_1 \Psi_0]' = \Psi_0' A_1'$
- Using the rules of matrix algebra, we get,

$$\Psi_0 = A_1 \Psi_0 A_1' + \Sigma_{\boldsymbol{u}}$$

- Solve for Ψ_0 with the Kronecker product and the vec operator to get [see Lutkepohl (2005)]

$$vec\Psi_0 = (I - A_1 \otimes A_1)^{-1} vec\Sigma_u$$

- Once Ψ_0 has been derived we can derive the autocovariances for $s > 0$ with recursive substitution

Mean, variance & autocovariance

- Then, to get the autocorrelation function, we need to normalize the autocovariances so that they have ones on the diagonal at $s = 0$
- Thus, we define the diagonal matrix ϑ, whose diagonal elements are the square roots of the diagonal elements of Ψ_0
- Then the autocorrelation function for the VAR is simply,

$$R_s = \vartheta^{-1} \Psi_h \vartheta^{-1}$$

- Note that while we only considered the case of a $VAR(1)$, we could have expressed a $VAR(p)$ as a $VAR(1)$ using the companion form
- In this case one would need to specify a selection matrix to extract the values of interest [see Lutkepohl (2005)]

Estimating VAR parameters

- VAR system can be estimated equation-by-equation using OLS
- Would be consistent, and with normality of errors, efficient
- Assuming that we have a sample size of T, with $\{y_1, \ldots, y_T\}$, for each of the K variables
- Can be shown that the estimator has the same efficiency as the generalized LS (GLS) estimator
- Following Lutkepohl (2005), we define $Y = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$, $A = [A_1, \ldots, A_p]$, $U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T]$ and $Z = [Z_0, \ldots, Z_{T-1}]$, where,

$$Z_{t-1} = \begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-p} \end{bmatrix}$$

Estimating VAR parameters

- The VAR model can then be written as,

$$Y = AZ + U$$

- And the OLS estimator of A is,

$$\hat{A} = [\hat{A}_1, \ldots, \hat{A}_p] = YZ'(ZZ')^{-1}$$

- This OLS estimator is consistent and asymptotically normally distributed,

$$\sqrt{T} vec(\hat{A} - A) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_e)$$

- where $\xrightarrow{d}$ implies convergence in distribution
- vec denotes the column stacking operator and $ZZ'/T \xrightarrow{d} \Gamma$

Choice of variable and lags

- When deciding on variables and lags, note that these models quickly become heavily parameterised, which could result in d.o.f. problems
- For example, with 6 variables and 8 lags, each equation will contain $(6 \times 8) + 1 = 49$ coefficients (including a constant)
- Hence, with limited observations the parameter estimates may be imprecise
- Choice of variables usually determined by economic theory or *a prior* ideas
- Lags can be determined by information criterion (BIC or AIC)

VAR forecasts

- VAR models are popular tools for forecasting variables
- VAR forecasts are derived as per the AR forecasts
- The observed h-step value of the $\boldsymbol{y}_t$ processes would be,

$$\boldsymbol{y}_{t+h} = A_1^h \boldsymbol{y}_t + \sum_{i=0}^{h-1} A_1^i \boldsymbol{u}_{t+h-i}$$

- Adding a constant, results in the following:

$$\boldsymbol{y}_{t+h} = (I + A_1 + \ldots A_1^{h-1})\mu + A_1^h \boldsymbol{y}_t + \sum_{i=0}^{h-1} A_1^i \boldsymbol{u}_{t+h-i}$$

VAR forecasts

- By employing the conditional expectation, we get the VAR point forecast,

$$\mathbb{E}\left[\boldsymbol{y}_{t+h} | \boldsymbol{y}_t\right] = (I + A_1 + \ldots + A_1^{h-1})\mu + A_1^h \boldsymbol{y}_t$$

- under the assumption of stationarity, this converges to the unconditional mean of the process when $h \to \infty$,

$$\mathbb{E}\left[\boldsymbol{y}_{t+h} | \boldsymbol{y}_t\right] = \frac{\mu}{I - A_1} \qquad \text{when } h \to \infty$$

- Again, these equations are just the multivariate extensions of the previous formulas used for the AR models

Mean Squared Forecasting Error

- Where $\mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t\right]$ is the predictor the VAR forecast error at horizon h is,

$$\boldsymbol{y}_{t+h} - \mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t\right] = \sum_{i=0}^{h-1} A_1^i \boldsymbol{u}_{t+h-i}$$

- The expectation of this provides the expected forecast error
- With the assumption that $\mathbb{E}[\boldsymbol{u}_t] = 0$

$$\mathbb{E}\left[\boldsymbol{y}_{t+h} - \mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t\right]\right] = \mathbb{E}\left[\boldsymbol{y}_{t+h}\right] - \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t\right]\right] = 0$$

- Thus the predictor $\mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t\right]$ is unbiased and the MSFE is simply the forecast error variance

Mean Squared Forecasting Error

- In the multivariate VAR setting the MSFE is,

$$\begin{aligned}\boldsymbol{\sigma}^f_{t+h} &= \mathbb{E}\left[\left(\boldsymbol{y}_{t+h} - \mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t\right]\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{i=0}^{h-1} A_1^i \boldsymbol{u}_{t+h-i}\right)\left(\sum_{i=0}^{h-1} A_1^i \boldsymbol{u}_{t+h-i}\right)\right]\end{aligned}$$

- where we can move the A_1 terms outside of the expectation and $\mathbb{E}(\boldsymbol{u}_{t+h-i}, \boldsymbol{u}_{t+h-i}) = \Sigma_{\boldsymbol{u}}$ for all h

- Hence,

$$\boldsymbol{\sigma}^f_{t+h} = \sum_{i=0}^{h-1} \left(A_1^j \Sigma_{\boldsymbol{u}} A_1^{j'}\right)$$

Uncertainty

- Assuming the errors of the VAR model are Gaussian, $\boldsymbol{u}_t \sim \text{i. i. d.}\mathcal{N}(0, \Sigma_{\boldsymbol{u}})$, and independent across time

- The forecast errors are normally distributed,

$$\frac{\boldsymbol{y}_{k,t+h} - \mathbb{E}\left[\boldsymbol{y}_{t+h} \middle| \boldsymbol{y}_{k,t}\right]}{\boldsymbol{\sigma}^f_{k,t+h}} \sim \mathcal{N}(0,1)$$

- where $\boldsymbol{y}_{k,t+h}$ and $\mathbb{E}\left[\boldsymbol{y}_{t+h} \middle| \boldsymbol{y}_{k,t}\right]$ are the k'th elements of the actual and predictor values

- and $\boldsymbol{\sigma}^f_{k,t+h}$ is the square root of the variance for the k'th equation (i.e. the square root of the k'th element on the diagonal on the $\boldsymbol{\sigma}^f_{t+h}$ matrix)

Uncertainty

- Forecast intervals can then be generated around the VAR point forecasts using,

$$\left[\mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{y}_{k,t}\right] - z_{\alpha/2}\, \boldsymbol{\sigma}^{f}_{k,t+h}\ ,\ \mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{y}_{k,t}\right] + z_{\alpha/2}\, \boldsymbol{\sigma}^{f}_{k,t+h}\right]$$

- which is equivalent to what was derived in the previous discussion on forecasts

Forecast failure in macroeconomics

- Since Sims (1980), examples of VARs that are used to forecast key economic variables such as output, prices, and the interest rates have been numerous
- However, some recent work suggests that VAR models may be prone to instabilities
- To improve the accuracy of forecasts with a VAR researchers use intercept correction, time varying-parameters, differencing data (that may help with mean shifts), model averaging, endogenous structural-breaks, etc.
- See Clement and Hendry (2011), Clark and McCracken (2008), Allen and Fildes (2005), and others for interesting discussions

Granger causality

- The idea is that a cause must precede the effect
- Hence, if variable $y_{2,t}$ Granger-causes behaviour in variable $y_{1,t}$, then $y_{2,t}$ should improve upon the predictions of $y_{1,t}$
- In the previous example,

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 \\ 1 & 0.2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

- Note, $y_{2,t}$ doesnt influence future values of $y_{1,t}$
- However, $y_{1,t}$ does influence future values of $y_{2,t}$

Granger causality

- The test for Granger causality is formally implemented by means of a joint hypothesis test
- For each equation in the VAR we compute $K-1$ restricted versions of the VAR model, which is compared with the unrestricted version
- Considers whether all the lags of the k'th variable in the system are jointly significantly different from zero
- This is simply a standard F-test, where the null hypothesis is no Granger causality

Granger causality

- In a $VAR(2)$ model, with $K = 3$, the first equation is,

$$\begin{aligned} y_{1,t} = \mu_1 + \alpha_{11} y_{1,t-1} + \alpha_{12} y_{2,t-1} + \alpha_{13} y_{3,t-1} + \ldots \\ \alpha_{14} y_{1,t-2} + \alpha_{15} y_{2,t-2} + \alpha_{16} y_{3,t-2} + e_{1,t} \end{aligned}$$

- Test for no Granger causality from $y_{2,t}$ to $y_{1,t}$ would be an F-test with null, $\alpha_{12} = \alpha_{15} = 0$
- Test for no Granger causality from $y_{3,t}$ to $y_{1,t}$ would be an F-test with null, $\alpha_{13} = \alpha_{16} = 0$
- Conducting similar tests on other equations of the VAR would give a complete test for no Granger causality
- Rejection of any of the null hypothesis indicates Granger causality
- Does not say anything about true causality, only used to infer a predictive relationship

Summary

- VAR is a multivariate version of the univariate AR
- $VAR(p)$ can be written as a $VAR(1)$ model by writing it in the companion form
- If all eigenvalues of the companion form matrix are less than 1 in absolute value, the VAR is stable
- Stable $VAR(p)$ model can be inverted and written as an infinite order vector moving average model
- VAR can be estimated by OLS equation-by-equation. Under standard assumptions, the OLS estimator will be similar to the maximum likelihood estimator of the whole system

Summary

- Forecasting with a stable VAR will converge towards the unconditional mean of the model, and the MSFE matrix of the forecast errors can be derived with relative ease
- Density forecasts and forecast intervals can be constructed based in a similar way to the AR models
- Granger causality tests involve testing whether or not lagged values of a given variable in the VAR system help predict one of the other endogenous variables in the system. Such a test can simply be conducted using a F-test, where the null hypothesis is no Granger causality