Time Series Analysis - Multivariate Stationary Models

ECON 722

Elena Pesavento — Emory University

Fall 2024

Introduction

- A multivariate time series $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{mt})'$ is an $m \times 1$ vector process observed in sequence over time, $t = 1, \ldots, n$.
- Multivariate time series models primarily focus on the joint modeling of the vector series $\boldsymbol{y}_t$. The most common multivariate time series models used by economists are vector autoregressions (VARs). VARs were introduced to econometrics by Sims (1980).
- Some excellent textbooks and review articles on multivariate time series include Hamilton (1994), Watson (1994), Canova (1995), Lütkepohl (2005), Ramey (2016), Stock and Watson (2006), and Kilian and Lütkepohl (2017).

Multiple Equation Time Series Models

To motivate vector autoregressions let us start by reviewing the autoregressive distributed lag model for the case of two series: $\boldsymbol{y}_t = (y_{1t}, y_{2t})'$ with a single lag.For y_{1t} :

$$y_{1t} = \alpha_0 + \alpha_1 y_{1t-1} + \beta_1 y_{2t-1} + e_{1t}.$$

For y_{2t} :

$$y_{2t} = \gamma_0 + \gamma_1 y_{2t-1} + \delta_1 y_{1t-1} + e_{2t}.$$

stacking the two equations together, and writing the vector error as $\boldsymbol{e}_t = (e_{1t}, e_{2t})'$ we can write

$$y_t = a_0 + A_1 y_{t-1} + e_t$$

where $\boldsymbol{a}_0$ is 2×1 and $\boldsymbol{A}_1$ is 2×2. This is a bivariate vector autoregressive model for $\boldsymbol{y}_t$.

VAR

In general for p lags we can writethe p^{th} order vector autoregressive (VAR(p)) model

$$y_t = a_0 + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t$$

$\boldsymbol{y}_t$ to be a vector of dimension m,$\boldsymbol{A}_\ell$ are $m \times m$ and $\boldsymbol{e}_t$ is $m \times 1$.

$$\boldsymbol{A}_\ell = \begin{bmatrix} a_{11,\ell} & a_{12,\ell} & \cdots & a_{1m,\ell} \\ a_{21,\ell} & a_{22,\ell} & \cdots & a_{2m,\ell} \\ \vdots & \vdots & & \vdots \\ a_{m1,\ell} & a_{m2,\ell} & \cdots & a_{mm,\ell} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = \mathbb{E}\left(\boldsymbol{e}_t \boldsymbol{e}_t'\right)$$

is non diagonal.

VAR as a linear projection

Theorem

If $\boldsymbol{y}_t$ is covariance stationary it has the projection equation

$$\boldsymbol{y}_t = \boldsymbol{a}_0 + \sum_{\ell=1}^{\infty} \boldsymbol{A}_\ell \boldsymbol{y}_{t-\ell} + \boldsymbol{e}_t.$$

The innovations $\boldsymbol{e}_t$ satisfy

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{e}_t\right) &= 0 & & \\
\mathbb{E}\left(\boldsymbol{y}_{t-\ell}\boldsymbol{e}_t'\right) &= 0 & \qquad \ell &\geq 1 \\
\mathbb{E}\left(\boldsymbol{e}_{t-\ell}\boldsymbol{e}_t'\right) &= 0 & \qquad \ell &\geq 1
\end{aligned}$$

and

$$\boldsymbol{\Sigma} = \mathbb{E}\left(\boldsymbol{e}_t\boldsymbol{e}_t'\right) < \infty.$$

If $\boldsymbol{y}_t$ is strictly stationary then $\boldsymbol{e}_t$ is strictly stationary.

Multivariate Wold Decomposition

We can write the model using the lag operator notation as

$$\mathrm{A}\ (\mathrm{L})\ y_t = a_0 + e_t$$

where

$$\boldsymbol{A}(z) = \boldsymbol{I}_m - \sum_{\ell=1}^{\infty} \boldsymbol{A}_\ell z^\ell.$$

The multivariate innovations $\boldsymbol{e}_t$ are mean zero and serially uncorrelated. This describes what is known as a multivariate white noise process.

Theorem - Multivariate Wold Decomposition

If y_t is covariance stationary and non-deterministic then it has the linear representation

$$y_t = \mu + \sum_{\ell=0}^{\infty} \boldsymbol{\Theta}_\ell e_{t-\ell}$$

where $\boldsymbol{e}_t$ are the white noise projection errors and $\boldsymbol{\Theta}_0 = \boldsymbol{I}_m$. The coefficient matrices $\boldsymbol{\Theta}_\ell$ are $m \times m$.

Vector MA Representation

We can write the moving average representation using the lag operator notation as

$$y_t = \boldsymbol{\mu} + \Theta(\mathrm{L})e_t$$

where

$$\boldsymbol{\Theta}(z) = \sum_{\ell=0}^{\infty} \boldsymbol{\Theta}_\ell z^\ell.$$

- If invertible, $\boldsymbol{\Theta}(z) = \boldsymbol{A}(z)^{-1}$ and $\boldsymbol{A}(z) = \boldsymbol{\Theta}(z)^{-1}$.
- Think in terms of MA matrices is easier for impulse response functions. You can calculate the moving average coefficient matrices $\boldsymbol{\Theta}_\ell$ from the projection coefficient matrices $\boldsymbol{A}_\ell$.
- While there is not closed-form solution there is a simple recursion by which the coefficients may be calculated. See Hansen's book.

VAR(1)

The first-order vector autoregressive process, denoted VAR(1), is

$$y_t = a_0 + A_1 y_{t-1} + e_t$$

where e_t is a strictly stationary and ergodic white noise process. Is it stationary?

Let $\lambda_{\max}(\boldsymbol{A})$ be the largest absolute eigenvalue of $\boldsymbol{A}$.

Theorem

If $\lambda_{\max}(\boldsymbol{A}_1) < 1$ then the VAR(1) process $\boldsymbol{y}_t$ is strictly stationary and ergodic.

VAR(p)

The $\mathbf{p}^{\text{th}}$ -order vector autoregressive process, denoted VAR($\mathbf{p}$), is

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + e_t$$

where e_t is a strictly stationary and ergodic white noise process. We can write the model using the lag operator notation as

$$\boldsymbol{A}(\mathrm{L})\boldsymbol{y}_t = a_0 + e_t$$

where

$$\boldsymbol{A}(z) = \boldsymbol{I}_m - \boldsymbol{A}_1 z - \cdots - \boldsymbol{A}_p z^p.$$

The condition for stationarity of the system can be expressed as a restriction on the roots of the determinantal equation.

Theorem

If all roots λ of $\det(\boldsymbol{A}(z)) = 0$ satisfy $|\lambda| > 1$ then the VAR(p) process $\boldsymbol{y}_t$ is strictly stationary and ergodic.

Regression Notation

Defining the $(mp+1) \times 1$ vector

$$x_t = \begin{pmatrix} 1 \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix}$$

and the $m \times (mp+1)$ matrix

$$\boldsymbol{A}' = \begin{pmatrix} a_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_p \end{pmatrix}.$$

Then the VAR system of equations can be written as

$$y_t = \boldsymbol{A}' x_t + e_t.$$

This is a multivariate regression model. The error has covariance matrix

$$\Sigma = \mathbb{E}\left(e_t e_t'\right)$$

Regression Notation

In general, if $\boldsymbol{y}_t$ is strictly stationary we can define the coefficient matrix $\boldsymbol{A}$ by linear projection.

$$\boldsymbol{A} = \left(\mathbb{E}\left(\boldsymbol{x}_t \boldsymbol{x}_t'\right)\right)^{-1} \mathbb{E}\left(\boldsymbol{x}_t \boldsymbol{y}_t'\right).$$

This holds whether or not $\boldsymbol{y}_t$ is actually a VAR(p) process. By the properties of projection errors

$$\mathbb{E}\left(x_t e_t'\right) = 0.$$

The projection coefficient matrix $\boldsymbol{A}$ is identified if $\mathbb{E}\left(\boldsymbol{x}_t \boldsymbol{x}_t'\right)$ is invertible.

Estimation

The systems estimator of a multivariate regression is least squares. The estimator can be written as

$$\widehat{\boldsymbol{A}} = \left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t' \right)^{-1} \left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{y}_t' \right).$$

Alternatively, the coefficient estimator for the j^{th} equation is (OLS equation by equation)

$$\widehat{\boldsymbol{a}}_j = \left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t' \right)^{-1} \left(\sum_{t=1}^{n} x_t y_{jt} \right).$$

The least squares residual vector is

$$\widehat{e}_t = y_t - \widehat{A}' x_t.$$

The estimator of the variance matrix is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{t=1}^{n} \widehat{\boldsymbol{e}}_t \hat{\boldsymbol{e}}_t'$$

Estimation

If $\boldsymbol{y}_t$ is strictly stationary and ergodic with finite variances then by the Ergodic Theorem

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{y}_t' \xrightarrow{p} \mathbb{E}\left(\boldsymbol{x}_t \boldsymbol{y}_t'\right)$$

and

$$\sum_{t=1}^{n} x_t x_t' \xrightarrow{p} \mathbb{E}\left(x_t x_t'\right).$$

Since the latter is positive definite, we conclude that $\widehat{\boldsymbol{A}}$ is consistent for $\boldsymbol{A}$. Standard manipulations show that $\widehat{\boldsymbol{\Sigma}}$ is consistent as well.

Asymptotic Distribution

Set

$$\boldsymbol{a} = \mathrm{vec}(\boldsymbol{A}) = \begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_m \end{pmatrix}, \quad \widehat{\boldsymbol{a}} = \mathrm{vec}(\widehat{\boldsymbol{A}}) = \begin{pmatrix} \widehat{\boldsymbol{a}}_1 \\ \vdots \\ \widehat{\boldsymbol{a}}_m \end{pmatrix}.$$

Theorem- MDS

If $\boldsymbol{y}_t$ follows the VAR(p) model with $\mathbb{E}\left(\boldsymbol{e}_t \mid \mathcal{F}_{t-1}\right) = 0$ $\mathbb{E}\left\|\boldsymbol{y}_t\right\|^4 < \infty$, and $\boldsymbol{\Sigma} > 0$, then as $n \to \infty$

$$\sqrt{n}(\widehat{\boldsymbol{a}} - \boldsymbol{a}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{V})$$

where

$$\begin{aligned} \boldsymbol{V} &= \overline{\boldsymbol{Q}}^{-1} \boldsymbol{\Omega} \overline{\boldsymbol{Q}}^{-1} \\ \overline{\boldsymbol{Q}} &= \boldsymbol{I}_m \otimes \boldsymbol{Q} \\ \boldsymbol{Q} &= \mathbb{E}\left(\boldsymbol{x}_t \boldsymbol{x}_t'\right) \\ \boldsymbol{\Omega} &= \mathbb{E}\left(\boldsymbol{e}_t \boldsymbol{e}_t' \otimes \boldsymbol{x}_t \boldsymbol{x}_t'\right). \end{aligned}$$

Asymptotic Distribution

- Notice that we added the stronger assumption that the innovation is a martingale difference sequence $\mathbb{E}\left(\boldsymbol{e}_t \mid \mathcal{F}_{t-1}\right) = 0$.
- This means that this distributional result assumes that the VAR(p) model is the correct conditional mean for each variable correct lags and there is no omitted nonlinearity.

If we further strengthen the MDS assumption to conditional homoskedasticity

$$\mathbb{E}\left(e_t e_t' \mid \mathcal{F}_{t-1}\right) = \boldsymbol{\Sigma}$$

then the asymptotic variance simplifies as

$$\begin{aligned}\boldsymbol{\Omega} &= \boldsymbol{\Sigma} \otimes \boldsymbol{Q} \\ \boldsymbol{V} &= \boldsymbol{\Sigma} \otimes \boldsymbol{Q}^{-1}\end{aligned}$$

Asymptotic Distribution

- If the VAR)p) is a an approximation, the asymptotic distribution can be derived under mixing conditions.
- In this case we do not require that the true process is a VAR: the coefficients are defined as those which produce the best (mean square) approximation.
- The only requirements on the true process are general dependence conditions.
- The distributional result shows that the coefficient estimators are asymptotically normal, with a covariance matrix which takes a "long-run" sandwich form.

Asymptotic Distribution

Theorem- Mixing

If $\boldsymbol{y}_t$ is strictly stationary, ergodic, $\boldsymbol{\Sigma} > 0$, and for some $r > 4, \mathbb{E}\left\|\boldsymbol{y}_t\right\|^r < \infty$ and the mixing coefficients satisfy $\sum_{\ell=1}^{\infty} \alpha(\ell)^{1-4/r} < \infty$, then as $n \to \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{a}} - \boldsymbol{a}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{V})$$

where

$$\boldsymbol{V} = \overline{\boldsymbol{Q}}^{-1} \boldsymbol{\Omega} \overline{\boldsymbol{Q}}^{-1}$$

$$\overline{\boldsymbol{Q}} = \boldsymbol{I}_m \otimes \boldsymbol{Q}$$

$$\boldsymbol{Q} = \mathbb{E}\left(\boldsymbol{x}_t \boldsymbol{x}_t'\right)$$

$$\boldsymbol{\Omega} = \sum_{\ell=-\infty}^{\infty} \mathbb{E}\left(\boldsymbol{e}_{t-\ell} \boldsymbol{e}_t' \otimes \boldsymbol{x}_{t-\ell} \boldsymbol{x}_t'\right).$$

Selection of Lags

For a data-dependent rule to pick the lag length p in a VAR it is recommended to minimize an information criterion. The formula for the AIC and BIC are

$$AIC(p) = n \log \det \widehat{\boldsymbol{\Omega}}(p) + 2K(p)$$

$$BIC(p) = n \log \det \widehat{\boldsymbol{\Omega}}(p) + \log(n)K(p)$$

$$\widehat{\boldsymbol{\Sigma}}(p) = \frac{1}{n}\sum_{t=1}^{n} \widehat{\boldsymbol{e}}_t(p)\widehat{\boldsymbol{e}}_t(p)'$$

$$K(p) = m(pm + 1)$$

where $K(p)$ is the number of parameters in the model, and $\widehat{\boldsymbol{e}}_t(p)$ is the OLS residual vector from the model with p lags. The log determinant is the criterion from the multivariate normal likelihood.

Impulse Response Functions

One of the most important concepts in applied multivariate time series is the impulse response function (IRF), which is defined as the change in $\boldsymbol{y}_t$ due to a change in an innovation or shock.

- There are various definitions of IRF. In linear model which definition you use does not matter.
- We will see later in the class that this matters a lot if the VAR is non linear.
- For right now, let's use the simplest definition
- We typically plot them as a function of h for each pair (i, j). The impulse response function $IRF_{ij}(h)$ is interpreted as how the i^{th} variable responds over time to the j^{th} innovation.

Impulse Response Functions

$$\mathrm{IRF}_{ij}(h) = \frac{\partial}{\partial e_{jt}} y_{it+h}.$$

There are m^2 such responses. Recall the Wold representation

$$y_t = \mu + \sum_{\ell=0}^{\infty} \mathbf{\Theta}_\ell e_{t-\ell}$$

We deduce that the impulse response matrix is

$$\mathrm{IRF}(h) = \mathbf{\Theta}_h$$

the h^{th} moving average coefficient matrix. The invididual impulse response is

$$\mathrm{IRF}\, F_{ij}(h) = \mathbf{\Theta}_{h,ij}$$

the ij^{th} element of $\mathbf{\Theta}_h$.

Impulse Response Functions

- In a linear vector autoregression, the impulse response function is symmetric in negative and positive innovations.
- The magnitude of the impact is linear in the magnitude of the innovation. Thus the impact of the innovation $e_{jt} = 2$ is $2IRF_{ij}(h)$ and the impact of the innovation $e_{jt} = -2$ is $-2IRF_{ij}(h)$.
- This means that the shape of the impulse response function is unaffected by the magnitude of the innovation
- The impulse response functions can be scaled as desired i.e. one unit of the impulse variable or one standard deviation size shock.

Cumulative Impulse Response Functions

$$\text{CIRF}(h) = \sum_{\ell=1}^{h} \frac{\partial}{\partial \varepsilon_t'} \boldsymbol{y}_{t+\ell}) = \sum_{\ell=1}^{h} \boldsymbol{\Theta}_\ell.$$

The limit of the cumulative impulse response as $h \to \infty$ is the long-run impulse response

$$\boldsymbol{C} = \lim_{h \to \infty} CIRF(h) = \sum_{\ell=1}^{\infty} \boldsymbol{\Theta}_\ell = \boldsymbol{\Theta}(1) = \boldsymbol{A}(1)^{-1}.$$

It is useful to observe that when a VAR is estimated on differenced observations $\Delta \boldsymbol{y}_t$ then cumulative impulse response is

$$\text{CIRF}(h) = \frac{\partial}{\partial \varepsilon_t'} \left(\sum_{\ell=1}^{h} \Delta \boldsymbol{y}_{t+\ell} \right) = \frac{\partial}{\partial \varepsilon_t'} \boldsymbol{y}_{t+h}$$

which is the impulse response function for the variable $\boldsymbol{y}_t$ in levels.

Estimation of Impulse Response Functions

- The impulse responses are determined by the VAR coefficients. We can write this mapping as $\boldsymbol{\Theta}_h = g_h(\boldsymbol{A})$. The plug-in approach suggests the estimator.
- Since $\widehat{\boldsymbol{A}}$ is random, so is $\widehat{IRF}(h)$.
- IRF is a non linear function of $\widehat{\boldsymbol{A}}$ so we just use the Delta Method to find the asymptotic distribution (a little messy).
- However, asymptotic approximation can be poor. Use Bootstrap instead.

Orthogonalized Impulse Response Functions

- Notice that the IRF as we have defined above is not very meaningful.
- The errors are correlated so when we shock one, the shock transmit.
- For economics interpretation, it is more useful to study the response to structural (orthogonal) shocks.
- We will discuss identification at length later on.

Local Projections

In some contexts (including prediction) it is useful to consider models where the dependent variable is dated multiple periods ahead of the right-hand-side variables. An h-step predictive VAR(p) takes the form

$$\boldsymbol{y}_{t+h} = b_0 + \boldsymbol{B}_1 \boldsymbol{y}_t + \cdots + \boldsymbol{B}_p \boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t.$$

There is an interesting relationship between a VAR model and the corresponding h-step predictive VAR model. This also means that $\widehat{IRF}(h) = \widehat{\boldsymbol{B}}_1$. This is known as the **local projection estimator**

I

f y_t is a VAR(p) process, then its h-step predictive regression is a predictive VAR(p) with $\boldsymbol{u}_t$ a $MA(h-1)$ process and $\boldsymbol{B}_1 = \boldsymbol{\Theta}_h = \text{IRF}(h)$.

We can see this with a simple AR(1) example....

- This method is currently extremely popular and we will discuss at length the good and the bad of local projections in the next weeks.

Granger Causality

- If a variable x does not help in predicting y we say that x *does not Granger-cause* y.
- In linear models this means that

$$\text{MSE}\left[\hat{E}\left(y_{t+s} \mid y_t, y_{t-1}, \ldots\right)\right] = \text{MSE}\left[\hat{E}\left(y_t + s \mid y_t, y_{t-1}, \ldots, x_t, x_{t-1}, \ldots\right)\right]$$

- Within a VAR it boils down to test that all coefficients on the variables x are equal to zero.
- IMPORTANT: this has nothing to do with causality. It is an unfortunate name. It only pertains to prediction.
- Hamilton has a nice counter-intuitive example you should look at.

What can VARs tell us about causality?

- To be able to discuss causality, we need to discuss identification.
- There is more and more interest in applying some of the terminology from micro to time series.
- Some debate on whether times series should be used for these questions or whether all we can learn is from reduced forms...
- we will discuss all this next.