Chapter 17

Multivariate Time Series

A multivariate time series $\boldsymbol{y}_t$ is a vector process $m \times 1$. Let $\mathcal{F}_{t-1} = (\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, ...)$ be all lagged information at time t. The typical goal is to find the conditional expectation $\mathbb{E}\left(\boldsymbol{y}_t \mid \mathcal{F}_{t-1}\right)$. Note that since $\boldsymbol{y}_t$ is a vector, this conditional expectation is also a vector.

17.1 Vector Autoregressions (VARs)

A VAR model specifies that the conditional mean is a function of only a finite number of lags:

$$\mathbb{E}\left(\boldsymbol{y}_t \mid \mathcal{F}_{t-1}\right) = \mathbb{E}\left(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_{t-k}\right).$$

A linear VAR specifies that this conditional mean is linear in the arguments:

$$\mathbb{E}\left(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_{t-k}\right) = \boldsymbol{a}_0 + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{A}_2 \boldsymbol{y}_{t-2} + \cdots \boldsymbol{A}_k \boldsymbol{y}_{t-k}.$$

Observe that $\boldsymbol{a}_0$ is $m \times 1$,and each of $\boldsymbol{A}_1$ through $\boldsymbol{A}_k$ are $m \times m$ matrices.

Defining the $m \times 1$ regression error

$$e_t = \boldsymbol{y}_t - \mathbb{E}\left(\boldsymbol{y}_t \mid \mathcal{F}_{t-1}\right),$$

we have the VAR model

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{a}_0 + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{A}_2 \boldsymbol{y}_{t-2} + \cdots \boldsymbol{A}_k \boldsymbol{y}_{t-k} + \boldsymbol{e}_t \\ \mathbb{E}\left(\boldsymbol{e}_t \mid \mathcal{F}_{t-1}\right) &= \boldsymbol{0}. \end{aligned}$$

Alternatively, defining the $mk + 1$ vector

$$\boldsymbol{x}_t = \begin{pmatrix} 1 \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-k} \end{pmatrix}$$

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

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_k \end{pmatrix},$$

then

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{e}_t.$$

The VAR model is a system of m equations. One way to write this is to let a_j' be the jth row of $\boldsymbol{A}$. Then the VAR system can be written as the equations

$$Y_{jt} = a_j' \boldsymbol{x}_t + e_{jt}.$$

Unrestricted VARs were introduced to econometrics by Sims (1980).

17.2 Estimation

Consider the moment conditions

$$\mathbb{E}\left(\boldsymbol{x}_t e_{jt}\right) = \mathbf{0},$$

$j = 1, ..., m$. These are implied by the VAR model, either as a regression, or as a linear projection.

The GMM estimator corresponding to these moment conditions is equation-by-equation OLS

$$\hat{\boldsymbol{a}}_j = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}_j.$$

An alternative way to compute this is as follows. Note that

$$\hat{\boldsymbol{a}}_j' = \boldsymbol{y}_j'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

And if we stack these to create the estimate $\hat{A}$, we find

$$\begin{aligned}\hat{\boldsymbol{A}} &= \begin{pmatrix} \boldsymbol{y}_1' \\ \boldsymbol{y}_2' \\ \vdots \\ \boldsymbol{y}_{m+1}' \end{pmatrix} \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1},\end{aligned}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_m \end{pmatrix}$$

the $T \times m$ matrix of the stacked $\boldsymbol{y}_t'$.

This (system) estimator is known as the SUR (Seemingly Unrelated Regressions) estimator, and was originally derived by Zellner (1962)

17.3 Restricted VARs

The unrestricted VAR is a system of m equations, each with the same set of regressors. A restricted VAR imposes restrictions on the system. For example, some regressors may be excluded from some of the equations. Restrictions may be imposed on individual equations, or across equations. The GMM framework gives a convenient method to impose such restrictions on estimation.

17.4 Single Equation from a VAR

Often, we are only interested in a single equation out of a VAR system. This takes the form

$$y_{jt} = \boldsymbol{a}_j'\boldsymbol{x}_t + e_t,$$

and $\boldsymbol{x}_t$ consists of lagged values of y_{jt} and the other $y_{lt}'s$. In this case, it is convenient to re-define the variables. Let $y_t = y_{jt}$, and $\boldsymbol{z}_t$ be the other variables. Let $e_t = e_{jt}$ and $\beta = a_j$. Then the single equation takes the form

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t, \tag{17.1}$$

and

$$\boldsymbol{x}_t = \left[\begin{pmatrix} 1 & \boldsymbol{y}_{t-1} & \cdots & \boldsymbol{y}_{t-k} & \boldsymbol{z}_{t-1}' & \cdots & \boldsymbol{z}_{t-k}' \end{pmatrix}'\right].$$

This is just a conventional regression with time series data.

17.5 Testing for Omitted Serial Correlation

Consider the problem of testing for omitted serial correlation in equation (17.1). Suppose that e_t is an AR(1). Then

$$\begin{aligned} y_t &= \boldsymbol{x}_t'\boldsymbol{\beta} + e_t \\ e_t &= \theta e_{t-1} + u_t \\ \mathbb{E}\left(u_t \mid \mathcal{F}_{t-1}\right) &= 0. \end{aligned} \tag{17.2}$$

Then the null and alternative are

$$\mathbb{H}_0 : \theta = 0 \qquad \mathbb{H}_1 : \theta \neq 0.$$

Take the equation $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t$, and subtract off the equation once lagged multiplied by θ, to get

$$\begin{aligned} y_t - \theta y_{t-1} &= \left(\boldsymbol{x}_t'\boldsymbol{\beta} + e_t\right) - \theta\left(\boldsymbol{x}_{t-1}'\boldsymbol{\beta} + e_{t-1}\right) \\ &= \boldsymbol{x}_t'\boldsymbol{\beta} - \theta \boldsymbol{x}_{t-1}\boldsymbol{\beta} + e_t - \theta e_{t-1}, \end{aligned}$$

or

$$y_t = \theta y_{t-1} + \boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{x}_{t-1}'\boldsymbol{\gamma} + u_t, \tag{17.3}$$

which is a valid regression model.

So testing $\mathbb{H}_0$ versus $\mathbb{H}_1$ is equivalent to testing for the significance of adding $(y_{t-1}, \boldsymbol{x}_{t-1})$ to the regression. This can be done by a Wald test. We see that an appropriate, general, and simple way to test for omitted serial correlation is to test the significance of extra lagged values of the dependent variable and regressors.

You may have heard of the Durbin-Watson test for omitted serial correlation, which once was very popular, and is still routinely reported by conventional regression packages. The DW test is appropriate only when regression $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + e_t$ is not dynamic (has no lagged values on the RHS), and e_t is iid $\mathrm{N}(0, \sigma^2)$. Otherwise it is invalid.

Another interesting fact is that (17.2) is a special case of (17.3), under the restriction $\boldsymbol{\gamma} = -\boldsymbol{\beta}\theta$. This restriction, which is called a common factor restriction, may be tested if desired. If valid, the model (17.2) may be estimated by iterated GLS. (A simple version of this estimator is called Cochrane-Orcutt.) Since the common factor restriction appears arbitrary, and is typically rejected empirically, direct estimation of (17.2) is uncommon in recent applications.

17.6 Selection of Lag Length in an VAR

If you want a data-dependent rule to pick the lag length k in a VAR, you may either use a testing-based approach (using, for example, the Wald statistic), or an information criterion approach. The formula for the AIC and BIC are

$$\begin{aligned} AIC(k) &= \log\det\left(\hat{\boldsymbol{\Omega}}(k)\right) + 2\frac{p}{T} \\ BIC(k) &= \log\det\left(\hat{\boldsymbol{\Omega}}(k)\right) + \frac{p\log(T)}{T} \\ \hat{\boldsymbol{\Omega}}(k) &= \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{e}}_t(k)\hat{\boldsymbol{e}}_t(k)' \\ p &= m(km+1) \end{aligned}$$

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

17.7 Granger Causality

Partition the data vector into $(\boldsymbol{y}_t, \boldsymbol{z}_t)$. Define the two information sets

$$\mathcal{F}_{1t} = \left(\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, ...\right)$$
$$\mathcal{F}_{2t} = \left(\boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{y}_{t-1}, \boldsymbol{z}_{t-1}, \boldsymbol{y}_{t-2}, \boldsymbol{z}_{t-2}, , ...\right)$$

The information set $\mathcal{F}_{1t}$ is generated only by the history of $\boldsymbol{y}_t$, and the information set $\mathcal{F}_{2t}$ is generated by both $\boldsymbol{y}_t$ and $\boldsymbol{z}_t$. The latter has more information.

We say that $\boldsymbol{z}_t$ does not *Granger-cause* $\boldsymbol{y}_t$ if

$$\mathbb{E}\left(\boldsymbol{y}_t \mid \mathcal{F}_{1,t-1}\right) = \mathbb{E}\left(\boldsymbol{y}_t \mid \mathcal{F}_{2,t-1}\right).$$

That is, conditional on information in lagged $\boldsymbol{y}_t$, lagged $\boldsymbol{z}_t$ does not help to forecast $\boldsymbol{y}_t$. If this condition does not hold, then we say that $\boldsymbol{z}_t$ Granger-causes $\boldsymbol{y}_t$.

The reason why we call this "Granger Causality" rather than "causality" is because this is not a physical or structure definition of causality. If $\boldsymbol{z}_t$ is some sort of forecast of the future, such as a futures price, then $\boldsymbol{z}_t$ may help to forecast $\boldsymbol{y}_t$ even though it does not "cause" $\boldsymbol{y}_t$. This definition of causality was developed by Granger (1969) and Sims (1972).

In a linear VAR, the equation for $\boldsymbol{y}_t$ is

$$\boldsymbol{y}_t = \alpha + \rho_1 \boldsymbol{y}_{t-1} + \cdots + \rho_k \boldsymbol{y}_{t-k} + \boldsymbol{z}'_{t-1}\gamma_1 + \cdots + \boldsymbol{z}'_{t-k}\gamma_k + e_t.$$

In this equation, $\boldsymbol{z}_t$ does not Granger-cause $\boldsymbol{y}_t$ if and only if

$$\mathbb{H}_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_k = 0.$$

This may be tested using an exclusion (Wald) test.

This idea can be applied to blocks of variables. That is, $\boldsymbol{y}_t$ and/or $\boldsymbol{z}_t$ can be vectors. The hypothesis can be tested by using the appropriate multivariate Wald test.

If it is found that $\boldsymbol{z}_t$ does not Granger-cause $\boldsymbol{y}_t$, then we deduce that our time-series model of $\mathbb{E}\left(\boldsymbol{y}_t \mid \mathcal{F}_{t-1}\right)$ does not require the use of $\boldsymbol{z}_t$. Note, however, that $\boldsymbol{z}_t$ may still be useful to explain other features of $\boldsymbol{y}_t$, such as the conditional variance.

Clive W. J. Granger

Clive Granger (1934-2009) of England was one of the leading figures in time-series econometrics, and co-winner in 2003 of the Nobel Memorial Prize in Economic Sciences (along with Robert Engle). In addition to formalizing the definition of causality known as Granger causality, he invented the concept of cointegration, introduced spectral methods into econometrics, and formalized methods for the combination of forecasts.

17.8 Cointegration

The idea of cointegration is due to Granger (1981), and was articulated in detail by Engle and Granger (1987).

Definition 17.8.1 *The $m \times 1$ series $\boldsymbol{y}_t$ is **cointegrated** if $\boldsymbol{y}_t$ is $I(1)$ yet there exists $\boldsymbol{\beta}$, $m \times r$, of rank r, such that $\boldsymbol{z}_t = \boldsymbol{\beta}'\boldsymbol{y}_t$ is $I(0)$. The r vectors in $\boldsymbol{\beta}$ are called the **cointegrating vectors**.*

If the series $\boldsymbol{y}_t$ is not cointegrated, then $r = 0$. If $r = m$, then $\boldsymbol{y}_t$ is $I(0)$. For $0 < r < m$, $\boldsymbol{y}_t$ is $I(1)$ and cointegrated.

In some cases, it may be believed that $\boldsymbol{\beta}$ is known a priori. Often, $\boldsymbol{\beta} = (1 \quad -1)'$. For example, if $\boldsymbol{y}_t$ is a pair of interest rates, then $\boldsymbol{\beta} = (1 \quad -1)'$ specifies that the spread (the difference in returns) is stationary. If $\boldsymbol{y} = (\log(C) \quad \log(I))'$, then $\boldsymbol{\beta} = (1 \quad -1)'$ specifies that $\log(C/I)$ is stationary.

In other cases, $\boldsymbol{\beta}$ may not be known.

If $\boldsymbol{y}_t$ is cointegrated with a single cointegrating vector ($r = 1$), then it turns out that $\boldsymbol{\beta}$ can be consistently estimated by an OLS regression of one component of $\boldsymbol{y}_t$ on the others. Thus $\boldsymbol{y}_t = (Y_{1t}, Y_{2t})$ and $\boldsymbol{\beta} = (\beta_1 \ \beta_2)$ and normalize $\beta_1 = 1$. Then $\hat{\beta}_2 = (\boldsymbol{y}_2'\boldsymbol{y}_2)^{-1}\boldsymbol{y}_2'\boldsymbol{y}_1 \xrightarrow{p} \beta_2$. Furthermore this estimation is super-consistent: $T(\hat{\beta}_2 - \beta_2) \xrightarrow{d} Limit$, as first shown by Stock (1987). This is not, in general, a good method to estimate $\boldsymbol{\beta}$, but it is useful in the construction of alternative estimators and tests.

We are often interested in testing the hypothesis of no cointegration:

$$\begin{aligned} \mathbb{H}_0 &: r = 0 \\ \mathbb{H}_1 &: r > 0. \end{aligned}$$

Suppose that $\boldsymbol{\beta}$ is known, so $\boldsymbol{z}_t = \boldsymbol{\beta}'\boldsymbol{y}_t$ is known. Then under $\mathbb{H}_0$ $\boldsymbol{z}_t$ is $I(1)$, yet under $\mathbb{H}_1$ $\boldsymbol{z}_t$ is $I(0)$. Thus $\mathbb{H}_0$ can be tested using a univariate ADF test on $\boldsymbol{z}_t$.

When $\boldsymbol{\beta}$ is unknown, Engle and Granger (1987) suggested using an ADF test on the estimated residual $\hat{z}_t = \hat{\boldsymbol{\beta}}'\boldsymbol{y}_t$, from OLS of y_{1t} on y_{2t}. Their justification was Stock's result that $\hat{\boldsymbol{\beta}}$ is super-consistent under $\mathbb{H}_1$. Under $\mathbb{H}_0$, however, $\hat{\boldsymbol{\beta}}$ is not consistent, so the ADF critical values are not appropriate. The asymptotic distribution was worked out by Phillips and Ouliaris (1990).

When the data have time trends, it may be necessary to include a time trend in the estimated cointegrating regression. Whether or not the time trend is included, the asymptotic distribution of the test is affected by the presence of the time trend. The asymptotic distribution was worked out in B. Hansen (1992).

17.9 Cointegrated VARs

We can write a VAR as

$$\begin{aligned} \boldsymbol{A}(\mathrm{L})\boldsymbol{y}_t &= \boldsymbol{e}_t \\ \boldsymbol{A}(\mathrm{L}) &= \boldsymbol{I} - \boldsymbol{A}_1\mathrm{L} - \boldsymbol{A}_2\mathrm{L}^2 - \cdots - \boldsymbol{A}_k\mathrm{L}^k \end{aligned}$$

or alternatively as

$$\Delta\boldsymbol{y}_t = \boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \boldsymbol{D}(\mathrm{L})\Delta\boldsymbol{y}_{t-1} + \boldsymbol{e}_t$$

where

$$\begin{aligned} \boldsymbol{\Pi} &= -\boldsymbol{A}(1) \\ &= -\boldsymbol{I} + \boldsymbol{A}_1 + \boldsymbol{A}_2 + \cdots + \boldsymbol{A}_k. \end{aligned}$$

Theorem 17.9.1 ***Granger Representation Theorem***
$\boldsymbol{y}_t$ *is cointegrated with* $m \times r$ $\boldsymbol{\beta}$ *if and only if* $\operatorname{rank}(\boldsymbol{\Pi}) = r$ *and* $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ *where* α *is* $m \times r$, $\operatorname{rank}(\boldsymbol{\alpha}) = r$.

Thus cointegration imposes a restriction upon the parameters of a VAR. The restricted model can be written as

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{D}(\mathrm{L})\Delta \boldsymbol{y}_{t-1} + \boldsymbol{e}_t$$
$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{z}_{t-1} + \boldsymbol{D}(\mathrm{L})\Delta \boldsymbol{y}_{t-1} + \boldsymbol{e}_t.$$

If $\boldsymbol{\beta}$ is known, this can be estimated by OLS of $\Delta \boldsymbol{y}_t$ on $\boldsymbol{z}_{t-1}$ and the lags of $\Delta \boldsymbol{y}_t$.

If $\boldsymbol{\beta}$ is unknown, then estimation is done by "reduced rank regression", which is least-squares subject to the stated restriction. Equivalently, this is the MLE of the restricted parameters under the assumption that $\boldsymbol{e}_t$ is iid $\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega})$.

One difficulty is that $\boldsymbol{\beta}$ is not identified without normalization. When $r = 1$, we typically just normalize one element to equal unity. When $r > 1$, this does not work, and different authors have adopted different identification schemes.

In the context of a cointegrated VAR estimated by reduced rank regression, it is simple to test for cointegration by testing the rank of $\boldsymbol{\Pi}$. These tests are constructed as likelihood ratio (LR) tests. As they were discovered by Johansen (1988, 1991, 1995), they are typically called the "Johansen Max and Trace" tests. Their asymptotic distributions are non-standard, and are similar to the Dickey-Fuller distributions.