University of York

Department of Economics

PhD Course 2006

VAR ANALYSIS IN MACROECONOMICS

Lecturer: Professor Mike Wickens

Lecture 2

VAR Models

Contents

1. Statistical v. econometric models

2. Statistical models

Data distributions

Static model

Dynamic model

Exogeneity and causality

Identification problem

3. Econometric models

Dynamic structure of simultaneous equation models

Final equation model

Final form model

VAR

4. Bayesian VAR analysis

5. Panel VAR models

6. Forward-looking rational expectations models

Single equation RE models

Simultaneous RE models

Impulse response functions

VAR analysis and the Lucas critique

1. Statistical v. Econometric models

"One definition of an economist is somebody who sees something happen in practice and wonders if it will work in theory." Ronald Reagan.

The main difference between a statistician and an economist in their use of time series analysis is that the statistician is trying to describe the data with a model, whereas the economist has in mind trying to learn about the economy or about economy theory. For the economist, being able to represent the data by a model is not sufficient; the model also has to make economic sense.

The quote seems to me illustrate this point. Reagan obviously intended his comment as a criticism of economists. Without intending to, it seems to me that he has made a profound observation. I interpret this quote to say: being able to find a good statistical model is not sufficient; the advancement of knowledge requires that we can explain why the statistical model fits the data. If all that were required was the construction of good statistical models of economic data then why study economics at all? We would all be able to save ourselves a lot of time.

A model that is misspecified - for example by wrongly omitting variables - may still fit the data very well, but the estimates may make little economic sense. The classic case is modelling the relation between quantity and price with a view to estimating demand. The estimates may show a statistically well-behaved positive relation. Does this mean that the demand curve slopes upwards? An alternative explanation might be that income has been omitted from the model and caused the coefficient on price to be biased upwards. If price and income are both increasing over time (as is often the case) then their correlation would be positive which would cause positive bias in the price coefficient.

We contrast the two approaches and conclude that in order to be of use to the economist, time series models must be identified by the use of appropriate restrictions.

Purely statistical approach

Can we derive the model from the data alone?

Suppose we start at a very primitive point with data from two series $\{y_t, x_t\}$ and T observations on each i.e. $\{t = 1, ..., T\}$.

We now consider the joint distribution for y_t and x_t,

i.e. the distribution of $z_t = (y_t,\ x_t)$.

Let this be

$$D(y_t, x_t; \theta) = D(z_t; \theta)$$

where θ are the parameters of the distribution.

We are interested in the relation between y_t and x_t.

Is it static or dynamic?

What is the causal structure, if any?

1. Static models

Using Bayes theorem we can write the joint distribution as

$$D(y_t, x_t; \theta) = D(y_t|x_t; \phi)D(x_t; \theta_2)$$

$D(y_t|x_t; \phi)$ = the conditional distribution of y_t given x_t and ϕ is a function of the parameters θ.

$D(x_t; \theta_2)$ = the marginal distribution of x_t and θ_2 is a sub-set of the parameters θ.

If $D(y_t, x_t; \theta) \sim N(\mu, \Sigma)$ then

the mean of the conditional distribution can be written

$$E(y_t|x_t) = \alpha + \beta x_t$$

If we define

$$e_t = y_t - E(y_t|x_t)$$

then

$$\begin{aligned}
y_t &= \alpha + \beta x_t + e_t \\
\beta &= \frac{\sigma_{xy}}{\sigma_{xx}} \\
\alpha &= \mu_y - \beta\mu_x \\
\sigma^2 &= \sigma_{yy} - \beta^2\sigma_{xx} \\
D(e_t|x_t) &\sim N(0, \sigma^2)
\end{aligned}$$

Thus the static linear model between y_t and x_t is obtained from the mean of the conditional distribution of y_t given x_t.

There is no requirement here that x_t be a "fixed" regressor; x_t can be stochastic, i.e. a random variable.

To estimate the model we could use OLS

Or we could calculate the sample moments of the data and then replace the unknown population parameters by their sample values.

The result would be the same, apart from possible degrees of freedom corrections.

The distribution of these estimates would be conditional on the particular sample values of x_t.

Inference would otherwise proceed as usual.

The key thing here is that there is no need to introduce the assumption of fixed x_t.

This result is of fundamental importance for economics. It implies that economics has every right to be categorised as a science even though controlled experiments may not be possible. The classical notion that science requires replication is therefore unnecessary.

In economics the economy produces just one set of values for x_t and corresponding values for y_t. In economics we just make our inferences *conditional* on the one set of data we have, i.e. on x_t.

The same thing occurs in cosmology.

2. Dynamic models

Consider now the joint distribution of all of the observations $\{t = 1, ..., T\}$, i.e. of $z = \{z_1, z_2, ..., z_T\}$

$$D(z; \theta) = D(z_1, ..., z_T; \theta)$$

Two cases can be distinguished

(i) The z_t are independent of each other

(ii) The z_t are not independent.

(i) z_t independent

If the observations are independent across time then

$$D(z_1, ..., z_T; \theta) = \prod_{t=1}^{T} D(z_t; \theta)$$

This implies that we can consider each distribution $D(z_t; \theta)$ on its own.

It would therefore give us the static model just considered.

(ii) z_t NOT independent

This is the usual situation in economics and is the key ingredient for time series modeling.

We can show that for $T = 3$

$$\begin{aligned} D(z_1, z_2, z_3) &= D(z_3|z_2, z_1)D(z_2, z_1) \\ &= D(z_3|z_2, z_1)D(z_2|z_1)D(z_1) \end{aligned}$$

Hence

$$D(z_1, ..., z_T; \theta) = \prod_{t=1}^{T} D(z_t|z_{t-1}, ..., z_1; \phi_t)$$

where $D(z_t|z_{t-1}, ..., z_1; \phi_t)$ is the conditional distribution of z_t given $z_{t-1}, ..., z_1$ and ϕ_t is a function of θ.

Note: if the z_t were independent then

$$D(z_t|z_{t-1}, ..., z_1) = D(z_t)$$

If we make the additional assumption that the distribution is Normal and given by

$$D(z; \theta) \sim N(\mu, \Sigma)$$

then the mean of the conditional distribution is *linear*:

$$E(z_t|z_{t-1}, ..., z_1) = \sum_{s=1}^{t-1} A_s z_{t-s}$$

If we define the deviation of z_t from the conditional mean as

$$e_t = z_t - E(z_t|z_{t-1}, ..., z_1)$$

then we have the VAR

$$z_t = \sum_{s=1}^{t-1} A_s z_{t-s} + e_t$$

Thus a VAR is a generic representation of the mean of the conditional distribution of z_t given $z_{t-1}, ..., z_1$.

But notice that this VAR has $t - 1$ lags!

This implies that there are more parameters to estimate than there are total observations! It is therefore impossible to estimate the parameters.

The important implication is that we need to restrict the order of the VAR to be much less than $t-1$ to make progress. Hence we need to introduce restrictions into the statistical approach. But statistics does not tell us how to do this.

Exogeneity and causality

Recall that $z_t = (y_t, x_t)$

Hence

$$D(z_t|z_{t-1},...,z_1;\phi_t) = D(y_t|x_t, z_{t-1},...,z_1;\varphi_1)D(x_t|z_{t-1},...,z_1;\varphi_2)$$

1. Weak exogeneity

x_t is said to be weakly exogenous for y_t if φ_1 is independent of φ_2.

T

2. Strong exogeneity

x_t is said to be strongly exogenous for y_t if it is weakly exogenous and

$$D(x_t|z_{t-1},...,z_1;\varphi_2) = D(x_t|x_{t-1},...,x_1;\varphi_2)$$

i.e. if lagged values of y_t don't help to determine x_t.

3. Granger causality

y_t is said to NOT Granger-cause x_t if, without restricting φ_1 and φ_2 to be independent,

$$D(z_t|z_{t-1},...,z_1;\phi_t) = D(y_t|x_t, z_{t-1},...,z_1;\varphi_1)D(x_t|x_{t-1},...,x_1;\varphi_2)$$

This implies that we are looking for block-exogeneity in the model for x_t.

Another implication is

$$weak\ exogeneity + Granger\ non-causality \equiv strong\ exogeneity$$

Identification

Suppose we restrict the lag structure. Will this be sufficient?

Consider the following equations derived from the conditional means

$$\begin{aligned} E(y_t|x_t, z_{t-1}) &= \beta x_t + \alpha_{11} y_{t-1} + \alpha_{12} x_{t-1} \\ E(x_t|z_{t-1}) &= \alpha_{21} y_{t-1} + \alpha_{22} x_{t-1} \end{aligned}$$

This can be rewritten as

$$\begin{aligned} y_t &= \beta x_t + \alpha_{11} y_{t-1} + \alpha_{12} x_{t-1} + e_{1t} \\ x_t &= \alpha_{21} y_{t-1} + \alpha_{22} x_{t-1} + e_{2t} \\ E(e_{1t} e_{2t}) &= 0 \end{aligned}$$

Note that the first equation involves the current value x_t. The uncorrelatedness of the errors follows from Bayes theorem and the factorization of the joint density function.

In general the first equation is not identified as we can take a linear combination of the two equations that will have the same variables as the first equation.

eg take (1)+$\lambda \times$ (2)

$$\begin{aligned} y_t &= (\beta - \lambda) x_t + (\alpha_{11} + \lambda a_{21}) y_{t-1} + (\alpha_{12} + \lambda a_{22}) x_{t-1} + u_t \\ u_t &= e_{1t} + \lambda e_{2t} \end{aligned}$$

If we estimate the equation for y_t how do we know which equation we have estimated, or what value λ has?

The implication of this is that just restricting the lag structure is not sufficient; we need further restrictions on the conditional distribution of y_t.

Note that we can form a VAR from these two equations as they can be written

$$\begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} a_{11} + \beta a_{21} & a_{12} + \beta a_{22} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} + \beta e_{2t} \\ e_{2t} \end{bmatrix} \\ &= \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ a_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \end{aligned}$$

where the α_{ij} and the covariance matrix of the errors are treated as unrestricted.

It follows that, even though the original model is not identified, the associated VAR is identified.

We have already shown that the original structural equation for y_t is not identified and hence we cannot learn about it from the data. The implication of this is that we cannot infer the original model from the VAR.

This is an example of the difficulty of trying to learn about structural economic models from a VAR.

To resolve the issue we need to introduce additional restrictions.

What type of identifying restrictions could we use?

We could

(i) delete a variable from the equation for y_t.

As a result we cannot form a linear combination of the two equations that has the same explanatory variables as the original equation.

Note that that this can only be done for the equation for y_t; deleting a variable from the x_t equation would not work.

(ii) restrict the error terms u_{1t} and u_{2t} in the VAR to be independent.

In forming the VAR we took a linear combination of the original errors e_{1t} and e_{2t}. As a result, the error terms in the VAR will in general be correlated even if e_{1t} and e_{2t} are uncorrelated.

Thus

$$E(u_{1t}^2) = \sigma_{11} + \beta^2\sigma_{22},\ E(u_{2t}^2) = \sigma_{22},\ E(u_{1t}u_{2t}) = \beta\sigma_{22}$$

Hence, we can recover β from the covariance matrix of the VAR errors as

$$\beta = \frac{E(u_{1t}u_{2t})}{E(u_{2t}^2)},\ \sigma_{11} = E(u_{1t}^2) - \beta^2 E(u_{2t}^2)$$

If we were to restrict the VAR errors u_{1t} and u_{2t} to be uncorrelated it would follow that $\beta = 0$. The structural y_t equation would then become the reduced form and the VAR equation for y_t. This would resolve the identification problem.

In general, if y_t and x_t consisted of a vector of variables then the model for y_t would be more complicated. Recovering all of the coefficients associated with the contemporaneous variables from the covariance matrix of the errors would not then be possible.

The model for y_t would take the form

$$\begin{aligned} By_t &= Cx_t + Dz_{t-1} + e_{1t},\ e_{1t} \sim N(0, \Sigma_{11}),\ B_{ii} = 1 \\ x_t &= Fz_{t-1} + e_{2t},\ e_{2t} \sim N(0, \Sigma_{22}) \\ E(e_{1t}e'_{2t}) &= 0 \end{aligned}$$

Hence

$$\begin{bmatrix} B & -C \\ 0 & I \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} D \\ F \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} B & -C \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} D \\ F \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} B & -C \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \\ &= \begin{bmatrix} B^{-1} & B^{-1}C \\ 0 & I \end{bmatrix} \begin{bmatrix} D \\ F \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \end{aligned}$$

where

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} B^{-1} & B^{-1}C \\ 0 & I \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

Thus we can recover

$$B^{-1}C = E(u_{1t}u'_{2t})E(u_{2t}u'_{2t})^{-1}$$

but not B and C.

More generally, when the model has more than two variables we need to use the standard rank and order identification results for simultaneous equation systems.

i.e. the order condition is that the identification of each equation can be achieved if each equation has $n-1$ exclusion (zero) restrictions.

Notice that the equations for x_t satisfy the order condition.

2. Relation with standard econometric models

1. The SEM and the VAR

If y_t and x_t are vectors of variables then the SEM can be written

$$B_{11}y_t + B_{12}x_t = A_{11}y_{t-1} + A_{12}x_{t-1} + e_{1t}$$

where x_t are (weakly) exogenous variables.

The reduced form can be written

$$y_t = \beta x_t + \alpha_{11}y_{t-1} + \alpha_{12}x_{t-1} + B_{11}^{-1}e_{1t}$$

where all the coefficients are matrices. For example, $\beta = -B_{11}^{-1}B_{12}$ etc

Thus the reduced form model defines $E(y_{it}|x_t, z_{t-1})$.

In general:

the SEM

(i) includes current values of other endogenous and exogenous variables in each equation

(ii) does not include equations for the exogenous variables.

the reduced form

(i) includes current values only of exogenous variables in each equation

(ii) does not include equations for the exogenous variables.

In contrast, the VAR

(i) does not include any other current variable

(ii) includes equations for x_t

(iii) makes no distinction between endogenous and exogenous variables.

Further, a VAR explains how the shocks (of the VAR) can be transmitted both to y_t and x_t.

The main problem for economists is how to give an economic interpretation to these shocks.

This is equivalent to saying that a VAR has a fundamental identification problem.

2. The SEM and dynamics

Starting with an SEM what can be said about the dynamic structure of the model and how the shocks get transmitted to the variables?

Final Equation

The SEM can be written

$$B(L)y_t = C(L)x_t + e_t$$

Noting that

$$B(L)^{-1} = \frac{adj[B(L)]}{|B(L)|}$$

we can write the SEM as

$$|B(L)|y_t = adj[B(L)]C(L)x_t + adj[B(L)]e_t$$

This is known as the FINAL EQUATION.

The equation

$$|B(L)| = 0$$

gives the internal dynamics of the SEM (i.e. of y_t).

This will be the same for each y_t unless there are cancelling factors.

Although the final equation also shows the dynamic response of y_t to shocks to the structural errors, this is better captured by the final form.

Final Form

This is given by

$$y_t = B(L)^{-1}C(L)x_t + B(L)^{-1}e_t$$

It shows the dynamic response of y_t to a change in x_t, or e_t. Unlike the final equation, the final form allows the dynamic behaviour of each element of y_t to be analysed without taking account of the responses of the other elements of y_t.

In both the final equation and the final form the responses of y_t to the shocks can be analysed independently of the x_t terms, and vice-versa.

3. VAR

To obtain the VAR we must add equations for x_t and combine these with the equations for y_t.

Suppose the equations for x_t are

$$D(L)x_t = F(L)y_{t-1} + \varepsilon_t$$

where $D_0 = I$ as we do not have a structural explanation for the x_t. (If we had then we would have included the x_t with the y_t. All variables would then be endogenous and there would no exogenous variables.)

Further, $E(e_t \varepsilon_t') = 0$ and ε_t is serially uncorrelated. This implies that x_t is weakly exogenous.

If in addition $F(L) = 0$ then x_t is strongly exogenous.

Adding the equations for x_t to the SEM gives a model that determines all of the variables. This can be written

$$\begin{aligned} \begin{bmatrix} B(L) & -C(L) \\ -F(L)L & D(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} e_t \\ \varepsilon_t \end{bmatrix} \\ G(L) \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} e_t \\ \varepsilon_t \end{bmatrix} \end{aligned}$$

Writing $G(L) = G_0 - G^*(L)L$, and noting that

$$\begin{aligned} G_0 &= \begin{bmatrix} B_0 & -C_0 \\ 0 & I \end{bmatrix} \\ G_0^{-1} &= \begin{bmatrix} B_0^{-1} & B_0^{-1}C_0 \\ 0 & I \end{bmatrix} \end{aligned}$$

we can obtain the VAR by solving $(y_t \ x_t)'$ in terms of their lags giving

$$\begin{aligned}\begin{bmatrix} y_t \\ x_t \end{bmatrix} &= G_0^{-1}G^*(L)\begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + G_0^{-1}\begin{bmatrix} e_t \\ \varepsilon_t \end{bmatrix} \\ &= A^*(L)\begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}\end{aligned}$$

or

$$A(L)\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where$A(L) = I - A^*(L)L$

From the structure of G_0^{-1} the equations for x_t are the same as the original x_t equations; only the equations for y_t are changed.

As previously noted, it is usually assumed that we cannot say which variables in a VAR are exogenous. All variables are treated in the same way.

Later we shall argue that it may be very useful to be able to make such a distinction.

Thus in a structural model we can derive the effect of changes in the x_t and the structural errors e_t on the y_t, but in a VAR all we can do is obtain the effect of the VAR shocks u_{1t} and u_{2t}.

If we knew which were the exogenous variables and ordered the variables appropriately then the u_{2t} could be interpreted as the shocks to the exogenous variables, the ε_t. But typically in a VAR we are not willing to make a distinction between endogenous and exogenous variables. As a result we are unable to obtain the effect of the exogenous on the endogenous variables or know which shocks are structural shocks.

The above analysis has made no distinction between stationary and non-stationary variables. We shall need to examine the consequences of this distinction and how to deal with non-stationary variables in a VAR. One implication is that some of the shocks will be permanent and others temporary. We will show that the temporary shocks are the structural shocks defined above, and the permanent shocks are all shocks to the exogenous variables.

4. Bayesian VAR analysis

It is a standard result in econometrics that including variables in a regression that have zero coefficients reduces the efficiency (raises the variances) of the estimates of the coefficients that are non-zero and the standard errors of out-of-sample forecasts.

Unrestricted VARs tend to have a large number of coefficients to estimate, many of which are close to zero. According to this argument it might therefore be better if one could set near-zero coefficients to zero.

Instead of setting the coefficients to zero, an alternative approach has been proposed by Doan, Litterman and Sims (1984) and Litterman (1986). This involves "shrinking" the badly determined coefficient estimates towards zero by the use of a prior weighting function. This is sometimes called a Minnesota or Litterman prior.

It is also an example of Bayesian estimation with a Normal prior.

Suppose that b is an estimator of β with covariance matrix V_b but there is a prior belief about β that it takes the value θ with a degree of uncertainty. Suppose that this uncertainty can be expressed by the prior Normal distribution $p(\beta) \sim N(\theta, \Omega)$. Then it follows from Bayes theorem that

$$p(\beta|z) = \frac{D(z|\beta)p(\beta)}{D(z)} = L(z|\beta)p(\beta)$$

where z is the data, $D(z|\beta)$ is the distribution of the data conditional on β. $D(z)$ is the unconditional or marginal distribution of the data over all possible values of β. $L(z|\beta)$ is the sample likelihood function. This can be replaced by the distribution of b to give.

$$p(\beta|z) = f(b|\beta)p(\beta)$$

$p(\beta|z)$ is the distribution of β conditional on the particular sample z. It is also known as the posterior distribution of β. The mean of this distribution can be used as the preferred estimator of β.

In the case where the distribution of the estimator b and the prior distribution are both Normal, the resulting Bayesian estimator of β is

$$\widehat{\beta} = [V_b^{-1} + \Omega^{-1}]^{-1}(V_b^{-1}b + \Omega^{-1}\theta)$$

This is a weighted average of b and θ where the weights are proportional to the inverse covariance matrices.

In implementing this for a VAR, in order to minimise the lag length, it is usual to set $\theta = 0$ and to make Ω a diagonal matrix. If the lag coefficients are expected to decline as the length of the lag increases then this can be reflected by making the corresponding diagonal element of Ω smaller. This would then give a larger weight to setting the coefficient to zero, thereby shrinking the estimate to zero compared with its regression estimate. This estimation procedure has been implemented in RATS, but not in EViews4. It can, however, be programmed in EViews.

5. Panel VAR Models (PVAR)

Reading:

Cheng Hsiao, *Analysis of Panel Data*, 2nd ed., Wiley, 2003

Introduction

A model can be defined to hold both over time and across different agents.

For example, one might have a model (consisting of a number of variables) for many different countries (where it is assumed that the model in each country is basically the same), and time series data for each variable in each country. Instead of analysing each country separately, one may wish to analyse the countries together in order to exploit the fact that they have the same economic structure. This is an example of panel data.

(i) When we have data over time for the same agent these are known as time-series data

(ii) When we have data at one point of time on different agents these are known as cross-section data

(iii) When we have data over time for different agents we have PANEL data.

Panel data can be balanced (i.e. we have the same number of time periods for each and every agent) or unbalanced (i.e. the number of time periods differs across agents).

The model can be static over time or dynamic.

Static model

$$
\begin{aligned}
y_{it} &= \alpha + \beta' x_{it} + \gamma' z_{it} + e_{it} \\
i &= 1, ..., N; \ t = 1, ..., T
\end{aligned}
$$

y_{it} = endogenous variable

x_{it} = vector of exogenous variables

z_{it} = vector of unobservable variables, uncorrelated with x_{it}.

e_{it} = vector of disturbances

α = intercept

β and γ = vectors of parameters

(a) Fixed effects model

Each agent (cross-section element) has a different constant mean:

$$y_{it} = \alpha + \beta' x_{it} + \delta_i + e_{it}, \quad Ex_{it} = Ee_{it} = 0$$

We could also interpret this as $\gamma = 0$ and there is one element of $x_{it} = \delta_i$.

For the purposes of estimation it is common to eliminate the fixed effects by taking first differences over time:

$$\begin{aligned} \Delta y_{it} &= \beta' \Delta x_{it} + \Delta e_{it}, \quad E\Delta x_{it} = E\Delta e_{it} = 0 \\ t &= 2, ..., T \end{aligned}$$

The problem is that the resulting disturbance term becomes serially correlated as $E(\Delta e_{it} \Delta e_{i,t-1}) = Ee_{i,t-1}^2$. We therefore need to use a GMM or IV estimator. A large literature is emerging on this problem.

(b) Random effects model

$$\begin{aligned} y_{it} &= \alpha + \beta' x_{it} + e_{it}, \quad Ex_{it} = Ee_{it} = 0 \\ e_{it} &= \delta_i + \lambda_t + \varepsilon_{it}, \text{ all random components} \\ E\delta_i &= E\lambda_t = E\varepsilon_{it} = 0 \text{ and mutually uncorrelated} \end{aligned}$$

Could also interpret this as z_{it} equal δ_i or λ_t, i.e. some vary over i and some over t.

Dynamic model

We now introduce dynamics into the model. A simple example for y_t, a scalar, is

$$y_{it} = \theta y_{i,t-1} + \alpha + \beta' x_{it} + \delta_i + \lambda_t + e_{it}$$

$\alpha, \delta_i, \lambda_t$ may be constant

Again fixed effects are commonly eliminated by first differencing

$$\Delta y_{it} = \theta \Delta y_{i,t-1} + \beta' \Delta x_{it} + \Delta \lambda_t + \Delta e_{it}$$

There is now an extra problem to the serial correlation introduced into the error term. This is that $\Delta y_{i,t-1}$ will be correlated with Δe_{it}. As a result OLS will give biased estimates. Unfortunately, this tends to be ignored in practice.

PVAR model

PVAR means a panel VAR. We now allow y_{it} to be an $m \times 1$ a vector of variables

1. Basic PVAR

$$\begin{aligned} A(L) y_{it} &= \alpha_i + e_{it} \\ i &= 1, ..., N;\ t = 1, ..., T \end{aligned}$$

α_i = vector of fixed effects for each element of y_{it}

Eliminating fixed effects by first differencing would imply that

$$A(L) \Delta y_{it} = \Delta e_{it}$$

Again this introduces serial correlation into the error term and so lags of Δy_{it} will be correlated with Δe_{it} and OLS will give biased estimates. Again, this tends to be ignored in practice.

Estimation is problematical. We need to find suitable instruments and then use GMM or IV. Extra lags of y_{it} may be a possibility as intstrument variables.

Note:

Recall the IV estimator for the model

$$y = X\beta + e$$

is

$$b_{IV} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y$$

where Z are the instrumental variables. If we choose $Z = X$ then IV≡OLS.

Although we now have an estimate of $A(L)$, we can't separate α_i from the error term.

And if we use OLS on the differenced version then we have the problem of only being able to calculate impulse response functions for Δe_{it}.

Estimation of the PVAR when $\alpha_i = 0$

The cross-section error terms will in general be correlated, .i.e. $E(e_{it}e_{jt}) = \Sigma_{ij}$.

Thus we write the model as

$$[A(L) \otimes I_N] \begin{bmatrix} y_{1t} \\ : \\ y_{Nt} \end{bmatrix} = \begin{bmatrix} e_{1t} \\ : \\ e_{Nt} \end{bmatrix}$$

or

$$[A(L) \otimes I]z_t = u_t$$

where $z'_t = [y'_{1t},..., y'_{Nt}]$ and $u'_t = [e'_{1t},..., e'_{Nt}]$, $E(u_t) = 0$ and $E(u_t u'_t) = \Sigma = \{\Sigma_{ij}\}$.

For example, a first-order PVAR, the model can be written

$$y_{it} = Ay_{i,t-1} + e_{it}$$

and hence

$$\begin{aligned} z_t &= [A \otimes I_N] z_{t-1} + u_t \\ &= B z_{t-1} + u_t \end{aligned}$$

Thus we obtain a first-order VAR in z_t but with the difference that the coefficient matrix B is restricted. This implies that we cannot use OLS. Instead, we should use maximum likelihood estimation in which we take account of the restrictions.

Impulse response functions

The problem now is that the errors across are correlated.

2. Cointegrated PVAR

We discuss cointegration in detail in a later lecture. It is, however, convenient to deal with its implications for PVARs here.

We now assume that the y_{it} are I(1). If we re-write

$$A(L) = A^*(L)(I - L) - A(1)L$$

where

$$A(1) = \alpha\beta'$$

and β is an $r \times m$ matrix of r cointegrating vectors, then the CPVAR model is

$$A^*(L)\Delta y_{it} = \alpha\beta' y_{i,t-1} + \delta_i + e_{it}$$

$\delta_i = m \times 1$ vector of parameters.

The implicit assumption here is that the r cointegrating vectors are the same for each cross-section unit i.

Tests for cointegration and the estimation of cointegrating vectors are typically made by first estimating the model for each $i = 1, ..., N$ and then averaging over the N cross-sections

Eliminating fixed effects by first differencing would imply that

$$A^*(L)\Delta^2 y_{it} = \alpha\beta'\Delta y_{i,t-1} + \Delta e_{it}$$

again creating problems for the interpretation of the impulse response functions of the CPVAR.

Estimation will again need to be undertaken using MLE or the equivalent.

6. Forward-looking rational expectations models

- It is increasingly common in modern monetary policy analysis to use small structural models instead of VAR models. These are often based on DSGE models and have New Keynesian price equations. A feature of these models is the presence of variables that are the rational expectation of future variables using current information. This is unlike the previous use of RE where the rational expectations were usually of *current* variables and based on past information.

- The assumption of rational expectations has proved highly controversial due to the extremely strong informational requirements they entail. Arguably, however, this criticism has been overdone. The assumption of RE is better thought of as analagous to that of perfect competition: it is a limiting but useful benchmark.

- In practice, of course, it is highly improbable that expectations can be fully rational. But the aim of these models is more to capture forward-lookingness in decision making than strong rationality. Expectations can be based on the organizing structure of the model and the optimal use (in a statistical sense) of current and past data. They are therefore better thought of as consistent than rational.

- We shall examine

(i) the solution of forward-looking RE models

(ii) the implications of such structural models for VAR analysis

(iii) the estimation of these models. The problem of estimating RE models where the expectation is of a current variable and the information set is dated in the past is well-known (see Wallis (1982) and Wickens (1982) and Pesaran (1987)).

First we consider single equation models to establish the issues, then we turn to simultaneous equation models.

Single equation forward-looking RE models.

Consider the model

$$\begin{aligned} y_t &= \alpha E_t[y_{t+1}] + \delta y_{t-1} + \beta x_t + e_t \\ x_t &= \theta_1 x_{t-1} + \theta_2 x_{t-2} + \varepsilon_t \end{aligned}$$

where e_t and ε_t are assumed to be serially and mutually uncorrelated.

Using the lag operator L which enables us to write $x_{t-n} = L^n x_t$ and $E_t[x_{t+n}] = L^{-n} x_t$, we can express the difference equation as

$$(-\alpha L^{-1} + 1 - \delta L) y_t = \beta x_t + e_t$$

This has the characteristic equation

$$-\alpha L^{-1} + 1 - \delta L = 0$$

implying that

$$-\delta L^{-1}[L^2 - \frac{1}{\delta}L + \frac{\alpha}{\delta}] = 0$$

The solution of

$$L^2 - \frac{1}{\delta}L + \frac{\alpha}{\delta} = 0$$

can be written

$$(L - \gamma_1)(L - \gamma_2) = 0$$

If the solution has saddlepath dynamics then one root is stable and the other is unstable. Let γ_1 be the stable root and γ_2 the unstable root, then $|\gamma_1| < 1$ and $|\gamma_2| > 1$. In this case

$$(L - \gamma_1)(L - \gamma_2)|_{L=1} < 0$$

Thus we can quickly check whether the difference equation has saddlepath dynamics from

$$\begin{aligned}(L-\gamma_1)(L-\gamma_2)|_{L=1} &= (L^2-\frac{1}{\delta}L+\frac{\alpha}{\delta})|_{L=1} \\ &= 1-\frac{1}{\delta}+\frac{\alpha}{\delta}<0\end{aligned}$$

We can now re-write the difference equation as

$$-\delta L^{-1}(L-\gamma_1)(L-\gamma_2)\,y_t=\beta x_t+e_t$$

or as

$$\left(1-\gamma_1 L^{-1}\right)\left(1-\gamma_2^{-1}L\right)y_t=\frac{\beta}{\delta\gamma_2}x_t+\frac{1}{\delta}e_t$$

Hence

$$\begin{aligned}y_t &= \frac{\beta}{\delta\gamma_2}\sum_{s=0}^{\infty}\gamma_1^s L^{-s}x_t+\frac{1}{\gamma_2}Ly_t+\frac{1}{\delta}e_t \\ &= \frac{\beta}{\delta\gamma_2}\sum_{s=0}^{\infty}\gamma_1^s E_t[x_{t+s}]+\frac{1}{\gamma_2}y_{t-1}+\frac{1}{\delta}e_t\end{aligned}$$

y_t can also be written as the partial adjustment model

$$\Delta y_t=(1-\frac{1}{\gamma_2})[y_t^*-y_{t-1}]+\frac{1}{\delta}e_t$$

where y_t^*, the long-run solution for y_t, is

$$y_t^*=\frac{\beta}{\delta(\gamma_2-1)}\sum_{s=0}^{\infty}\gamma_1^s E_t[x_{t+s}]$$

We now need to replace $E_t[x_{t+s}]$ using the equation for x_t. If x_t is an AR(2) then we can write

$$E_t[x_{t+s}]=\phi_{1s}x_t+\phi_{2s}x_{t-1}$$

The solution for y_t then takes the form

$$y_t=\lambda_1 x_t+\lambda_2 x_{t-1}+\lambda_3 y_{t-1}+u_t$$

where $u_t=\frac{1}{\delta}e_t$.

We have therefore transformed the RE model to a standard dynamic structural model without RE variables of the sort that we considered previously.

We can use the solution to solve for $E_t[y_{t+1}]$. This is

$$\begin{aligned} E_t[y_{t+1}] &= \lambda_1 E_t[x_{t+1}] + \lambda_2 x_t + \lambda_3 y_t \\ &= (\lambda_1\theta_1 + \lambda_2)x_t + \lambda_1\theta_2 x_{t-1} + \lambda_3 y_t \end{aligned}$$

Substituting this into the original model gives

$$\begin{aligned} y_t &= \alpha[(\lambda_1\theta_1 + \lambda_2)x_t + \lambda_1\theta_2 x_{t-1} + \lambda_3 y_t] + \delta y_{t-1} + \beta x_t + e_t \\ &= \frac{\alpha(\lambda_1\theta_1 + \lambda_2) + \beta}{1 - \alpha\lambda_3} x_t + \frac{a\lambda_1\theta_2}{1 - \alpha\lambda_3} x_{t-1} + \frac{\delta}{1 - \alpha\lambda_3} y_{t-1} + \frac{1}{1 - \alpha\lambda_3} e_t \end{aligned}$$

If this is solved for y_t then we simply return to the previous solution.

We note two further things about the solution:

1. Identification:

   The presence of x_{t-2} in the equation for x_t is crucial for identification. In its absence $E_t[y_{t+1}]$ would be determined just by x_t and y_t, implying that $E_t[y_{t+1}]$ would be perfectly correlated with the other variables in the model for y_t. Estimation of α would then be impossible.

2. Estimation:

   The presence of y_t in $E_t[y_{t+1}]$ implies that it will correlated with e_t. Thus one can't simply substitute a forecast for $E_t[y_{t+1}]$ based on regressing y_{t+1} on x_t, x_{t-1} and y_t. It will be necessary to instrument the forecast too. Valid instruments are x_t, x_{t-1} and y_{t-1}.

Thus, if the forecast for y_{t+1} is $\hat{y}_{t+1}$we substitute this into the original equation to obtain

$$
\begin{aligned}
y_t &= \alpha \hat{y}_{t+1} + \delta y_{t-1} + \beta x_t + v_t \\
v_t &= e_t + \alpha(E_t[y_{t+1}] - \hat{y}_{t+1})
\end{aligned}
$$

which can be consistently estimated by IV (GMM) using (x_t, x_{t-1}, y_{t-1}) as instruments.

Simultaneous systems with future expectations

It is possible to generalise the methodology to the case of a system of equations

$$\begin{aligned} FE_t[y_{t+1}] + B(L)y_t + C(L)x_t &= e_t \\ D(L)x_t &= \varepsilon_t \end{aligned}$$

Using the lag operator the model can be written

$$A(L)y_t = -C(L)x_t + e_t$$

where

$$A(L) = B(L) + C(L) + FL^{-1}$$

We denote the roots of

$$|A(L)| = 0$$

as λ_i $(i = 1, .., p)$ and $\gamma_j = (j = 1, ..., q)$ where $p + q = n$, λ_i are the unstable roots and γ_j are the stable roots. Thus $|\lambda_i| \leq 1$ and $|\gamma_j| > 1$ and

$$\begin{aligned} |A(L)| &= aL^{-p} \prod_{i=1}^{p}(L - \lambda_i) \prod_{i=1}^{q}(L - \gamma_j) \\ &= b \prod_{i=1}^{p}(1 - \lambda_i L^{-1}) \prod_{i=1}^{q}(1 - \gamma_j^{-1} L) \end{aligned}$$

where

$$b = a \prod_{i=1}^{q}(-\gamma_j)$$

Hence we can write

$$\begin{aligned} A(L) &= \frac{adj A(L)}{b \prod_{i=1}^{p}(1 - \lambda_i L^{-1}) \prod_{i=1}^{q}(1 - \gamma_j^{-1} L)} \\ &= \frac{adj A(L)}{b \prod_{i=1}^{q}(1 - \gamma_j^{-1} L)} \sum_{i=1}^{p} \frac{b_i}{1 - \lambda_i L^{-1}} \end{aligned}$$

Thus the model can be written

$$\prod_{i=1}^{q}(1-\gamma_j^{-1}L)y_t = \frac{adjA(L)}{b}\sum_{i=1}^{p}\frac{b_i}{1-\lambda_i L^{-1}}[-C(L)x_t+e_t]$$

or

$$y_t = G(L)y_{t-1} + \frac{1}{b}\sum_{i=1}^{p}\sum_{s=0}^{\infty} b_i\lambda_i^s L^s adjA(L)[-C(L)x_t+e_t]$$

If $D(L)$ is of order n then

$$E_t[x_{t+s}] = D_s(L)x_t$$

and so

$$y_t = G(L)y_{t-1} + H(L)x_t + K(L)e_t$$

Hence, $E_t[y_{t+1}]$ is obtained from

$$E_t[y_{t+1}] = G(L)y_t + H^*(L)x_t + K^*(L)e_t$$

We note that due to the presence of $K^*(L)e_t$, in general this is not a VAR for y_{t+1}, but a VARMA.

Substituting $E_t[y_{t+1}]$ into the original structural model gives

$$B(L)y_t + C(L)x_t + F[G(L)y_t + H^*(L)x_t + K^*(L)e_t] = e_t$$

implying that the solution can be written

$$B^*(L)y_t + C^*(L)x_t = M(L)e_t$$

To estimate the model we use

$$B(L)y_t + C(L)x_t + FE_t[y_{t+1}] = e_t$$

replacing $E_t[y_{t+1}]$ with the predicted value from the VARMA and instrumenting it due to the presence of y_t and e_t.

We have shown, therefore, that the solution is a SEM *but* with a VMA error structure.

Pre-multiplying by $M(L)^{-1}$ and then normalising would produce a VAR of infinite order.

If all of the variables are I(0), pre-multiplying by $B^*(L)^{-1}$ would yield an infinite order VMA.

If some of the variables are I(1) and there is cointegration then the Granger Representation theorem would need to be used to obtain the VMA.

It follows that it is possible to use the VAR models discussed previously even when the underlying structural model has forward-looking RE variables.

Impluse response function analysis

Consider the two equation model again. This is

$$
\begin{aligned}
y_t &= \alpha E_t[y_{t+1}] + \delta y_{t-1} + \beta x_t + e_t \\
x_t &= \theta_1 x_{t-1} + \theta_2 x_{t-2} + \varepsilon_t
\end{aligned}
$$

and has the solution for y_t:

$$
y_t = \lambda_1 x_t + \lambda_2 x_{t-1} + \lambda_3 y_{t-1} + \lambda_4 e_t
$$

Thus the VAR in (y_t, x_t) is

$$
\begin{aligned}
\begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} 1 & -\lambda_1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \lambda_3 & \lambda_2 \\ \theta_1 & \theta_2 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & -\lambda_1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \lambda_4 e_t \\ \varepsilon_t \end{bmatrix} \\
&= \begin{bmatrix} \lambda_3 + \lambda_1\theta_1 & \lambda_2 + \lambda_1\theta_2 \\ \theta_1 & \theta_2 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \lambda_4 & \lambda_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_t \\ \varepsilon_t \end{bmatrix}
\end{aligned}
$$

The error term of the VAR therefore has a recursive structure, but is upper triangular and not lower triangular, implying that causality goes from x_t to y_t, and not vice-versa. If a Cholseki decomposition is employed, the ordering of the variables in the VAR therefore needs reversing.

According to the structural solution, y_t responds instantaneously to changes in x_t; part of this is transmitted directly, and part through $E_t[y_{t+1}]$. The latter is picking up the fact that current changes in x_t have a predictable effect on future values of x_t, and these affect $E_t[y_{t+1}]$, and hence y_t. As a result, the VAR accurately captures impact effects on y_t: the impact effect from the solution is λ_1, as is the impact of ε_t on y_t in the VAR. Lagged effects will also be captured correctly by the VAR.

VAR analysis and the Lucas critique

If the Lucas critique dealt a severe blow to the SEM of the Cowles Commission, the blow to the VAR was even more severe.

To see why, suppose that one of the equations is a policy rule. For example, in the single equation model, x_t might be a policy instrument and the equation for x_t a policy rule. Suppose that at some point in the sample the rule was changed.

Consider the two equation model again. A change in the x_t equation (that was known) would affect y_t directly, as before, and through $E_t[y_{t+1}]$. The direct effect could not be distinguished from a shock to ε_t, but the change in the x_t equation would cause the relation between $E_t[y_{t+1}]$ and x_t forecasts of future values of x_t to change. This would cause the forecast for y_{t+1} to be revised. All of this would occur the moment the change is anticipated and, due the forward-looking expectations, this could be prior to the change occurring. As result, the backward-looking solution for y_t would change. This implies that the VAR would become structurally unstable.

In contrast, the original structural equation for y_t would remain unchanged. As a consequence, when it is thought that the underlying structural equations involve expectations variables, it might be better to estimate the structural model than a VAR and to carry out impulse response function analysis on a model formed from the structural equation and whatever equation for x_t is expected to be in force.