Bent E. Sørensen
January 25, 2022

1 Teaching notes on structural VARs.

1.1 Vector MA models:

1.1.1 Probability theory

The simplest (to analyze, estimation is a different matter) time series models are the moving average (MA) models:

$$x_t \;=\; \mu \;+\; u_t \;+\; B_1 u_{t-1} \;+\; ... \;+\; B_l u_{t-l} \;=\; \mu \;+\; B(L)u_t,$$

where the innovation u_t is white noise and the lag-polynomial is defined by the equation. The positive integer l is called the **order** of the MA-process. MA processes are quite easy to analyze because they are given as a sum of independent (or uncorrelated) variables. However, they may not always be easy to estimate: since it is only the x_ts that are observed, the u_ts are unobserved; i.e., latent variables.

Consider the simple scalar MA(1)-model (I leave out the mean for simplicity)

$$(*)\; x_t \;=\; u_t \;+\; b u_{t-1}\;.$$

If u_t is an independent series of $N(0, \sigma_u^2)$ variables, then this model really only says that x_t has mean zero and and autocovariance function: $\gamma(0) = (1 + b^2)\sigma_u^2;\; \gamma(1) = \gamma(-1) = b\sigma_u^2;\; \gamma(k) = 0;\; k \neq -1, 0, 1$. (Notice that I here figured out what the model says about the distribution of the observed x's. Therefore this is what the model "really" says. The autocorrelation function for the MA(1) model is trivial except for $\rho(-1) = \rho(1) = \frac{b}{1+b^2}$. By elementary calculus, it is easy to show that the MA(1) has a maximum of 0.5 for the first order autocorrelation function.

This same statistical model could also be modeled

$$(**)\ x_t \;=\; v_t \;+\; bv_{t+1}\ .$$

with v_t having the same distribution as u_t. Typically one will rule out the model (**) by assumption, but even the model (*) does not identify the parameter b uniquely. If you only observe the data then you can calculate the empirical autocovariance function and if the data are normally distributed then this is basically all the information that the data is is going to give you (apart from the mean). Assume that the autocovariance function for x_t is $\gamma_x(0) = 1$ (you can always normalize) and $\gamma_x(1) = r$ and the higher order autocorrelations are all 0. Then we find the equation

$$\frac{c}{1+c^2} \;=\; r$$

to determine c. But this is a second order polynomial, which has the solution $c = 0$ if $r = 0$ and $c = \frac{1}{2r} \pm .5\sqrt{r^{-2} - 4}$, otherwise. Notice that this in general gives one solution in the interval $[-1, 1]$ and one solution outside this interval.

Consider equation (*) again. In lag-operator notation it reads

$$x_t \;=\; (1 \;+\; bL)u_t\ ,$$

which can be inverted to

$$u_t \;=\; (1 \;+\; bL)^{-1}x_t \;=\; x_t \;-\; bx_{t-1} \;+\; b^2x_{t-2} \;+\; ...$$

It is quite obvious that this expression is not meaningful if $|b| \geq 1$ since the power term blows up. In the case where $|b| < 1$ the right hand side converges to a well defined random variable. sequences:

Definition: The scalar MA(q) model is called *invertible* if all the roots of the lag-polynomial b(L) (strictly speaking the corresponding z-transform b(z)) are outside the unit circle.

In the invertible model the innovation term u_t is a function of (the infinite past of) the observable x_ts. This is often the most sensible assumption, and in any event the invertibility assumption is almost always imposed in estimations in order for the maximum likelihood estimator to have a unique maximum (i.e. for the model to be identified). Of course you may have an economic model where the u_ts have another interpretation—there is nothing inherently wrong with the non-invertible MA model, but then you have to be careful with identification if you are estimating the model.

1.1.2 Structural interpretation of Vector MA models

The model structure

$$x_t \;=\; \mu \;+\; u_t \;+\; B_1 u_{t-1} \;+\; ... \;+\; B_l u_{t-l} \;=\; \mu \;+\; B(L) u_t,$$

is sometimes used for *structural MA-models* where each u_t vector consists of variables such as productivity shocks, monetary shocks, fiscal policy shocks, etc. Because the u-terms have mean zero (also when conditioned on any lagged observations) the term "shock" is highly appropriate in this model.

The effect on the i'th element, x_t^i, of x_t of a unit shock to j'th element of u_{t-k} is b_k^{ij} where the subscript k implies that we are looking at the i,j'th element of B_k, where we interpret B_k to be a matrix of 0's if $k > l$. Another way of expression this is

$$\frac{\partial x_t^i}{\partial u_{t-k}^j} = b_k^{ij} .$$

Notice that the notation can vary a lot for the ijth element of matrix B_k—some authors prefer $b_{k,ij}$ or other variants. We assume that the u_t terms are stationary and independent and the stationarity assumption implies that we can write the previous equation as

$$\frac{\partial x_{t+k}^i}{\partial u_t^j} = b_k^{ij} \;;$$

i.e., as an equation that show the predicted effect of current shocks. If you plot b_k^{ij} as function of k, you get the so-called impulse response function. (The shock to u_t^j is the "impulse" and the predicted response of x_{t+k}^i is the "response.") So, if x_t is, say, 3 dimensional the structural MA-model implies 3 times n impulse response function if the B_k matrices have dimension $3 \times n$. In practice, the matrices are most often chosen to be quadratic, with the number of endogenous variables equal to the number of "impulses."

Impulse response functions are great tools to analyze the workings of models. It variable 1, say, is a productivity shock, the impulse response functions for the x-variables related to this shock very clearly show how the productivity shock reverberates through the economy. In particular for engineers working with linear systems, this is a useful tool because typically the can actually select an "impulse" (feed some input into the physical system) and measure the response. In macroeconomics we do not have that ability although in isolated case you might be able to observe natural experiments that act as exogenous impulses.

Vector MA-models are very convenient to use for forecasting. Denote the expectation of x_t conditional on $x_{t-h}, x_{t-h-1}, ...$ (The notation $E(x_t|I_{t-h})$ where "I" means "information set" is also used sometimes.) For the Vector MA

$$x_{t|t-h} = B_h u_{t-h} + ...B_k u_{t-k} \ ,$$

interpreted as 0, if $h > k$. The "h-period forecast error" $x_t - x_{t|t-h}$ is then

$$x_t - x_{t|t-h} = u_t + B_1 u_{t-1} + ...B_{h-1} u_{t-h+1} \ ,$$

which, of course, converges to x_t is $h \to \infty$.

A related tool is *variance decompositions*: If the variance-covariance matrix Σ_u is diagonal and u has p elements:

$$\Sigma_u \ = \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ & \ddots & \\ 0 & & \sigma_p^2 \end{pmatrix}$$

then

$$Var(x_t^i - x_{t|t-h}^i) = \Sigma_{l=1}^p \ (\sigma_l^2 \, \Sigma_{k=0}^{h-1} \, (b_k^{i,l})^2 \ )$$

The contribution from innovation j to the variance of $x_t^i - x_{t|t-h}^i$ is therefore

$$f_j^i = \frac{\sigma_j^2 \, \Sigma_{k=0}^{h-1} \ (b_k^{i,j})^2}{\Sigma_{l=1}^p \ \sigma_l^2 \ \Sigma_{k=0}^{h-1} \, (b_k^{i,l})^2} \ .$$

(Here, interpret B_0 as the identity matrix. In some applications B_0 will be not be the identity matrix.) Notice that $\Sigma_j f_j^i = 1$ and that f_j^i is the fraction of the variance of the h-period ahead forecast error of variable i that is explained (or caused) by the innovations to shock j. This can be very a very useful way to describe, for example, which variable is "most important" in generation the business cycle because the variance of output (or output growth) is a reasonable measure of the "size" of the business (or any other) cycle, and the proportion of that explained by, say, productivity shocks, tells us exactly how important productivity shocks are for the business cycle on average, at a given forecast horizon. It is possible that, say, demand shocks are more important in the short run and supply shocks in the large run—if you have the economy described by a linear vector-MA with independent innovations, you can answer such a question.

1.2 AR models:

The most commonly used type of time series models are the auto regressive (AR) models. In vector form it is usually denoted a VAR process:

$$x_t \;=\; \mu \;+\; A_1 x_{t-1} \;+\; \ldots \;+\; A_k x_{t-k} \;+\; u_t \;,$$

where the innovation u_t is a martingale difference sequence (or white noise). Here k is a positive integer called the order of the AR-process. Such a process is usually referred to as an VAR(k) process.

If a finite order VAR-model is invertible, then if $x_t = A(L)u_t$ (where I suppress the constant term for convenience) then x_t satisfies the (infinite order) Vector MA-model

$$x_t = A^{-1}(L)u_t \;.$$

So for invertible VAR-models we can use the methods outlined above to calculation impulse response functions and variance decompositions.

In practice, if you want to plot impulse response functions (IRFs), it is much easier. Just as in the scalar case, if

$$x_t - Ax_{t-1} = u_t \;.$$

you can invert the lag polynomial (the proof is the same as in the scalar case) and get

$$x_t = u_t + Au_{t-1} + A^2 u_{t-2} + ... \ .$$

assuming the process is stable; i.e., that the right hand convergest. The impuls response functions are then contained in $I, A, A^2, ..$ when you multiply each by, say, $(1, 0, ..., 0)'$ for the the first error. For example, the impulse response of variable i from an impulse to variable j after 3 periods is the (i, j)th element of A^3.
But you don't even need to know this. You can just do a recursion. Say that x_{t+k} contains a vector of impact of a unit shock to the first error term. You do $x_0 = (1, ...0)'$. Then $x_1 = A\,x_0$ and so on $x_t = Ax_{t-1}$, which you just do in a loop and collect the values. You can always do this. The eigenvalue of A doesn't even have to be smaller then one for this to work. And you can do it for any VAR. For example, if

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + u_t \ ,$$

you can again do the recureion, for say the first innovation. We have $x_0 = (1, ...0)'$, $x_1 = A_1 x_0$ and

$$x_2 = A_1 x_1 + A_2 x_0$$

and so on. This is the same loop you would use to generate the VAR in Monte Carlo, except you do not add error terms beyond the first one. You can generalize this to a VARMA, which I may ask you to do for simple example in a home work.

In practice, empirical economists often estimate a VAR-model (MA-models are hard to estimate). Say, you have a time series for $x_t' = y_t, p_t, m_t$, where the variables could be output growth, productivity growth, and money growth, respectively. Then you can estimate a VAR and invert this to perform variance decompositions and impulse response functions etc., but notice that this only makes sense if the error terms can be *interpreted as exogenous shocks* which is often quite a leap of faith.

This brings me to the last topic of identification. As for the MA-processes, "VAR-modeling" can is used in a much more specific sense where the innovations (the error terms) are interpreted at shock to particular "driving" processes (such as productivity shocks). I will refer to this as *structural VAR* modeling which originally suggested by Christopher Sims as an alternative to the big Keynesian macro econometric models. The basic philosophy was that the usual macro models only can be identified under extensive a priori restrictions (in order for individual equations to be identified it is usually assumed that a given endogenous variable only depends on a limited number of other endogenous variables). Sims finds many of these a priori restrictions "incredible" and suggest that one starts with an unrestricted VAR model instead. Sims' article is called "Macroeconomics and Reality" and is reprinted various places but there are by now many surveys and books about structural VAR modelling.

Consider the case of the 1-order VAR for a p-dimensional vector. What you can estimate from the data is $p \times p + p(p+1)/2$ parameters (ignoring the vector of constants for simplicity), namely the matrix A and the different parameters of the variance matrix. To make this explicit, we can write

$$x_t = A\,x_{t-1} + \Sigma_u v_t \ , \ \ (*)$$

where v_t is a vector white noise process and Σ_u is lower triangular (you can always choose a Cholesky triangular matrix for for square root of the variance matrix).

The "Cowles commission" way of identifying models was to assume (typically) that the error matrix was diagonal and the economic theorizing would lead to assumptions on which variables were function of each other, e.g., output depending on money and productivity but, e.g., these independent of each other. This could be written as

$$A_0 x_t = A\,x_{t-1} + v_t \ .$$

Here you might let the elements of v_t have separate variances but then you would need to put enough restrictions on A_0 to not have more than $p \times p + p(p+1)/2$ parameters. Another way of putting it is what you cannot estimate a model

$$A_0 x_t = A\,x_{t-1} + \Sigma_u v_t \ ,$$

because this is equivalent to

$$x_t = A_0^{-1} A\,x_{t-1} + A_0^{-1} \Sigma_u v_t \ ,$$

and one cannot untangle both A_0 and A from an estimated $p \times p$ coefficient to x_{t-1}. Many people in the VAR tradition choose to estimate the model in the form (*), assuming the estimated error terms corresponds to innovations to exogenous driving variables. They would not assume that each error term is exogenous but, if

$$\Sigma_u = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ \sigma_{21} & \sigma_2^2 & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix},$$

that u_1 is an exogenous innovation (shock, impulse, etc.), that $u_2 - \sigma_{21} u_1$ is an exogenous innovation, and $u_3 - \sigma_{31} u_1 - \sigma_{32} u_2$ is an exogenous innovation. So, e.g., σ_{21} allows variable number 2, in the example: money, to be a function of variable number 1 output, within the current period t, and variable number 3 (productivity) to be a function of the other two innovations within the current period. (Note that you need to modify the impulse response functions and variance decompositions to allow for non-diagonal covariance matrix, but that should be straight-forward.) So you might hear an economist say that he or she chooses an "ordering;" here, they might choose to order the x vector as $x_t' = p_t, m_t, y_t$ such that money within the current period can react to productivity shocks (but not the other way around) and output to the other two (but not the other way around)—this is basically the idea in RBC models. (Of course, in this case here this simply corresponds the choosing an upper triangular Cholesky matrix, but people prefer to stick with a low triangular and talk about "orderings" of the variables.

This might all make perfect sense. (Although, I do not believe that monetary policy would ever be a function of just productivity shocks.) However, there seems to be a very unfortunate tendency in this literature to not involve much economic argumentation. While people in the Cowles tradition (whatever problems they might have had otherwise) at least thought hard about what restriction to put on A_0, you might hear "structural VAR" modelers just spend one line, stating something like "I choose an ordering and the impulse response functions look about the same with other orderings..." which basically means (in my view) that they have given up on economics.

A second critical issue is that if you involve such variables as money then you are assuming (sadly, this is again often done implicitly, without discussion) that output is ONLY a function of these variables or at least that no variable that might affect money is correlated within the current period output (only productivity shocks are allowed to impact money in the current period). I cannot see that happen. Also, to use this model (or variations of it) you have to assume that you can measure productivity shocks in the sense of increased knowledge/technique/blue prints/etc. and something like Solow residuals, which are often applied, are not, in my opinion, very good measures of this. In general, one has to make "brave" assumption to represent the whole economy as a low order VAR. (Higher order VAR's might be better, but you then end up having to estimate a large amount of parameters which likely will lead to imprecise estimates in the type of samples available to macroeconomists.) However, in other applications, the methodology might work fine. What you have to do, is to argue more in terms of economics for whatever restrictions you use whether you put restriction on A_0 on Σ or some other combination. I outline an alternative methodology below (it is just an outline, see the original paper if you want more details):

A clever alternative way of identifying models was suggested by Blanchard and Quah. They argued that supply shocks would affect output forever but demand shocks only temporary. Writing (say) a two-dimensional process as

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} + \begin{pmatrix} c_1^{11} & c_1^{12} \\ c_1^{21} & c_1^{22} \end{pmatrix} \begin{pmatrix} u_{1t-1} \\ u_{2t-1} \end{pmatrix} + \begin{pmatrix} c_2^{11} & c_2^{12} \\ c_2^{21} & c_2^{22} \end{pmatrix} \begin{pmatrix} u_{1t-2} \\ u_{2t-2} \end{pmatrix} + \ldots$$

then the hypothesis that demand shocks (say variable 2) has no long run effects can be implemented as the restriction that the sum of the impulses with respect to variable 2 is zero (it is here essential that the variables are written in growth rates). I won't go through the model in details but verify for yourself that

$$\frac{\partial x_{t+h}^i}{\partial u_t^j} = \Sigma_{k=0}^h b_k^{ij} \ ;$$

which simply follows from the definition of of the impulse response function—here applied to the variable Δx_t—and the fact that

$$x_{t+h} = \Delta x_{t+h} + \Delta x_{t+h-1} + ... + \Delta x_{t+1} + x_t \ .$$

The idea is clever and illustrates that sometimes theoretical considerations can result in non-obvious ways of identifying the model. For empirical research, the infinite sum of the impulse responses are likely to not be robustly estimated, so this approach has not had a significant influence on empirical practise although it pops up in research once in a while.