ECONOMICS 7395, Spring, 2005
Bent E. Sørensen
March 1, 2005

1 Granger Causality.

1.1 Linear Prediction.

Assume that you want to predict the value of y_{t+k} based on the information set $\mathcal{F}_t$. How do you do that in the best possible way? This depends on your cost of making a wrong prediction, so if you have a formal model for your cost of making an error of a given size, then you should minimize that function (in statistics this is usually called a *loss function*). In econometrics it is usual to choose to minimize the mean square error (MSE) of the forecast, i.e.

$$\min E\{(y_{t+k} - \hat{y}_{t+k})^2\}$$

where $\hat{y}_{t+k}$ is the predictor of y_{t+k}. One can show that the conditional mean $E\{y_{t+k}|\mathcal{F}_t\}$ is the best mean square predictor. If the information set $\mathcal{F}_t$ consists of a vector of observations z_t (which would usually include $y_t, y_{t-1}, ..., y_1$), then the conditional mean in the case of normally distributed variables is linear (as we know). In the case where the observations are not normally distributed the conditional mean is not a linear function of the conditioning varibles, so if you can find the true conditional mean you may want to do that, however, timeseries analysis is, as mentioned, mostly in the 2nd order tradition, so often people use the *best linear predictor* rather than the conditional mean. You find the best linear predictor as that linear function of the conditioning variables that would give you the conditional mean if the data had been normally distributed.

Assume that your data are described by a VAR(2) model:

$$y_t \;=\; \mu \;+\; A_1 y_{t-1} \;+\; A_2 y_{t-2} \;+\; u_t \;.$$

What would be the best (linear) forecast of y_{t+1} based on $y_1, ..., y_t$? Obviously,

$$\hat{y}_{t+1} \;=\; \mu \;+\; A_1 y_t \;+\; A_2 y_{t-1}.$$

It turns out that we can iterate this formula to find

$$\hat{y}_{t+k} \;=\; \mu \;+\; A_1 \hat{y}_{t+k-1} \;+\; A_2 \hat{y}_{t+k-2}.$$

for any k. Another approach would be to reformulate the model as a higher dimensional VAR(1) system, since it is easy to see that

$$\hat{y}_{t+k} \;=\; (I + A + ... + A^{k-1})\mu \;+\; A^k y_t \;,$$

in this case. (Note that the best linear predictor in the stable case converges (for $k \to \infty$) to the unconditional mean of the process).

For models with MA components things are harder. Recall that one can write the ARMA model as a high order VAR(1) (the state-space representation), so one can use the formula above, but the complication is that even at time t one does not know u_t. The Kalman filter does however, as a byproduct, give you the best guess of $u_t, u_{t-1}, ...$, (namely as part of $\alpha_{t|t}$), so you can use the Kalman filter to generate $\alpha_{t|t}$ and then you can use the formula above. For more elaborations, see Harvey (1989).

1.2 Granger Causality.

Assume that the information set $\mathcal{F}_t$ has the form $(x_t, z_t, x_{t-1}, z_{t-1}, ..., x_1, z_1)$, where x_t and z_t are vectors (that includes scalars of course) and z_t usually will include y_t and z_t may or may not include other variables than y_t.

Definition: We say that x_t is Granger causal for y_t wrt. $\mathcal{F}_t$ if the variance of the optimal linear predictor of y_{t+h} based on $\mathcal{F}_t$ has smaller variance than the optimal linear predictor of y_{t+h} based on $z_t, z_{t-1}, ...$ - for any h. In other word x_t is Granger causal for y_t if x_t helps predict y_t at some stage in the future.

Often you will have that x_t Granger causes y_t and y_t Granger causes x_t. In this case we talk about a *feedback system.* Most economists will interpret a feedback system as simply showing that the variables are related (or rather they do not interpret the feedback system).

Sometimes econometrians use the shorter terms "causes" as shorthand for "Granger causes". You should notice, however, that Granger causality is not causality in a deep sense of the word. It just talk about linear prediction, and it only has "teeth" if one thing happens before another. (In other words if we only find Granger causality in one direction). In economics you may often have that all variables in the economy reacts to some unmodeled factor (the Gulf war) and if the response of x_t and y_t is staggered in time you will see Granger causality even though the real causality is different. There is nothing we can do about that (unless you can experiment with the economy) - Granger causality measures whether one thing happens before another thing and helps predict it - and nothing else. Of course we all secretly hope that it partly catches some "real" causality in the process. In any event, you should try and use the full term Granger causality if if is not obvious what you are refering to

The definition of Granger causality did not mention anything about possible instantaneous correlation between x_t and y_t. If the innovation to y_t and the innovation to x_t are correlated we say there is *instantaneous causality.* You will usually (or at least often) find instantaneous correlation between two time series, but since the causality (in the "real" sense) can go either way, one usually does not test for instantaneous correlation. However, if you do find Granger causality in only one direction you may feel that the case for "real" causality is stronger if there is no instantaneous causality, because then the innovations to each series can be thought of as actually being generated from this particular series rather than part of some vector innovations to the vector system. Of course, if your data is ampled with a long sampling period, for example annually, then you would have to explain why one variable would only cause the other after such a long lag (you may have a story for that or you may not, depending on your application).

Granger causality is particularly easy to deal with in VAR models. Assume that our data can be described by the model

$$\begin{bmatrix} y_t \\ z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} A_{11}^1 & A_{12}^1 & A_{13}^1 \\ A_{21}^1 & A_{22}^1 & A_{23}^1 \\ A_{31}^1 & A_{32}^1 & A_{33}^1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \\ x_{t-1} \end{bmatrix} + \ldots + \begin{bmatrix} A_{11}^k & A_{12}^k & A_{13}^k \\ A_{21}^k & A_{22}^k & A_{23}^k \\ A_{31}^k & A_{32}^k & A_{33}^k \end{bmatrix} \begin{bmatrix} y_{t-k} \\ z_{t-k} \\ x_{t-k} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

Also assume that

$$\Sigma_u = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12}' & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{13}' & \Sigma_{23}' & \Sigma_{33} \end{bmatrix}.$$

This model is a totally general VAR-model - only the data vectors has been partitioned in 3 subvectors - the y_t and the x_t vectors between which we will test for causality and the z_t vector (which may be empty) which we condition on.

In this model it is clear (convince yourself!) that x_t does *not* Granger cause y_t with respect to the information set generated by z_t if either $A_{13}^i = 0$ and $A_{23}^i = 0;\ i = 1, ..., k$ or $A_{13}^i = 0$ and $A_{12}^i = 0;\ i = 1, ..., k$. Note that this is the way you will test for Granger causality. Usually you will use the VAR approach if you have an econometric hypothesis of interest that states that x_t Granger causes y_t but y_t does not Granger cause x_t. Sims (1972) is a paper that became very famous because it showed that money Granger causes output, but output does not Granger cause money. (This was in the old old days when people still took monetarism seriously, and here was a test that could tell whether the Keynesians or the monetarists were right!!). Later Sims showed that this conclusion did not hold if interest rates were included in the system. This also shows the major drawback of the Granger causality test - namely the dependence on the right choice of the conditioning set. In reality one can never be sure that the conditioning set has been choosen large enough (and in short macro-economic series one is forced to choose a low dimension for the VAR model), but the test is still a useful (although not perfect) test.

I think that the Granger causality tests are most useful in situations where one is willing to consider 2-dimensional systems. If the data are reasonably well described by a 2-dimensional system ("no z_t variables") the Granger causality concept is most straightforward to think about and also to test. By the way, be aware that there are special problems with testing for Granger causality in co-integrated relations (see Toda and Phillips (1991)).

In summary, Granger causality tests are a useful tool to have in your toolbox, but they should be used with care. It will very often be hard to find any clear conclusions unless the data can be described by a simple "2-dimensional" system (since the test may be between 2 vectors the system may not be 2-dimensional is the usual sense), and another potentially serious problem may be the choice of sampling period: a long sampling period may hide the causality whereas for example VAR-systems for monthly data may give you serious measurement errors (e.g. due to seasonal adjustment procedures).

Extra reference:

Toda, H.Y. and P.C.B. Phillips (1994) : "Vector Autoregressions and Causality: A Theoretical Overview and Simulation Study", *Econometric Reviews* 13, 259-285.