VAR'S - THEORY

CONSIDER BIVARIATE GRANGER CAUSALITY TESTS:

IN A BIVARIATE VAR DESCRIBING $x$ AND $y$, $y$ DOES NOT
GRANGER-cause $x$ IF THE COEFFICIENT MATRICES $\phi_j$ ARE ALL
LOWER TRIGONAL FOR ALL $j$:

$$ \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_1^{(x)} & 0 \\ \phi_1^{(y)} & \phi_2^{(y)} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \phi_1^{(x)} & 0 \\ \phi_2^{(x)} & \phi_2^{(x)} \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \cdots + \begin{bmatrix} \phi_1^{(y)} & 0 \\ \phi_2^{(y)} & \phi_2^{(y)} \end{bmatrix} \begin{bmatrix} x_{t-p} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} $$

CLEARLY THE ZERO IN THE UPPER RIGHT CORNER IMPLIES THAT
OPTIMAL ONE-PERIOD AHEAD FORECASTS OF $x$ DO NOT DEPEND ON
LAGGED $y$.

NOTE: THE ZERO (OR A ZERO RESTRICTION) IMPLIES CONFIDENCE IN A
CASUAL RELATIONSHIP.

NEXT: ASSUME $x_t$ IS COVARIANCE STATIONARY OF LENGTH $n$.
BY THE WOLD DECOMPOSITION THEOREM (1938), $x_t$
CAN BE REPRESENTED BY

$$ x_t = \mu + \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j} $$

WHERE $\epsilon_t$ IS A WHITE NOISE ERROR PROCESS, $\phi_0 = 1,$
AND $\sum_{j=0}^{\infty} \phi_j < \infty$. 
This can be generalized if \( x_t \) is a covariance stationary vector of length \( n \), then \( x_t \) has a moving average representation by the Wold decomposition theorem.

If this is true, then certainly \( x_t \) can be written as:

\[(1) \quad x_t = A(L)x_t + B(L)\varepsilon_t \]

where \( A(L) \) and \( B(L) \) are matrices of polynomials in the lag operator, and the coefficient on \( L^0 \) is zero for all elements of \( A(L) \).

If \( B(L) \) is invertible, then (1) can be written as a VAR:

\[(2) \quad x_t = C^u(L)x_t + \varepsilon_t \]

Note: There are only lagged values of \( x \) on the RHS.

The coefficient on \( L^0 \) is zero for all elements of \( C^u(L) \).

In theory the elements of \( C^u \) are polynomials that may be of infinite degree, but in practice the lagged is truncated to some finite length \( m \).

Hence:

\[(3) \quad x_t = C(L)x_t + \varepsilon_t = \sum_{s=1}^{M} C_s x_{t-s} + \varepsilon_t \]

\( C_s \) is a \( n \times n \) coefficient matrix associated with the \( s \)th lag.

(The VAR has \( n \) equations in which each element of \( x_t \) is a linear function of its own lagged values and the lagged values of all the other variables.)
THE MODEL (3) CAN BE CONSISTENTLY AND EFFICIENTLY ESTIMATED BY APPLYING OLS TO EACH EQUATION.

IT IS THE EQUIVALENT MOVING AVERAGE REPRESENTATION THAT IS USED TO EXAMINE THE DYNAMIC RESPONSE TO SHOCKS.

IF \([I - C(L)]\) HAS AN INVERSE, THEN (3) CAN BE WRITTEN AS

\[
(4) \quad x_t = [I - C(L)]^{-1} \epsilon_t = D(L) \epsilon_t = \sum_{s=0}^{\infty} D_s \epsilon_{t-s}
\]

WHERE \(D_s\) IS \(N \times N\).

FROM (4) IF \(\epsilon\) NO CONTEMPORANEOUS CORRELATION AMONG THE ELEMENTS OF \(\epsilon\), \(d_{ij}\), THE \(i,j\) TH ELEMENT OF \(D_s\) TRACES OUT THE RESPONSE OF THE \(i\) TH ELEMENT OF \(x_t\) OVER TIME(S) TO AN INNOVATION IN THE \(j\) TH ELEMENT OF \(x_t\).

NORMALLY THESE "INNOVATIONS" WILL BE NON-CORRELATED, THEREFORE IT IS IMPOSSIBLE TO UNIQUELY DECOMPOSE THE VARIANCE OF EACH VARIABLE INTO COMPONENTS ACCOUNTED FOR BY EACH INNOVATION.

SOLUTION REQUIRES A TANGENTIAL ORTHOGONALIZATION TRANSFORMATION TO THE VECTOR OF INNOVATIONS, \(\epsilon_t\), TO PRODUCE A NEW VECTOR OF INNOVATIONS HAVING THE IDENTITY MATRIX AS ITS COVARIANCE MATRIX.

WE WANT A LOWER TRIANGULAR MATRIX \(G\) SO THAT

\[
(5) \quad \epsilon_t = G \epsilon_t \quad \text{AND THE}
\]

COVARIANCE MATRIX OF \(\epsilon_t\) IS THE IDENTITY MATRIX.
Once we find $G$,

\[(5) \quad \mathbf{x}_k = DCLG^{-1}\mathbf{e}_k = FCL\mathbf{v}_k = \sum_{s=0}^{\infty} F_s \mathbf{v}_{k+s}\]

The transformation $G$ is not unique and essentially imposes a particular causal ordering on the variables.

$G$ is lower triangular, so $G^{-1}$ is upper triangular.

$G^{-1}$ creates the zero restrictions that imply causality.

**Triangular Decomposition**

When we perform Gaussian elimination on a matrix $A$, in the $k^{th}$ step, we form $A^{(k)}$ from $A^k$ by subtracting multiples of the $k^{th}$ row from rows $k+1$ to $n$.

This can be written $A^{(k+1)} = M_k A^{(k)}$.

$M_k$ is nonsingular, so it follows that

$$A^{(k)} = M_k^{-1} A^{(k+1)}.$$ 

By induction

$$A^{(1)} = M_1^{-1} M_2^{-1} \ldots M_{n-1}^{-1} A^{(n)}$$

It can be shown that

$$M_1^{-1} M_2^{-1} \ldots M_{n-1}^{-1} = L$$

is lower triangular with unitary diagonal elements.

Put $A^{(1)} = A$, $A^{(n)}$ is upper triangular, so

$$A = LU, \quad u = A^{(n)}$$
SO GAUSSIAN ELIMINATION IS EQUIVALENT TO A TRIANGULAR DECOMPOSITION.

IF A IS SYMMETRIC, THEN THIS DECOMPOSITION DEISTRYS THE SYMMETRY. FOR EXAMPLE,

\[
\begin{pmatrix}
16 & 4 & 8 \\
4 & 5 & -4 \\
8 & 4 & 22
\end{pmatrix}
= \begin{pmatrix}
1 \\
0.5 & 1 \\
0.5 & -1.5 & 1
\end{pmatrix}
\begin{pmatrix}
16 & 4 & 8 \\
4 & -4 \\
9
\end{pmatrix}
\]

A SLIGHTLY DIFFERENT FORM OF DECOMPOSITION IS

\[ A = LDL^T \]

USING THE SAME EXAMPLE,

\[
\begin{pmatrix}
16 & 4 & 8 \\
4 & 5 & -4 \\
8 & 4 & 22
\end{pmatrix}
= \begin{pmatrix}
1 \\
0.5 & 1 \\
0.5 & -1.5 & 1
\end{pmatrix}
\begin{pmatrix}
16 & 4 \\
4 & 9 \\
1 & -1.5
\end{pmatrix}
\]

WE COULD CONSTRUCT A MATRIX \( D^{1/2} = \begin{pmatrix} \sqrt{2} \end{pmatrix} \)

SO THAT \( D = D^{1/2} D^{1/2} \)

THIS ALLOWS \( A = LDLD^T = LDL^{1/2} D^{1/2} L^T = \tilde{L} \tilde{L}^T \)

WHEN \( \tilde{L} = LD^{1/2} \)

THE DECOMPOSITION \( A = \tilde{L} \tilde{L}^T \) IS THE CHOLESKI DECOMPOSITION.

(IF A IS NOT P.D., \( L \) CONTAINS COMPLEX ELEMENTS)

\[
\begin{pmatrix}
16 & 4 & 8 \\
4 & 5 & -4 \\
8 & -4 & 22
\end{pmatrix}
= \begin{pmatrix}
4 & 1 & 2 \\
1 & 2 & -3 \\
2 & -3 & 3
\end{pmatrix}
\]
SIDE NOTE: THE CHOLESKI DECOMPOSITION TAKES ABOUT $\frac{1}{2}$ THE COMPUTATIONS THAT THE TRIANGULAR (GAUSSIAN) DECOMPOSITION.

THE FORMULAE FOR THE CHOLESKI DECOMPOSITION ARE:

$$L_{ii} = (a_{ii} - \sum_{k=1}^{i-1} L_{ik}^2)^{\frac{1}{2}}$$

$$L_{ij} = \frac{a_{ij} - \sum_{k=1}^{i-1} L_{ik}L_{jk}}{L_{jj}}, \quad j < i$$

**Notice:** Reordering the matrix rows or columns results in a different decomposition.

WHEN WE TRACE OUT THE RESPONSE OF AN ELEMENT OF $X_t$ TO AN ORTHOGONALIZED INNOVATION, THIS IS THE IMPULSE RESPONSE AND REMAINS UNIQUE.


THIS IS THE VARIANCE DECOMPOSITION.

Most researchers argue that only variance decomposition results that are robust to re-ordering can be considered significant.
Kang (1985) has argued that (VAR) Granger tests are quite sensitive to method of detrending.

It appears that ordering is less important for high frequency data, less for temporally aggregated data.

Lag lengths may also be important.