Part II — Granger Causality

Granger causality (Granger 1969) is an important concept in econometrics.

Let \( \mathcal{L}(U|V) \) denote the conditional distribution of \( U \) given \( V \).

**Definition.** Time series \( Z_t \) is said to Granger cause time series \( Y_t \) if

\[
\mathcal{L}(Y_t|Y_{t-1}, Z_{t-1}, Y_{t-2}, Z_{t-2}, \ldots) \neq \mathcal{L}(Y_t|Y_{t-1}, Y_{t-2}, \ldots). 
\]

\( Z_t \) is said to Granger cause in mean \( Y_t \) if

\[
E(Y_t|Y_{t-1}, Z_{t-1}, Y_{t-2}, Z_{t-2}, \ldots) \neq E(Y_t|Y_{t-1}, Y_{t-2}, \ldots). 
\]
Check Granger causality in mean by VAR: Let \( X_t = (Z_t, Y_t)' \) and \( X_t \sim \text{AR}(p) \),

\[
X_t = c + \sum_{j=1}^{p} A_j X_{t-j} + \varepsilon_t,
\]

where \( \varepsilon_t \sim \text{WN}(0, \Sigma_\varepsilon) \).

If in addition \( X_t \) is a Gaussian process, then

\[
E(Y_t|Y_{t-1}, Z_{t-1}, Y_{t-2}, Z_{t-2}, \ldots) = c_2 + \sum_{j=1}^{p} (a_{21}(j) z_{t-j} + a_{22}(j) y_{t-j}) + \varepsilon_{t,2}
\]

Thus \( Y_t \) is Granger causes in mean by \( Z_t \) if \( a_{21}(j) \neq 0 \) at least for one \( j \geq 1 \).
**Granger causality test** is to test the hypothesis

\[ H_0 : a_{21}^{(1)} = \cdots = a_{21}^{(p)} = 0 \]

When \( H_0 \) is rejected, \( Z_t \) is regarded as Granger causing \( Y_t \).

To test hypothesis \( H_0 \), let

\[ \text{RSS} = \sum_{t=p+1}^{T} \left\| X_t - \hat{c} - \sum_{j=1}^{p} \hat{A}_j X_{t-j} \right\|^2 \]

where \( \hat{c}, \hat{A}_1, \cdots, \hat{A}_p \) are the LSE for the parameters. Let \( \text{RSS}_r \) be the residual sum squares under \( H_0 \). The contribution of \( Z_t \) to \( Y_t \) is then measured by the standard \( F \)-test statistic:

\[ F = \frac{(\text{RSS}_r - \text{RSS})/p}{\text{RSS}/(2T - 4p - 2)}. \]

Under \( H_0 \), \( pF \sim \chi^2_p \), or \( F \sim F_{p, 2T-4p-2} \) asymptotically.
Let $\hat{\Sigma}_\varepsilon$ be the estimated covariance matrix of $\varepsilon_t$ under the full VAR($p$) model, and $\hat{\Sigma}_{\varepsilon,r}$ be such an estimator with the restriction $a_{21}^{(1)} = \cdots = a_{21}^{(p)} = 0$.

Then the likelihood ratio test statistic is

$$(T - p) \log \left( \frac{|\hat{\Sigma}_{\varepsilon,r}|}{|\hat{\Sigma}_\varepsilon|} \right).$$

which is asymptotically $\chi^2_p$ under $H_0$. 
**Instantaneous causality.** For \( X_t = (Z_t, Y_t)' \), and

\[
X_t = c + \sum_{j=1}^{p} A_j X_{t-j} + \epsilon_t,
\]

where \( \epsilon_t \sim WN(0, \Sigma_\epsilon) \). If the off-diagonal element \( \Sigma_\epsilon \) is not zero, then

\[
\text{Corr}(Z_t, Y_t|X_{t-1}, \ldots, X_{t-p}) \neq 0.
\]

In this case, \( Z_t \) and \( Y_t \) have instantaneous Granger causality.

The instantaneous Granger causality can be tested with the null hypothesis \( H_0 : \sigma_{21} = 0 \), where \( \sigma_{21} \) denotes the off-diagonal element of \( \Sigma_\epsilon \). Under \( H_0 \),

\[
\text{Cov}(Z_t, Y_t|X_{t-1}, \ldots, X_{t-p}) = 0.
\]

A Wald test statistic can be constructed based on the estimate for \( \sigma_{21} \) and its asymptotic normality.
The function `causality` in the package `vars` implements the $F$-test for Granger causality and the Wald test for instantaneous causality. It also offers the option to evaluate the $P$-values of the tests by a wild bootstrap method instead of the asymptotic distributions.

Let $X_{12}$ be a data matrix containing the daily log returns of FTSE 100 and FTSE Midcap as its two columns. We run

```r
> m12 = VAR(X12, ic="SC")
> causality(m12)
```

**H0:** FTSE100 do not Granger-cause FTSE.MidCap

$F$-Test = 13.3076, df1 = 1, df2 = 490, p-value = 0.0002927

**H0:** No instantaneous causality between FTSE100 and FTSE.MidCap

Chi-squared = 116.7708, df = 1, p-value < 2.2e-16
Since the null hypothesis of no Granger causality is rejected with the $P$-value 0.0003, there is significant evidence indicating that FTSE 100 Granger causes FTSE MidCap.

On the other hand, the null hypothesis of no instantaneous causality is rejected with the $P$-value 0. Hence, there exists instantaneous causality between FTSE 100 and FTSE MidCap.

Applying the above test to different data sets, we also find that FTSE 100 Granger causes FTSE SmallCap with the $P$-value 0.004, and that there exists no significant evidence indicating that FTSE SmallCap Granger causes FTSE 100 as the $P$-value is 0.914.
Impulse response functions: another way to investigate the effect of a change in one component series on the other components.

A causal VAR process can be written as VMA(∞):

\[ X_t = c + \varepsilon_t + \sum_{k=1}^{\infty} B_k \varepsilon_{t-k}, \]

with the elements of \( B_j \) decaying to 0 exponentially as \( j \to \infty \).

An increase in the first component of \( X_t \) by a unit at time \( t \) can only be caused by the same increase in the first component of \( \varepsilon_t \).

The impact of such an increase on the \( i \)-th component at time \( t+k \), i.e. the change in \( X_{t+k,i} \), is \( b_{i1}^{(k)} \) – the \((i,1)\)-th element of \( B_k \).
Hence

the \((i, j)\)-th element of \(B_k\) is the impulse response of \(X_{t+k,i}\),
the \(i\)-th component at the \(k\) units of time ahead, from one
unit of extra shock in the \(j\)-component of \(X_t\).

The above analysis is valid only if the components of \(\varepsilon_t\) are inde-
dependent with each other, i.e. no Granger instantaneous causality
among the components of \(X_t\).

To alleviate the correlations among the components of \(\varepsilon_t\), the
following transformation is applied to the model:

\[
\varepsilon_t = \Psi_0 e_t, \quad \text{for} \quad \Psi_0 \Psi_0' = \Sigma_{\varepsilon}.
\]
Then $\text{Var}(e_t) = I_d$, i.e. the components of $e_t$ are uncorrelated with each other, and the VMA($\infty$) model can be written as

$$X_t = c + \Psi_0 e_t + \sum_{k=1}^{\infty} \Psi_k e_{t-k},$$

where $\Psi_k = B_k \Psi_0$ for $k \geq 1$.

The matrices $\Psi_0, \Psi_1, \cdots$ are called the impulse response functions.

More precisely the plot of $\psi_{ij}^{(k)}$ against $k$, for $k = 0, 1, \cdots$, is called the response function of $X_{ti}$ to the impulse on the $j$-th component of $e_t$, where $\psi_{ij}^{(k)}$ denotes the $(i,j)$-th element of $\Psi_k$.

The computation of $\Psi_k$ can be carried out using R-function Psi in the package vars.
Unfortunately $\psi_{ij}^{(k)}$ measures the responses to the change in the component of $e_t$ instead of $X_t$ — an innate problem when the components of $X_t$ are dependent with each other, as then it may be impossible to isolate the change of one component from those of the others.

Furthermore, the definition of $e_t$ is not unique, as $\Psi_0$ can be replaced by $\Psi_0 H$ for any $d \times d$ orthogonal matrix $H$.

Most packages, including vars, adopt the Cholesky decomposition of $\Sigma_\epsilon$, leading to a lower triangular $\Psi_0$. 
To fit vector AR models to the returns of these two price series,

```r
> fitX=VAR(X, ic="AIC"); summary(fitX)
```

Estimation results for equation S.P500:
```
=======================================
S.P500 = S.P500.l1 + JP.Morgan.l1 + const

        Estimate Std. Error t value Pr(>|t|)
S.P500.l1   -0.07625   0.06887  -1.107    0.269
JP.Morgan.l1 -0.02099   0.03720  -0.564    0.573
const         0.10375   0.04358   2.381    0.018 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 0.6825 on 247 degrees of freedom
Multiple R-Squared: 0.009656,   Adjusted R-squared: 0.001637
```
F-statistic: 1.204 on 2 and 247 DF,  p-value: 0.3017

Estimation results for equation JP.Morgan:

\[
\text{JP.Morgan} = \text{S.P500}.l1 + \text{JP.Morgan}.l1 + \text{const}
\]

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| S.P500.l1 | 0.52797 | 0.12227 | 4.318 | 2.28e-05 *** |
| JP.Morgan.l1 | -0.17623 | 0.06604 | -2.669 | 0.00812 ** |
| const | 0.08523 | 0.07737 | 1.102 | 0.27172 |

---

Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 1.212 on 247 degrees of freedom
Multiple R-Squared: 0.07407,   Adjusted R-squared: 0.06657
F-statistic: 9.879 on 2 and 247 DF,  p-value: 7.456e-05
The fitted model for S&P 500 returns is not significant; indicating that it is almost impossible to predict future returns for S&P 500. Nevertheless, the fitted model for JP Morgan returns is highly significant with $P$ value smaller than 0.0001. Furthermore, the estimated coefficient for the lagged S&P 500 return is 0.52797 which is statistically highly significant. The estimated covariance matrix for the innovation is

$$\hat{\Sigma}_\varepsilon = \begin{pmatrix} 0.4658 & 0.3515 \\ 0.3515 & 1.4681 \end{pmatrix}.$$ 

The estimated correlation coefficient between the two innovation components is 0.425.

To estimate impulse response functions for this fitted AR(1) model

> Psi(fitX).
Impulse response functions (the $2 \times 2$ matrices as a function of $k$) of the fitted AR(1) model for the daily log returns (in percentage) of S&P 500 index and JP Morgan stocks in 2013.
Since $\hat{\Sigma}_\varepsilon$ is not a diagonal matrix, the estimated responses were measured with respect to the impulses in the two components of the standardized innovation $e_t$.

There exist hardly any responses from S&P 500 return to both components of $e_t$ at non-zero lags. (Note that the response at zero lag has no predictive values.)

On the other hand, the response from JP Morgan to the first component of $e_t$ is greater than 0.5 at lag 1. Since there exists no significant autocorrelations at non-zero lags for this JP Morgan return series, this must be the response to the impulse in the return of S&P 500 on the previous day. This analysis indicates that S&P 500 index carries a lead effect on JP Morgan stock at lead 1, but not vice versa.