Lecture 4: VAR models

Dr. Joao B. Duarte[1]

[1]Nova School of Business and Economics
University of Cambridge

Masters, Economics: Macroeconometrics
Lisbon
Spring 2017

Lecture Objectives:

- Motivation for using VAR in macro.
- Structural VAR vs Reduced-form VAR.
- VAR properties including conditions for stationarity.
- VAR lag selection and estimation.
- VAR forecast and identification.
- SVAR using the recursive identification strategy and the related Choleski Decomposition.
- Impulse response function and their confidence intervals construction and the forecast error variance decomposition.
- Granger Causality.

Secondary Readings:

- Chapter 5, Applied Econometric Time Series, Enders, Walter, Fourth Edition
- Chapters 10 and 11, Time Series Analysis, Hamilton, James, first edition

Motivation

- Two main approaches:

1. Structural Models

   - Traditionally, macroeconometric hypothesis tests and forecasts were conducted using large-scale macroeconometric models. Example of a small structural model:

$$
\begin{aligned}
P_t &= \phi_0 + \phi_1 z_t + \varepsilon_{1t} \\
Q_t &= \theta_0 + \theta_1 a_t + \varepsilon_{2t}
\end{aligned}
$$

   - Macroeconomists used to estimate systems like above but with a very high number of equations.
   - Implicit in these systems are identification restrictions. In this example for instance, it is assumed z_t only affects P_t and that a_t only affects Q_t. Moreover, it is assumed that Q_t and P_t are unrelated.

Motivation

2. Reduced-form Models

- These models were primarily used to study the effect of macro policies on a particular variable of interest. Many monetarists used such models to study the effect of money on the economy. Example:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 m_t + \phi_3 m_{t-1} + \varepsilon_t$$

Motivation

- There were serious problems with both of the previous approaches.
- For the structural approach it was very hard to defend ad hoc restrictions to identify the systems. This was particularly true for high dimensional systems.
- Moreover, variables typically affect each other, i.e. they are endogenous.
- For the reduced-form approach, for instance, it is hard to defend that only money affects GDP and not the other way around.
- Finally and more importantly, the Lucas critique showed that such models were not suited to analyze policy effects in the macroeconomy.

Motivation

- Sims (1980) proposes the use of VARs and saves the day.
- VAR models continue to be the work-horse of empirical macroeconomics and finance.
- They are extremely useful as a descriptive tool of the dynamics of a set of variables.
- They are also a powerful tool in forecast.
- The ability of structural representations of VAR models to differentiate between correlation and causation, in contrast, has remained contentious

Motivation

- Structural interpretations of VAR models require additional identifying assumptions that must be motivated based on institutional knowledge, economic theory, or other extraneous constraints on the model responses.
- Only after decomposing forecast errors into structural shocks that are mutually uncorrelated and have an economic interpretation can we assess the causal effects of these shocks on the model variables.
- The structural VAR model literature has continuously evolved since the 1980s. Even today new ideas and insights are being generated. In the next lecture we review many of these alternative identification schemes.

Intro to VAR

- Lets start with a two variable case with just one lag.

$$
\begin{aligned}
y_t &= b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt} && (1)\\
z_t &= b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt} && (2)
\end{aligned}
$$

  Where $\varepsilon_{yt}, \varepsilon_{zt}$ are both white noise processes with variance σ_y, σ_z respectively. For now, we assume y_t and z_t are stationary.

- (1) and (2) constitute a Structural two-variable and one-lag VAR model.

- The structure of the system incorporates feedback because y_t and z_t are allowed to affect each other.
- $\varepsilon_{yt}, \varepsilon_{zt}$ are pure innovations. They do not depend on any other shock and are easy to interpret.

Intro to VAR

- Equations (1) and (2) cannot be estimated using OLS because of the simultaneous equation bias.
- However, we can manipulate the system using matrix algebra.

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

Or more compactly:

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

- where all variables are in vector notation.

Intro to VAR

- We can pre-multiply both sides by B^{-1} to obtain the Reduced-form VAR:

$$x_t = A_0 + A_1 x_{t-1} + e_t \tag{3}$$

Where $A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1$ and $e_t = B^{-1}\varepsilon_t$

- Note that (3) is just a vector version of an AR(1) process. Hence the VAR name.
- Notice that the inverse of the matrix B connects the structural shocks and reduced-form shocks through:

$$\boxed{e_t = B^{-1}\varepsilon_t}$$

Intro to VAR

- We can rewrite (3) as

$$
\begin{aligned}
y_t &= a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t} && (4) \\
z_t &= a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t} && (5)
\end{aligned}
$$

- (1) and (2) is called the Structural VAR while (4) and (5) is called the Reduce-form VAR.
- Note that e_{1t}, e_{2t} are both combinations of the structural shocks via $e_t = B^{-1}\varepsilon_t$. Hence

$$
\begin{aligned}
e_{1t} &= \frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}} && (6) \\
e_{2t} &= \frac{\varepsilon_{zt} - b_{21}\varepsilon_{yt}}{1 - b_{12}b_{21}} && (7)
\end{aligned}
$$

Intro to VAR

- Note that e_{1t}, e_{2t} are themselves also white noise processes because they are nothing but linear combinations of white noise processes.
- However, they are generally correlated with each other:

$$E(e_{1t}e_{2t}) = \frac{-(b_{21}\sigma_y^2 + b_{12}\sigma_z^2)}{(1 - b_{12}b_{21})^2}$$

- Moreover, notice that there is no clear interpretation for e_t in contrast with ε_t. Lets write the variance-covariance matrix of e:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \tag{8}$$

VAR Stability and Stationarity

- It should not come as a surprise that the stability conditions are similar to those of AR models.
- Moreover, they are also closely connected with stationarity. All we need is that $A_1^n x_{t-n-1} \to 0$ as $n \to \infty$ in (3).
- The difference is that before A_1 was a real number and now it is a Matrix.
- However, it can be shown that if each eigenvalue λ of A satisfies $|\lambda| < 1$, then for any vector x,

$$\lim_{n \to \infty} A^n x = 0 \tag{9}$$

VAR Stability and Stationarity

- Hence, once again it is about the characteristic polynomial:

$$(a_{11}-\lambda)(a_{22}-\lambda)-a_{12}a_{21}$$

$$\lambda^2-\lambda(a_{22}+a_{11})+a_{11}a_{22}-a_{12}a_{21}$$

- Hence, the VAR is stationary if both roots λ_1, λ_2 lie within the unit circle.

VAR Stability and Stationarity

- Again we can alternatively do it by using the Lag-operator in (4) and (5):

$$\begin{aligned} y_t &= a_{10} + a_{11} L y_t + a_{12} L z_t + e_{1t} \\ z_t &= a_{20} + a_{21} L y_t + a_{22} L z_t + e_{2t} \end{aligned}$$

- One can solve this system and find that:

$$y_t = \frac{K}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}$$

- Where K depends on e_t and coefficients. We can find z_t and it turns out it has exactly the same denominator.
- Hence the stability and stationarity condition amounts for the solution to the denominator to be outside the unit circle.

VAR Stability and Stationarity

- Example:Panel (a): $a_{10} = a_{20} = 0$, $a_{11} = a_{22} = 0.7$ and $a_{12} = a_{21} = 0.2$
  Panel (b) $a_{12} = a_{21} = -0.2$
  Panel(c) $a_{11} = a_{22} = a_{12} = a_{21} = 0.5$
  Panel (d) $a_{10} = 0.5$

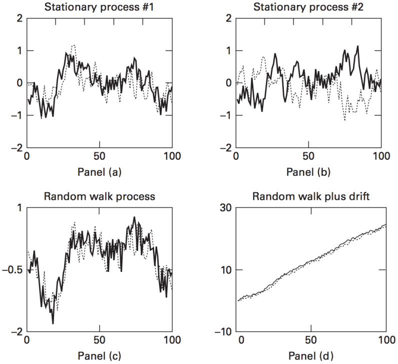


Figure: Enders VAR examples

VAR Stability and Stationarity

- If the individual variables are stationary, the VAR will also be stationary because it is just linear combinations of stationary processes/
- If the VAR is stationary, it is also ergodic and standard hypothesis testing can be done.
- However, there is a debate on whether the series should be stationary or not. Sims for instance recommends against differencing on the grounds that information is lost.
- If the purpose is to understand dynamic interrelationships between series then, the variables do not need to be stationary.
- If on the other hand, one cares about hypothesis testing and forecast, then the series should be stationary.

VAR Lag Selection

- We generally use the information criteria to select the number of lags in a VAR. There are other methods however.
- It is also important to take into account two things when selecting the number of lags:
  1. Parsimonious models. That is, smaller number of lags are preferred. The models are more accurate.
  2. Seasonality. Although there are some ways we will discuss later on how to deal with it, some lags are still more important than others. Example: The lag 4 is usually very important in a quarterly series.

VAR Estimation

- Note that the Reduced-form of the VAR (4) and (5) contains only pre-determined variables.
- Hence, we can use OLS estimation equation by equation.
- Moreover, note that (4) and (5) also forms a SUR due to the correlation between the errors
- However, since all regressions have identical right-hand side variables, we would not gain any efficiency by exploring this correlation.
- Typically, we estimate VARs using OLS equation by equation.

VAR Estimation

- One thing that you probably have noticed by now is that the number of coefficients increases very fast as one increases the number of lags in a VAR
- Even for a small VAR, the VAR is typically overparameterized in that many of these coefficient estimates will be insignificant
- However, the goal is to find the important interrelationships among the variables.
- Improperly imposing zero restrictions may waste important information.
- Moreover, the regressors are likely to be highly collinear so that the t-tests on individual coefficients are not reliable guides for paring down the model.

VAR Forecast

- We can use the techniques learned in the AR processes to analyse the properties of the VAR forecast using (3).
- However, for forecast it is useful to throw away the insignificant parameters. Moreover, it is essential that the series are stationary.
- After throwing away the insignificant coefficient, one can estimate the near-VAR using SUR (note that after throwing away coefficient, the right-hand side variables becomes different across equation and higher efficiency can be obtain via SUR).

VAR Identification

- Lets go back to our structural VAR:

$$
\begin{aligned}
y_t &= b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt} && (10) \\
z_t &= b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt} && (11)
\end{aligned}
$$

- We know we cannot directly estimate (10) and (11).
- OLS of the reduced-form representation of (10) and (11) give us the two elements of A_0, the four elements of A_1 and the four elements of the variance-covariance matrix Σ using the residuals. However, note that $\sigma_{12} = \sigma_{21}$.
- In total the OLS of the reduced-form give us 9 estimates.

VAR Identification

- Identification of the VAR as to with our ability to Identify (10) and (11) with our estimates of the reduced-form (4) and (5).
- In other, words can we recover all the parameters of (10) and (11) with the OLS estimates of the reduced-fom VAR?
- Without any further assumptions, the answer is no.
- The structural system has 10 parameters while the reduced-form estimates only give us 9.

SVAR - Recursive VAR

- The first structural VARs used the recursive identification strategy proposed by Sims. Sometimes, they are just called recursive VARs.
- Lets take a look at the matrix, B^{-1},that connected both representations of a VAR:

$$e_t = B^{-1}\varepsilon_t = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}^{-1} \varepsilon_t \Rightarrow$$

$$e_t = \begin{bmatrix} \frac{1}{1-b_{12}b_{21}} & -\frac{b_{12}}{1-b_{12}b_{21}} \\ -\frac{b_{21}}{1-b_{12}b_{21}} & \frac{1}{1-b_{12}b_{21}} \end{bmatrix} \varepsilon_t$$

SVAR - Recursive VAR

- Note that if we either restrict $b_{12} = 0$ or $b_{21} = 0$ we can identify ε_t from e_t. Lets see why:
- Suppose $b_{21} = 0$:

$$e_t = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \varepsilon_t \Rightarrow$$

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

SVAR - Recursive VAR

- So:

$$
\begin{aligned}
e_{1t} &= \varepsilon_{yt} - b_{12}\varepsilon_{zt} \\
e_{2t} &= \varepsilon_{zt}
\end{aligned}
$$

- Hence the variance and covariance of e_t estimated, (3 parameters), is enough to recover σ_y^2, σ_z^2 plus b_{12} (3 parameters):

$$
\begin{aligned}
var(e_{1t}) &= \sigma_y^2 + b_{12}^2\sigma_z^2 \\
var(e_{2t}) &= \sigma_z^2 \\
cov(e_{1t}, e_{2t}) &= -b_{12}\sigma_z^2
\end{aligned}
$$

SVAR - Recursive VAR

- These results can be generalized. Without going to much into detail, lets go back to the mapping between structural and the reduced-form model:

$$e_t = B^{-1}\varepsilon_t \tag{12}$$

- Hence,

$$\Sigma = B^{-1}\Sigma_{\varepsilon}(B^{-1})^T \tag{13}$$

- The identification problem comes from the fact there are $\frac{n^2+n}{2}$ known parameters since Σ is symmetric (remember $\sigma_{12} = \sigma_{21}$ etc) and we need to recover n^2 unknown parameters.
- The unknown parameters are $n^2 - n$ from B plus n values of Σ_{ε}. Hence the n^2

SVAR - Recursive VAR

- In order to exactly identify the system, we need to impose $n^2 - \frac{n^2+n}{2} = \frac{n^2-n}{2}$ restrictions.
- A recursive formulation of the structural VAR imposes exactly that number of restrictions on B by imposing zero restrictions on the structural parameters in a recursive manner. In general for a VAR with n variables:

$$B = \begin{bmatrix} 1 & b_{12} & \cdots & b_{1n} \\ b_{21} & 1 & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & 1 \end{bmatrix} \tag{14}$$

- The recursive assumption imposes:

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ b_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & 1 \end{bmatrix} \tag{15}$$

SVAR - Recursive VAR

- Note that the number of restrictions is exactly $\frac{n^2-n}{2}$. Thus the recursive structure exactly identifies
- There is an alternative way of doing the recursive approach that is numerically easier to compute (This is what software packages use) using the Choleski Decomposition and is equivalent
- One can always do a Choleski decomposition of the residuals variance-covariance matrix $\Sigma = LL^T$ where L is triangular and use it to recover the structural errors.

$$\varepsilon_t = L^{-1} e_t \tag{16}$$

SVAR - Recursive VAR

- However L has positive elements in its diagonal different from 1. Hence, L has exactly $\frac{n^2-n}{2}$. What about the Σ_ε? No need, they will be by construction normalized to 1:

$$var(\varepsilon_t) = L^{-1} var(e_t)(L^{-1})^T$$

$$\Sigma_\varepsilon = L^{-1}\Sigma(L^{-1})^T$$

$$\Sigma_\varepsilon = L^{-1}LL^T(L^{-1})^T$$

$$\Sigma_\varepsilon = I$$

- It turns out both approaches are equivalent.

SVAR - Recursive VAR

- Caveat Note: To impose the $\frac{n^2-n}{2}$ in a recursive way is sufficient to recover all the parameters. But in general this is just a necessary condition and not sufficient.
- The SVAR main tools are:
  1. Impulse response Functions
  2. Variance Decomposition
- Note that both of them only make sense in a structural VAR (why?)

SVAR - Impulse response Function

- Back to our reduced-form VAR

$$x_t = A_0 + A_1 x_{t-1} + e_t \tag{17}$$

- Just like in the AR model, we can also rewrite the reduced-form VAR as VMA(∞):

$$x_t = \mu + e_t + A_1 e_{t-1} + A_1^2 e_{t-2} + ... \tag{18}$$

  Where μ is the unconditional mean. The coefficients A_1 can be seen as impact multipliers

- However, remember that e_t include many structural shocks. Hence, it is hard to interpret (18)

SVAR - Impulse response Function

- After identification, the structural shocks can be recovered $L\varepsilon_t = e_t$. Given L from the Choleski decomposition we can use it to get the impulse-response function that can be interpreted:

$$x_t = \mu + L\varepsilon_t + A_1 L\varepsilon_{t-1} + A_1^2 L\varepsilon_{t-2} + \ldots \tag{19}$$

- If the VAR is stationary, the impact multipliers vanish.
- Moreover, given that for each order of variables we have a different unique L. Depending on the order we can get different impulse response functions.
- This is a problem and has generated until today many proposals to overcome this identification problem.

SVAR - Impulse response Function

- Example:

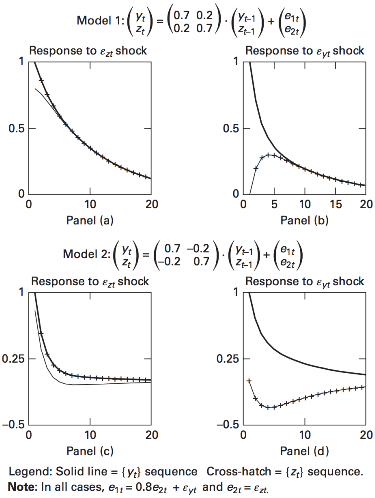


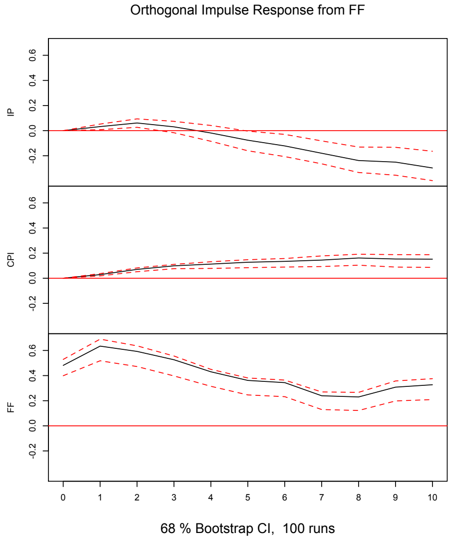
Figure: Enders Figure 5.7

SVAR - Impulse response Function Confidence Intervals

- The impulse response functions are very complex objects (non-linear functions of coefficients)
- Generally, we don't know their distribution
- We need to rely on the following procedures:
  1. Bootstrap bias corrected
  2. Monte Carlo integration
  3. Delta method
- Kilian (1998) showed Bootstrap bias corrected is the best for small sample analysis. R uses this procedure.

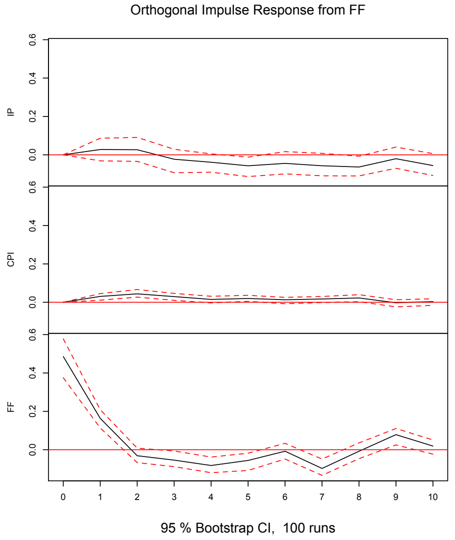
SVAR - Impulse response Function Confidence Intervals

- Example:



SVAR - Impulse response Function Confidence Intervals

- Example: First Differences



SVAR - Variance Decomposition

- Lets start again with the VMA representation of the VAR:

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i L \varepsilon_{t-i}$$

- To simplify the algebra, lets define $\phi_i = A_1^i L$. Then, taking n steps forward:

$$x_{t+n} = \mu + \sum_{i=0}^{\infty} \phi_i \varepsilon_{t+n-i}$$

- Using this representation, we can easily find the *n*-period forecast error $x_{t+n} - E(x_{t+n})$:

$$x_{t+n} - E(x_{t+n}) = \sum_{i=0}^{n-1} \phi_i \varepsilon_{t+n-i} \tag{20}$$

SVAR - Variance Decomposition

- In our two example of two variables VAR y, z, lets focus on the y portion of (20).

$$\begin{aligned} y_{t+n} - E(y_{t+n}) =& \phi_{11}(0)\varepsilon_{yt+n} + \phi_{11}(1)\varepsilon_{yt+n-1} + ... + \phi_{11}(n-1)\varepsilon_{yt+1} \\ &+ \phi_{12}(0)\varepsilon_{zt+n} + ... + \phi_{12}(n-1)\varepsilon_{zt+1} \end{aligned}$$

- Denote the n-step-ahead forecast error variance of y_{t+n} as $\sigma_y^2(n)$

$$\begin{aligned} \sigma_y^2(n) =& \sigma_y^2(\phi_{11}(0)^2 + .. + \phi_{11}(n-1)^2) \\ &+ \sigma_z^2(\phi_{12}(0)^2 + .. + \phi_{12}(n-1)^2) \end{aligned} \tag{21}$$

- Hence, we can decompose the forecast variance into the variance of the various structural shocks.

SVAR - Variance Decomposition

- Example:

```
> fevd(var1, n.ahead=60)
$IP
             IP          CPI          FF
 [1,] 1.0000000 0.0000000000 0.000000000
 [2,] 0.9971536 0.0018241303 0.001022235
 [3,] 0.9956089 0.0016622187 0.002728920
 [4,] 0.9961732 0.0016631747 0.002163637
 [5,] 0.9970295 0.0013135475 0.001656963
 [6,] 0.9963075 0.0010992775 0.002593202
 [7,] 0.9941806 0.0009277913 0.004891592
 [8,] 0.9897229 0.0008684191 0.009408670
 [9,] 0.9821272 0.0014920637 0.016380695
[10,] 0.9752455 0.0020445960 0.022709919
[11,] 0.9673315 0.0018998714 0.030768594
[12,] 0.9584242 0.0018111208 0.039764647
[13,] 0.9487683 0.0018784060 0.049353338
[14,] 0.9382502 0.0022943854 0.059455396
[15,] 0.9273537 0.0030767393 0.069569608
[16,] 0.9168430 0.0042904459 0.078866537
[17,] 0.9060854 0.0058998486 0.088014706
[18,] 0.8952551 0.0076177892 0.097127104
[19,] 0.8845347 0.0093762891 0.106089030
[20,] 0.8735740 0.0114725073 0.114953459
[21,] 0.8620931 0.0142408262 0.123666102
[22,] 0.8501713 0.0177479933 0.132080714
[23,] 0.8377954 0.0218993101 0.140305293
[24,] 0.8252133 0.0264833545 0.148303372
```

Granger Causality

- Granger causality has to do with the ability of a current variable y in predicting future values of another variable z.
- Note that is only future and not current! Also the word causality should not be taken serious in the sense of the usual causality in microeconometrics.
- If say z Granger causes y and not the other way around, then there is stronger grounds for causation but still not definite.
- To test for Granger causality of z on y we use a F-test on the coefficients on the z lags in the y equation. If we can reject that they are zero then z granger causes y

Granger Causality

- Example:

```
> causality(var1, cause="FF")
$Granger

	Granger causality H0: FF do not Granger-cause IP CPI

data:  VAR object var1
F-Test = 2.6091, df1 = 24, df2 = 1512, p-value = 3.541e-05


$Instant

	H0: No instantaneous causality between: FF and IP CPI

data:  VAR object var1
Chi-squared = 21.864, df = 2, p-value = 1.787e-05
```

Figure: Does FF Granger cause IP and CPI

Summary of VAR in Practice

1. If we want to do hypothesis testing and forecast, we need to check if the variables included in the VAR are stationary. (ACF can be helpful, we will see more formal tests on this issue next lecture)
2. Choose the VAR lag p. Generally we can use the AIC and SBC to do this.
3. Estimate the VAR with OLS.
4. Again, for hypothesis testing and forecast you need to check the parameter estimates for significance. Otherwise, just skip this step.
5. Choose an identification strategy and compute the IRF and the FEVD.

Questions to think about

- How can we get from a Reduced-Form VAR to a SVAR and vice-versa?
- Why can't we identify the SVAR parameters with the reduced-form VAR?
- Does the Recursive identification always deliver exact identification?
- How can we interpret the IRF? What is the scale of the IRF at each step?
- Can we infer causality from the FEVD?