VARs and Granger Causality

- Economic theories usually consist of a set of endogenous variables that are determined, in equilibrium, by exogenous parameters

- We are almost always interested in these endogenous outcomes, though sometimes we are interested in fundamental parameters.

  - Eg. Demand and supply elasticities, capacity constraints.

- However, it is extremely difficult to determine causality

  - You know this already since 216 was focused on much of these identification problems

- Time series has a number of techniques that side-step or re-cast issues of causation

  - **VAR**: Vector autoregression

  - Granger Causality

Vector autoregression (VAR)

- VARs are basically systems of equations with outcome variables that depend on other outcome variables

- Consider the following simply time series model

$$y_t = \beta_{10} - \beta_{12}x_t + \gamma_{11}y_{t-1} + \gamma_{12}x_{t-1} + u_{yt}$$
$$x_t = \beta_{20} - \beta_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}x_{t-1} + u_{xt}$$

- VARs are meant to get around the obvious concerns in the above system of equations. What causes what?

- Instead, we again focus on predicting endogenous variables as a function of lags. The goal is exploiting useful information and making predictions.

Vector autoregression (VAR)

- VARs treat the x's and y's the same, and view the problem was one of forecasting as opposed to empirical identification.

- Again, consider the model:

$$
\begin{aligned}
y_t &= \beta_{10} - \beta_{12}x_t + \gamma_{11}y_{t-1} + \gamma_{12}x_{t-1} + u_{yt} \\
x_t &= \beta_{20} - \beta_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}x_{t-1} + u_{xt}
\end{aligned}
$$

- Solve for y_t and x_t on the LHS:

$$
\begin{aligned}
y_t + \beta_{12}x_t &= \beta_{10} + \gamma_{11}y_{t-1} + \gamma_{12}x_{t-1} + u_{yt} \\
\beta_{21}y_t + x_t &= \beta_{20} + \gamma_{21}y_{t-1} + \gamma_{22}x_{t-1} + u_{xt}
\end{aligned}
$$

- Arranging the system of equations in matrix notation:

$$
\begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{yt} \\ u_{xt} \end{bmatrix}
$$

- Invert:

$$
\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \tilde{u}_{yt} \\ \tilde{u}_{xt} \end{bmatrix}
$$

Vector autoregression (VAR)

- Define the vector of constants as

$$\mathbf{B}_0 = \begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix}$$

- Define the matrix of lag coefficients as:

$$\mathbf{B}_1 = \begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$

- Solving for $\begin{bmatrix} y_t \\ x_t \end{bmatrix}$, we have the "Reduced form VAR":

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \mathbf{B}_0 + \mathbf{B}_1 \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \widetilde{u}_{yt} \\ \widetilde{u}_{xt} \end{bmatrix}$$

- Current observables, y_t and x_t, are functions of lagged observables.

- Some points about this VAR:

  - For purposes of forecasting, this is as convenient as anything we've done thus far.

  - This VAR cannot recover the original parameters.

Estimating a VAR in R

- Estimating a VAR in R is as simple as running a linear regression, but there are also handy packages to do this:

- First, let's download the data we need

```
library(quantmod)
getSymbols('GOOG',from='2014-11-12',to='2015-11-12')
vol<-GOOG$GOOG.Volume
price<-GOOG$GOOG.Open
```

- Question: Do prices drive volume or volume drive prices?

- We first run a reduced for VAR by directly programming into OLS

```
n<-length(price)
regPrice<-lm(price[2:n]~price[1:(n-1)]+vol[1:(n-1)])
regVol<-lm(vol[2:n]~price[1:(n-1)]+vol[1:(n-1)])
summary(regPrice)
summary(regVol)
```

- Make sure the indexing is correct!

Estimating a VAR in R

- The package MSBVAR has a load of functions that help in time-series analysis and forecasting

- First, load the required library

```
library(MSBVAR)
```

- Initiate the data as a time-series object

```
y2<-ts(data.frame(price,volume))
```

- Run the VAR in reduced form

```
pricevol_var<-reduced.form.var(y2,p=1)
```

  - $p = 1$ specifies one period lag. Increase for bigger lag lengths.

- Summarize like our earlier versions

```
summary(pricevol_var)
```

Causality Tests

- Causality in time series is defined differently from applied micro (as in 216)

- In applied micro "style", we would usually approach a problem as follows:

    - Find an "instrument" for x_t

    - Use IV to find the precise estimate for β_{12}.

    - Other techniques could also be used (randomly allocate x_t's, use discontinuities if they exist, etc..)

- "Causality" in time series usually refers to the "information" contained within a time series for some variable.

- "Granger Causality"

    - Controling for the history of y's, the history of x's help predict y.

    - x "Granger Causes" y.

    - This is not the same as casual inference - omitted variables could still affect this conclusion.

Granger Causality Test

- Granger Causality Test is performed by the following three-step procedure (which is essentially a F-test)

- Step 1: Regress y on y lags without x lags (restricted model)

$$y_t \;=\; a_1 + \sum_{j=1}^{m} \gamma_j y_{t-j} + e_t$$

- Step 2: Add in x lags and regress again (unrestricted model)

$$y_t \;=\; a_1 + \sum_{i=1}^{n} \beta_i x_{t-i} + \sum_{j=1}^{m} \gamma_j y_{t-j} + e_t$$

- Step 3: Test null hypothesis that $\beta_i = 0 \; \forall i$ using a F-test

Granger Causality Test in R

- Running a Granger test in R is quite simple

- Using the same time series object as in the previous example:

  ```
  granger.test(y2,p=2)
  ```

- If $p = 1$, we only have one lag. Do we need a Granger testing this case?

Sims Causality Test

- "Sims Causality"

  - Controling for lag y's and x's, do future x's predict current y's?

- Sims Causality tests for the effect of "leading terms"

- **Step 1:** Regress y on y lags and x lags (restricted model)

$$y_t = a_1 + \sum_{i=1}^{n} \beta_i x_{t-i} + \sum_{j=1}^{m} \gamma_j y_{t-j} + e_{1t}$$

- **Step 2:** Add in x leading terms $(t + \rho)$ and regress again (unrestricted model)

$$y_t = a_1 + \sum_{i=1}^{n} \beta_i x_{t-i} + \sum_{j=1}^{m} \gamma_j y_{t-j} + \sum_{\rho=1}^{m} \xi_j x_{t+\rho} + e_{1t}$$

- **Step 3:** Test null hypothesis that $\xi_j = 0$ $\forall i$ using a F-test

- If we reject this null hypothesis, then y causes x since the future cannot predict the present.