Time After Time II

Statistical Issues in Multivariate Time-Series Data

Jan Rovny

Sciences Po, Paris

Where are we going?

- Earlier we saw the issues in univariate time-series
- Now, let us look at time-series with two or more variables
- Will assess (causal) flows between variables in time $x_t \rightarrow y_t$:
  1. Prewhitening
  2. Granger Causality
  3. Error Correction Models

Prewhitening 1

- If we have two series that are white noise processes, we can see their association (causation) by looking at how the two series are correlated at different lags and leads
- Causation is by definition asymmetrical if $x \rightarrow y$, then it is not true that $y \rightarrow x$. That implies that lagged values of x will be correlated with y, but not the reverse.
- Real world time-series, however, have various error aggregation processes (see Time-Series I)
- It is thus the error process in x that leads to y
- We thus need to remove the filter to observe x and y as white noise – this is the logic of prewhitening

Prewhitening 2

- Need to thus identify, estimate, and diagnose both series x and y using ARIMA (see Time-Series I)
- Then we need to capture and correlate the residuals:
  - Estimate the model for x
    ```
    m1<-arima(x, order=c(#,#,#))
    ```
  - Capture the residuals of x
    ```
    x.res<-m1$residuals
    ```
  - Estimate the model for y
    ```
    m2<-arima(y, order=c(#,#,#))
    ```
  - Capture the residuals of y
    ```
    y.res<-m2$residuals
    ```
- Now cross correlate the residuals:
  ```
  ccf(x.res,y.res)
  ```
- Positive lags: here lagged values of x correlate with current values of y, thus $x \rightarrow y$
- Negative lags: here lagged values of y correlate with current values of x, thus $y \rightarrow x$ WRONG WAY!

Granger Causality

- Granger causality posits that if $x \to y$, change in x leads to later changes in y.
- Modeling y as caused by its previous lags, and seeing if adding information about the history of x improves the prediction of y
- The logic is the following:
  1) $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_k y_{t-k} + e_t$
  2) $y_t = \beta_0 + \beta_1 y_{t-1} + ... + \beta_k y_{t-k} + \gamma_1 x_{t-1} + ... + \gamma_k x_{t-k} + e_t$
- The process is then reversed, in the attempt to test the opposite causal flow from y to x.
- In R use package "lmtest":

  ```
  grangertest(y~x, order=#, data=" ")
  ```

  - Here"order" decides the number of lags to consider
  - The test compares the effect of lags of x and y on y (eq 2) with just lags of y on y (eq 1)
  - If the reported F-test has a significant p-value, then the causal order is correct

Multiple Granger Causality

- We can use Vector Autoregressive models, which allow data to atheoretically speak for themselves to test relationships between many variables simultaneously
- The model is appealing due to its simplicity
- But it is a travesty in estimation: all variables on both sides of the equation...
- See R demonstration

Cointegration 1: The Logic

- Cross-correlation and granger causality tests work only for stationary data!
- What if our x and y are not stationary, but integrated together – cointegrated?
- In cointegration, series x sets a target level to which series y responds. If the series are truly causally related, then a mismatch between the series must be subsequently corrected.

Cointegration 2: The Mechanics

- Cointegrated series move in tandem
- Regressing one on the other thus estimates the movement, and leaves the residuals stationary.
- Finally, to ascertain causality $x \to y$, the stationary errors get corrected in the future values of y.

Cointegration 3: The Estimation

- Engel and Granger Two-Step method:
  1) $y_t = \beta_0 + \beta_1 x_t + u_t$ estimate the simple regression to get its residuals z
  2) $\Delta y_t = \alpha \Delta x_t - \pi z_{t-1} + \upsilon_t$ estimate the error correction in z
  Here a negative coefficient π suggests error correction
- Error Correction Model:
  $\Delta y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \beta_3 \Delta x_t + e_t$
  - y_{t-1} is the lagged dependent variable (LDV)
  - The coefficient β_3 (on Δx) suggests Granger causality whereby change in x causes change in y
  - A negative coefficient β_1 (on the LDV) suggests error correction.
- See R demonstration