Lecture Note 1

- Spurious correlation in time series analysis

Spurious correlation in time series analysis

Consider two vectors of random variables $\{x_t\}_{t=1}^T$ and $\{y_t\}_{t=1}^T$ that are **independently** generated according to the following data generating processes:

$$
\begin{aligned}
y_t =& \lambda y_{t-1} + \varepsilon_{1t}, \text{ where } \varepsilon_{1t} \sim \mathcal{N}(0,1), \ |\lambda| < 1 \text{ and } y_0 = 0, \\
x_t =& \rho x_{t-1} + \varepsilon_{2t}, \text{ where } \varepsilon_{2t} \sim \mathcal{N}(0,1), \ |\rho| < 1 \text{ and } x_0 = 0.
\end{aligned}
$$

An investigator estimates the coefficient β of the following regression

$$y_t = \beta x_t + u_t$$

by the least square (LS) method. She wants use t-statistics to test the null hypothesis that $\beta = 0$ against the alternative one that $\beta \neq 0$, that is

$$\mathcal{H}_0 : \beta = 0, \text{ v.s. } \mathcal{H}_a : \beta \neq 0 \overset{t-stat}{\longrightarrow} \mathbf{t} = \frac{\hat{\beta}}{s.e.(\hat{\beta})}.$$

If there exists no serial correlation, i.e. $\lambda = \rho = 0$, then, under the null hypothesis, the test statistics asymptotically has a normal distribution with mean zero and variance one, i.e. $\mathbf{t} \overset{a.s.}{\sim} \mathcal{N}(0,1)$. But, if $\lambda \neq 0$ and $\rho \neq 0$, it can be shown that under the null hypothesis, the test statistics has an asymptotic distribution with non-zero mean, i.e. the mean is equal to $\frac{1+\lambda\rho}{1-\lambda\rho}$.

So, if we ignore the time series dependence across observations, we are prone to reject the null hypothesis more often than $\alpha\%$ of the times, while the null hypothesis is true.