治理技术专题

定量政治分析方法

Quantitative Analysis II

苏 毓 淞

清华大学社会科学院政治学系

第十二讲　时间序列分析 (II)

分析时间序列数据的难题

- 大部分分析时间序列的模型，前提是数据是平稳的 (Stationary)。
- 但大部分的时间序列是自相关的 (autocorrelated)，而其中自相关有部分原因就是数据不是平稳的。
- 如果回归方程的左右项（因变量、自变量）都是时间序列数据时，通常回归的 R^2 会很高，即便是这两个变量并没有实质的关联。称之为谬误 (spurious) 回归或无意义 (nonsense) 回归。**当时间序列不平稳时，这个问题尤其严重。**
- 金融时间序列数据（例如股票价格），通常具有随机漫步 (random walk) 的性质，也就是说预测明天股价最好的变量是今天的股价加上随机误差项。
- 使用时间序列数据分析因果关系时，数据必须是平稳的。

重要概念

- 随机过程 (Stochastic Processes)
- 平稳过程 (Stationarity Processes)
- 纯随机过程 (Purely Random Processes)
- 不平稳过程 (Nonstationary Processes)
- 整合的变量 (Integrated Variable)

随机过程 (Stochastic Processes)

- 定义：依时间排序的随机变量。
- 平稳随机过程：它的均值和方差是不随时间变化（固定的），两个时间点间变量的协方差 (covariance) 只与它们的时间差有关，而不与时间点有关。
- 又称为弱平稳 (weakly stationary)、协方差平稳 (covariance stationary)、二阶平稳 (second-order stationary)

$$
\begin{aligned}
Mean: &\quad \mathsf{E}(Y_t) = \mu \\
Variance: &\quad \mathsf{var}(Y_t) = \mathsf{E}(Y_t - \mu)^2 = \sigma^2 \\
Covariance: &\quad \gamma_\kappa = \mathsf{E}\left[(Y_t - \mu)(Y_{t+k} - \mu)\right]
\end{aligned}
$$

- 协方差在这里又称自相关协方差。

平稳随机过程 (Stationary Stochastic Processes)

- 平稳随机过程即均值、方差和协方差不随时间变化的过程。
- 这类数据会绕着均值上下波动，波动幅度跟方差大小有关，而且波动总会回到均值。
- 不平稳随机过程：均值或和方差会随时间改变的过程。
- 纯随机（白噪, white noise）过程 (purely random processes)：均值为 0，方差固定，没有序相关 (serial correlation) 的过程

不平稳随机过程 (Nonstationary Stochastic Processes)

- 随机漫步模型 (random walk model)：典型的不平稳随机过程（股价、汇率）。
  1. 无漂移 (drift) 的随机漫步 (没有常数项)
  2. 有漂移的随机漫步（有常数项）

无漂移的随机漫步

- $$Y_t = Y_{t-1} + u_t$$
- u_t 是白噪，$N(0, \sigma^2)$

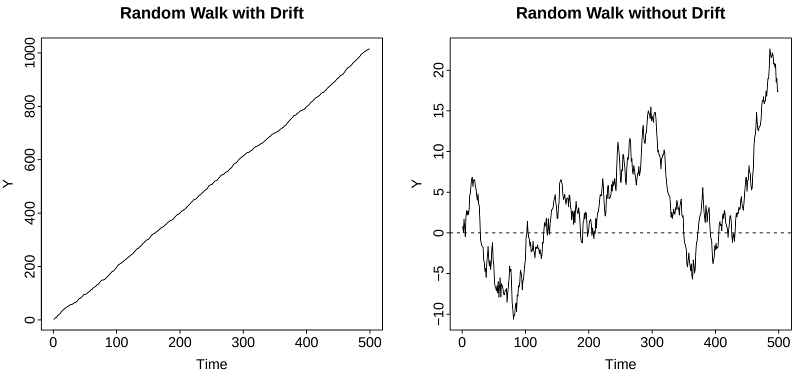
$$\begin{aligned}
Y_1 &= Y_0 + u_1 \\
Y_2 &= Y_1 + u_2 = Y_0 + u_1 + u_2 \\
Y_3 &= Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3
\end{aligned}$$

- 如果起点是 0，$Y_t = Y_0 + \sum u_t$
- $\mathrm{E}(Y_t) = \mathrm{E}\left(Y_0 + \sum u_t\right) = Y_0$
- $\mathrm{var}(Y_t) = t\sigma^2$

有漂移的随机漫步

- $$Y_t = \delta + Y_{t-1} + u_t$$
- 称 δ 为漂移参数，ΔY_t 随着漂移参数上下波动
$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t$$
- $\mathrm{E}(Y_t) = Y_0 + t\delta$
- $\mathrm{var}(Y_t) = t\sigma^2$
- 有漂移的随机漫步的均值和方差随着时间增加，违反了平稳性条件

图示有和无漂移的随机漫步



单位根 (Unit Root) 随机过程

$$Y_t = \rho Y_{t-1} + u_t, \qquad -1 < \rho < 1$$

- 如果 $\rho = 1$，存在单位根（从 $\rho = 1$ 得此名称）问题，等于无漂移的随机漫步，这个随机过程也是不平稳的。
- 如果 $|\rho < 1|$，则这个随机过程也是不平稳的

整合的 (integrated) 过程

- 随机漫步是整合的 (integrated) 过程的特殊情况。
- 在一阶的情况下，无漂移的随机漫步是平稳的，称之为一阶整合的过程。$\Delta Y_t = u_t$
- 如果 Y_t 是平稳的，我们令它为 $Y_t \sim I(0)$；如果 Y_t 在一阶的情况下是平稳的，我们令它为 $Y_t \sim I(1)$，$\ldots$；如果 Y_t 在 d 阶的情况下是平稳的，我们令它为 $Y_t \sim I(d)$

整合的 (integrated) 过程的特性

- $X_t \sim I(0), Y_t \sim I(1)$，则 $Z_t = X_t + T_t \sim I(1)$，平稳的过程和不平稳的过程相加后的过程是不平稳的。
- $X_t \sim I(d)$，则 $Z_t = (a + bX_t) \sim I(d)$，$I(d)$ 随机过程的线性和还是属于 $I(d)$。
- $X_t \sim I(d_1), Y_t \sim I(d_2)$，则 $Z_t = (aX_t + bY_t) \sim I(d_2)$，如果 $(d_1 < d_2)$。
- $X_t \sim I(d), Y_t \sim I(d)$，则 $Z_t = (aX_t + bY_t) \sim I(d^*)$，一般情况下 $d^* = d$,，但有时候 $d^* < d$

谬误回归

$$y_t = y_{t-1} + u_t, \qquad u_t \sim (0, 1)$$
$$x_t = x_{t-1} + v_t, \qquad v_t \sim (0, 1)$$

```
> summary(M1)

Call:
lm(formula = y ~ x)

Multiple R-squared:  0.2314,    Adjusted R-squared:  0.2298
F-statistic: 149.6 on 1 and 497 DF,  p-value: < 2.2e-16

> library(lmtest)
> dwtest(M1)

        Durbin-Watson test

data:  M1
DW = 0.036172, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

通常 $R^2 > d - statistic^2$ 就可能是谬误回归

Durban Watson 统计量

$$d = \frac{\sum_{t=2}^{T}(\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=1}^{T}(\epsilon_t)^2}$$

- 介于 $0-4$。
- 等于 2 表示没有自相关。
- 接近 0 表示负自相关。
- 接近 4 表示正自相关。

单位根检验

$$
\begin{aligned}
Y_t &= \rho Y_{t-1} + u_t \\
Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\
&= (\rho - 1) Y_{t-1} + u_t \\
\Delta Y_{t-1} &= \delta Y_{t-1} + u_t
\end{aligned}
$$

- Dickey-Fuller test：$H_0 : \delta = 0$ 也就是说过程是平稳的。

```
> adf.test(na.exclude(y))

	Augmented Dickey-Fuller Test

data:  na.exclude(y)
Dickey-Fuller = -1.7005, Lag order = 7, p-value = 0.7051
alternative hypothesis: stationary
```

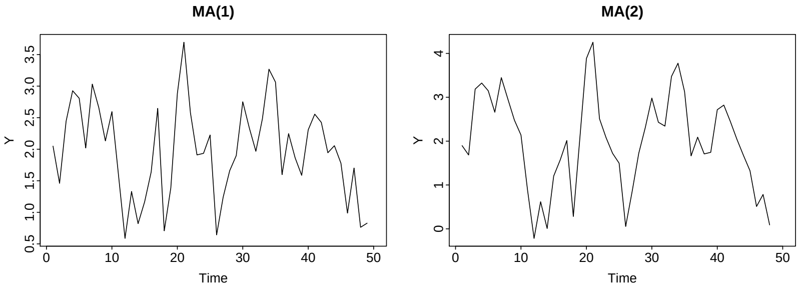
协整性回归

- 两个具有单位根的时间序列相回归，称为协整回归
- 两个序列可能共享同样的趋势，所以这个回归不一定是谬误的。
- 如果两个经济变量长期会达成均衡，那么他们之间一定有协整性。
- 检验？略

移动平均 (Moving Average) 过程

- $Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1}$
- μ 是常数项，u 是白噪随机误差项。
- Y 是当前 $+$ 滞后一阶的误差项的移动平均。
- 所以 Y_t 称之为一阶移动平均过程 MA(1)。
- MA(2):$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2}$

图示移动平均 (Moving Average) 过程



自相关移动平均 (ARMA) 过程

- ARMA(1,1): $Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1}$
- ARMA 假设的是过程是平稳的。
- 如果过程不是平稳的，也就是它是整合的，我们需要 ARIMA 模型。