八、多元时间序列分析

目录

ARIMAX 模型

- ARIMAX 模型
  - ARIMAX 模型结构：$y_t = \beta_0 + \sum_{i=1}^{k} \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x_{it} + \frac{\Theta(B)}{\Phi(B)} a_t$
  - 延迟 k 阶互相关函数：

    $\mathrm{Cov}_k = \mathrm{Cov}(y_t, x_{t-k}) = E[(y_t - E(y_t))(x_{t-k} - E(x_{t-k}))]$
  - 延迟 k 阶互相关系数：$C\rho_k = \frac{\mathrm{Cov}(y_t, x_{t-k})}{\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(x_{t-k})}}$
  - 如果 $k > 0$，计算的是 y 序列滞后于 x 序列 k 期的相关系数
  - 如果 $k < 0$，计算的是 x 序列滞后于 y 序列 k 期的相关系数
  - $C\rho_k \sim N\left(0, \frac{1}{n-|k|}\right)$

干预分析

- 干预分析
  - 定义：评估外部事件对序列产生的影响的分析
  - 实际上是带虚拟变量回归的 ARIMAX 模型
  - 根据作用机制可以分为三种类型：

阶梯干预：$\begin{cases} 0, & \text{干预事件发生之前}(x < T) \\ 1, & \text{干预事件发生之后}(x \geq T) \end{cases}$

脉冲干预：$\begin{cases} 0, & \text{干预事件发生时}(x = T) \\ 1, & \text{其他时刻}(x \neq T) \end{cases}$

其他类型的干预：用阶梯干预和脉冲干预的转换函数或组合来生成

伪回归

- 伪回归
  - 在非平稳的场合，参数显著性检验犯第一类错误的概率远远大于显著性水平，伪回归显著成立
  - 原因：在非平稳场合，参数的 t 检验统计量不再服从 t 分布。统计量真实的抽样分布 $t(\hat{\beta}_1)$ 尾部肥，方差大，比 t 分布要扁平很多。如果继续使用 t 分布的临界值做方程显著性判断，则会导致很大的犯第一类错误的概率

协整

- 单整
  - $x_t \sim I(0)$：如果序列平稳，说明序列不存在单位根，序列为零阶单整序列
  - $x_t \sim I(d)$：原序列至少需要进行 d 阶差分才能实现平稳，说明原序列存在 d 个单位根，原序列为 d 阶单整序列
  - 若 $x_t \sim I(d)$，则 $\forall a,b \neq 0$，有 $a + bx_t \sim I(d)$
  - 若 $x_t \sim I(0), y_t \sim I(0)$，则 $\forall a,b \neq 0$，有 $z_t = ax_t + by_t \sim I(0)$
  - 若 $x_t \sim I(d), y_t \sim I(d)$，则 $\forall a,b \neq 0$，有 $z_t = ax_t + by_t \sim I(k)$，$k \leq \max(d,c)$

协整

- 协整的概念与检验
  - 自变量序列：$\{x_1\},\cdots,\{x_k\}$，响应变量序列：$\{y_t\}$
  - 回归模型：$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{it} + \varepsilon_t$
  - $\{y_t\}$ 与 $\{x_1\},\cdots,\{x_k\}$ 协整：回归残差序列 $\{\varepsilon_t\}$ 平稳
  - 若非平稳序列之间具有协整关系，不会产生伪回归问题
  - H_0 : 多元序列之间不存在协整关系 ($\{\varepsilon_t\}$ 非平稳，$\varepsilon_t \sim I(k)$)，H_1 : 多元序列之间存在协整关系 ($\{\varepsilon_t\}$ 平稳，$\varepsilon_t \sim I(0)$)
  - EG 检验：

    $y_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \cdots + \hat{\beta}_k x_{kt} + \varepsilon_t$

    对回归残差序列进行平稳性检验

协整

- 误差修正模型
  - 形式：$\nabla y_t = \beta \nabla x_t - \mathrm{ECM}_{t-1} + \varepsilon_t$
  - ECM_{t-1} 是上一期的误差
  - 误差修正系数：$\hat{\beta}_1$
  - 负反馈机制：$\beta_1 < 0$
  - 当 $\mathrm{ECM}_{t-1} > 0 (< 0)$ 时，即上期真实支出比估计支出大 (小)，这种误差反馈回来，会导致下期支出适当压缩 (增加)，即 $\nabla y_t < 0 (> 0)$

Ganger 因果检验

- Ganger 因果关系定义
  - x 是 y 的 Ganger 原因：$\sigma^2(y_{t+1}|I_t) < \sigma^2(y_{t+1}|I_t - X_t)$
  - $I_t = \{x_t, x_{t-1}, \cdots, y_t, y_{t-1}, \cdots\}$
  - x 和 y 独立：(x, y)
  - x 是 y 的 Ganger 原因：$(x \rightarrow y)$
  - y 是 x 的 Ganger 原因：$(x \leftarrow y)$
  - x 和 y 互为因果 (存在反馈关系)：$(x \leftrightarrow y)$

Ganger 因果检验

- Ganger 因果检验
  - $y_t = \beta_0 + \sum_{k=1}^{p} \beta_k y_{t-k} + \sum_{k=1}^{q} \alpha_k x_{t-k} + \sum_{k=1}^{l} \gamma_k z_{t-k} + \varepsilon_t$
  - $H_0 : (x, y)$ 或 $\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$，$H_1 : x \to y$ 或 $\exists \alpha_i \neq 0$
  - 统计量：

    原假设成立下：$y_t = \beta_0 + \sum_{k=1}^{p} \beta_k y_{t-k} + \sum_{k=1}^{l} \gamma_k z_{t-k} + \varepsilon_{1t}$，得

    $SST = SSR_{yz} + SSE_r$

    备择假设成立下：$y_t = \beta_0 + \sum_{k=1}^{p} \beta_k y_{t-k} + \sum_{k=1}^{q} \alpha_k x_{t-k} + \sum_{k=1}^{l} \gamma_k z_{t-k} + \varepsilon_t$，

    得 $SST = SSR_{xyz} + SSE_u$

    $$F = \frac{(SSE_r - SSE_u)/q}{SSE_u/(n-p-q-1)} \sim F(q, n-p-q-1)$$
  - Granger 因果检验即使显著拒绝原假设，也不能说明两个序列间具有真正的因果关系