五、自相关性

目录

自相关性及其产生的原因

- 自相关性及其产生的原因
  - 自相关性 (序列相关性)：$\mathrm{Cov}(u_t, u_s) \neq 0$
  - 一阶自相关性：$u_t = \rho u_{t-1} + v_t$；$\hat{\rho} = \dfrac{\sum u_t u_{t-1}}{\sum u_{t-1}^2}$
  - p 阶自相关：$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + v_t$
  - 产生原因：

    经济变量惯性的作用；

    经济行为的滞后性；

    随机偶然因素的干扰；

    模型设定误差；

    观测数据处理不当

自相关性的影响

- 一元线性回归模型下自相关性的影响 (OLS)
  - $\hat{b}_1, \hat{b}_0$ 是 b_1, b_0 的无偏估计量，保持线性
  - $\hat{b}_1$ 不再有最小方差，$\operatorname{Var}(\hat{b}_1^*) = \operatorname{Var}(\hat{b}_1) + 2\sum\limits_{t \neq s} k_t k_s E(u_t u_s) > \operatorname{Var}(\hat{b}_1)$
  - $E\left(\dfrac{\sum e_t^2}{n-2}\right) < \sigma^2$，低估了随机项的方差
  - RSS 虚假缩小，ESS 虚假增大，$F = \dfrac{\text{ESS}/k}{\text{RSS}/(n-k-1)}$ 虚假增大 (高估)，检验失效

- 自相关性的检验
  - 图示法：

    绘制 tOe_t 图 (残差图) 与 $e_{t-1}Oe_t$ 散点图；

    若存在有规律的变动或系统性变动，表明存在自相关性
  - 德宾-沃森检验 (DW 检验)：$H_0:\rho=0$ (不存在一阶自相关性)，

    $H_1:\rho\neq 0$ (存在一阶自相关性)

    统计量 $\mathrm{DW}=\dfrac{\sum\limits_{t=2}^{n}(e_t-e_{t-1})^2}{\sum\limits_{t=1}^{n}e_t^2}\approx 2(1-\hat{\rho})$；

    $\mathrm{DW}\in[0,4]$，下极限分布 d_L 的上临界值为 $4-d_U$，上极限分布 d_U 的下临界值为 $4-d_L$

自相关性的检验

- 自相关性的检验
  - 德宾-沃森检验 (DW 检验):

    $0 \leq \mathrm{DW} \leq d_L$，拒绝 H_0，认为存在一阶正自相关；

    $4-d_L \leq \mathrm{DW} \leq 4$，拒绝 H_0，认为存在一阶负自相关；

    $d_U \leq \mathrm{DW} \leq 4-d_U$，接受 H_0，认为不存在一阶自相关；

    $d_L < \mathrm{DW} < d_U$ 或 $4-d_U < \mathrm{DW} < 4-d_L$，不能确定；

    只能用于判断一阶自相关性，有两个无法判定的区域，不能用于滞后被解释变量 y_{t-i}，要求 $n \geq 15$

  - DW 检验的修正 (h 检验):

    Durbin-h 统计量 $h = \left(1-\frac{\mathrm{DW}}{2}\right)\sqrt{\frac{n}{1-n\mathrm{Var}(\hat{b}_2)}} \to N(0,1)$

自相关性的检验

- 自相关性的检验
  - 回归检验法：根据 OLS 法求 e_t，对 e_t 建立回归模型
  - 高阶自相关性检验：

    相关图检验：自相关系数 $r_k = \dfrac{\sum\limits_{t=k+1}^{n}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sum\limits_{t=1}^{n}(y_t-\bar{y})^2} \in [-1,1]$，偏自

    相关系数 $\varphi_{k,k} = \begin{cases} r_1, & k=1 \\ \dfrac{r_k-\sum\limits_{j=1}^{k-1}\varphi_{k-1,j}r_{k-j}}{1-\sum\limits_{j=1}^{k-1}\varphi_{k-1,j}r_k}, & k=2,3,\cdots \end{cases}$,

    $\varphi_{k,j} = \varphi_{k-1,j} - \varphi_{k,k}\varphi_{k-1,k-j}, j=1,2,\cdots,k-1$;

- 自相关性的检验
  - 高阶自相关性检验：

    Q 统计量检验：$Q_{LB} = n(n+2)\sum_{j=1}^{p}\frac{r_j^2}{n-j} \sim \chi^2(n-p)$；

    $Q_{LB} < \chi_\alpha^2(n-p)$，接受 H_0，不存在 p 阶自相关；$Q_{LB} > \chi_\alpha^2(n-p)$，接受 H_0，存在 p 阶自相关；

    LM 乘数检验：根据 OLS 法求 e_t，对辅助回归模型

    $e_t = b_0 + b_1x_{1t} + \cdots + b_kx_{kt} + \rho_1e_{t-1} + \rho_2e_{t-2} + \cdots + \rho_pe_{t-p} + v_t$ 进行回归分析，计算决定系数 R^2，统计量 $\mathrm{LM}(p) = nR^2 \sim \chi^2(p)$

自相关性的解决方法

- 广义差分法
  - 一元一阶自相关：

    广义差分变换 $y_t^* = y_t - \rho y_{t-1}$，$x_t^* = x_t - \rho x_{t-1}$，$A = b_0(1-\rho)$，则

    $y_t^* = A + b_1 x_t^* + v_t$；

    $\hat{b}_0 = \dfrac{A}{1-\rho}$；

    损失的观测值可以作变换 $y_1^* = y_1\sqrt{1-\rho^2}, x_1^* = x_1\sqrt{1-\rho^2}$；

    $\rho = 1$，则此时是差分变换

自相关性的解决方法

- 广义差分法
  - 多元一阶自相关：

    广义差分变换 $y_t^* = y_t - \rho y_{t-1}$，$x_{it}^* = x_{it} - \rho x_{i,t-1}$，$A = b_0(1-\rho)$，则

    $y_t^* = A + b_1 x_{1t}^* + b_2 x_{2t}^* + \cdots + b_k x_{kt}^* + v_t$
  - 一元 p 阶自相关：

    广义差分变换 $y_t^* = y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} - \cdots - \rho_p y_{t-p}$，

    $x_t^* = x_t - \rho_1 x_{t-1} - \rho_2 x_{t-2} - \cdots - \rho_p x_{t-p}$，则 $y_t^* = A + b_1 x_t^* + v_t$

自相关性的解决方法

- 自相关系数的估计方法
  - 广义差分法：

    大样本：$\hat{\rho} = 1 - \dfrac{\mathrm{DW}}{2}$；

    小样本：$\hat{\rho} = \dfrac{n^2(1-\mathrm{DW}/2)+(k+1)^2}{n^2-(k+1)^2} \xrightarrow{n\to\infty} 1 - \dfrac{\mathrm{DW}}{2}$；

    也可以是 $\hat{\rho} = \dfrac{\sum e_t e_{t-1}}{\sum e_t^2}$
  - Durbin 两步估计法：

    第一步：对 $y_t = a_0 + \rho y_{t-1} + a_1 x_t + a_2 x_{t-1} + v_t$ 进行回归分析，y_{t-1} 的回归系数就是 ρ；

    第二步：用 $\hat{\rho} = \rho$ 进行广义差分变换，求得原模型

自相关性的解决方法

- 自相关系数的估计方法
  - 迭代估计：
    利用 OLS 法进行回归分析得残差 $e_t(1)$，根据 $\hat{\rho}(1) = \dfrac{\sum e_t(1)e_{t-1}(1)}{\sum e_t^2(1)}$，利用 $\hat{\rho}(1)$ 进行广义差分变换，求得 $e_t(2)$，重复上述步骤，直到 $|\hat{\rho}(n+1) - \hat{\rho}(n)| < \delta$
  - 搜索估计法：略
- 广义最小二乘法与广义差分法的关系
  - 对存在自相关性的模型，广义最小二乘法与广义差分法是等价的
  - $\hat{\boldsymbol{B}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{Y}$ 是无偏、有效估计量

代码输出结果分析

- 回归分析结果

同第二章：

常数和解释变量	参数估计值	参数标准误差	t统计量	双侧概率
C(b_0)	331.5264	57.16954	5.799003	0.0000
PI(b_1)	0.692812	0.006279	110.3337	0.0000
决定系数	0.997297	被解释变量均值		4662.514
调整的决定系数	0.997215	被解释变量标准差		4659.100
回归标准误差	245.8925	赤池信息准则		13.90311
残差平方和	1995283.	施瓦兹信息准则		13.99199
对数似然函数	−241.3044	汉南准则		13.93379
F统计量	12173.53	DW统计量		0.180221
F统计量的概率	0.000000			

代码输出结果分析

- 各种检验的输出结果分析

同第四章：

英文	含义	英文	含义
`Heterpskedasticity Test`	检验方法	`F-statistic`	回归模型的 F 统计量
`Obs*R-squared`	F 检验统计量	`Prob. Chi-Square(2)`	F 统计量对应的 p 值
`Prob. F(a, b)`	自由度为 a,b 的 F 分布临界值	`Scaled explained SS`	LM 统计量