



计量经济学II (Econometrics II)

胡华平

西北农林科技大学

经济管理学院数量经济教研室

huhuaping01@hotmail.com

2022-09-25

西北农林科技大学

模块1：计量经济学基础

Chapter 01. 经典模型

Chapter 02. 矩阵分析

Chapter 03. 放宽假设

Chapter 04. 扩展方法

CLRM
NCLARM

放宽

对 X_i 放宽:

- ① 多重共线性
- ② X_i 随机

对 u_i 放宽:

- ① 异方差
- ② 自相关

OLS

OLS

其他



BLUE

BLUE

定义

影响

诊断

纠正

多元回归

BLUE

辅助回归

变量变换

LE

内生性/
外生性

IV

$E(u_i^2) = \sigma_i^2$

LUE

e_i e_i^2

WLS

$E(u_t u_{t-s}) \neq 0$

LUE

e_t e_{t-s}

GDE

3.3 自相关 (Auto-correlation) 问题

3.3.1 序列自相关性的定义和来源

3.3.2 序列自相关性的影响和后果

3.3.3 序列自相关性问题的诊断

3.3.4 序列自相关性问题的矫正

3.3.1 序列自相关性的定义和来源



自相关的概念与内涵

滞后变量 (lag variable) : 对某个时间序列变量进行**滞后变换**后得到的新变量。

例如对时间序列变量 $Z_{2,t}$ ($t \in 1, 2, \dots, T$) 进行滞后变换可以得到多个**滞后变量**: $Z_{2,t-1}; Z_{2,t-2}; \dots; Z_{2,t-p}; \dots; Z_{2,t-(T-1)}$ 。



自相关的概念与内涵

时间序列数据的两种相关关系：

- **序列相关** (serial correlation)：两个不同时间序列变量之间（或/及它们的滞后变量之间）的（线性）相关关系。

$$\text{cor}(Z_{2,t}, Z_{3,t})$$

$$\text{cor}(Z_{2,t}, Z_{3,t-1}); \dots; \text{cor}(Z_{2,t}, Z_{3,t-p}). \quad p \in 1, 2, \dots, T-1$$

$$\text{cor}(Z_{2,t-1}, Z_{3,t}); \dots; \text{cor}(Z_{2,t-p}, Z_{3,t}). \quad p \in 1, 2, \dots, T-1$$

- **自相关** (autocorrelation)：某个时间序列变量与其自身滞后变量的（线性）相关关系。

$$\text{cor}(Z_{2,t}, Z_{2,t-1}); \text{cor}(Z_{2,t}, Z_{2,t-2}); \dots; \text{cor}(Z_{2,t}, Z_{2,t-p}). \quad p \in 1, 2, \dots, T-1$$



自相关的概念与内涵

k变量总体回归模型(PRM)

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t$$

经典线性回归模型假定 (CLRM) 在随机干扰项 u_i 之间不存在自相关, 也即:

$$E(u_t u_{t-p}) = 0; \quad p \in 1, 2, \dots, T-1$$

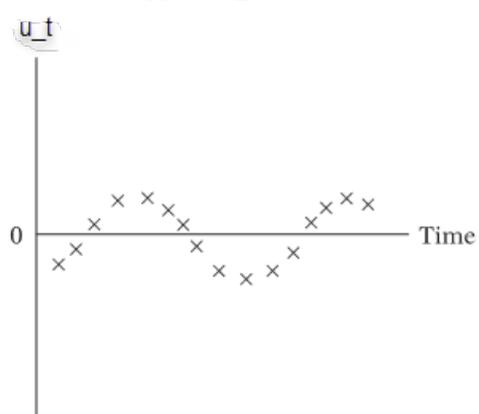
如果随机干扰项 u_i 出现自相关时, 则意味着**违背**了CLRM假设 (其他假设仍旧成立) :

$$E(u_t u_{t-p}) \neq 0; \quad p \in 1, 2, \dots, T-1$$

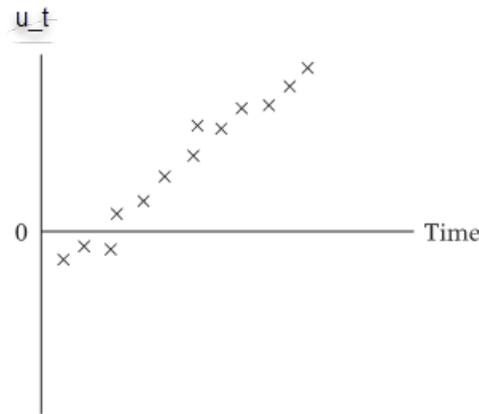


自相关的概念与内涵 (模拟演示)

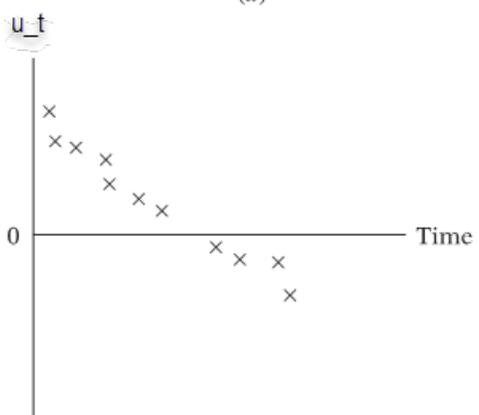
随机干扰项 u_t 的几种自相关模式和非自相关模式可以表示为:



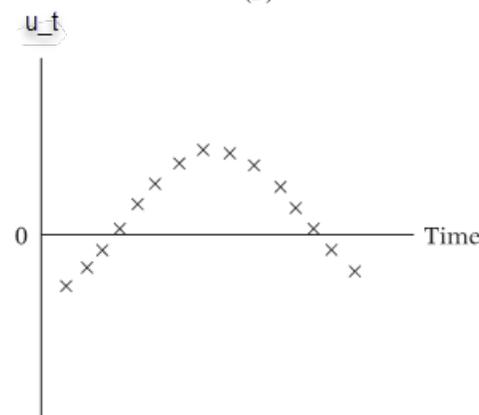
(a)



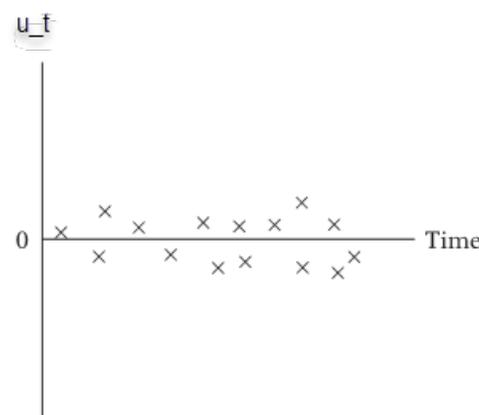
(b)



(c)



(d)



(e)



产生自相关的原因1

产生自相关的原因1：时间惯性的普遍存在。大多数经济时间序列变量都有时间惯性。

- GNP、价格指数、生产、就业和失业等时间序列变量都呈现出商业循环。



产生自相关的原因2

产生自相关的原因2-1: 模型设定偏误——应含而未含某些重要变量 (excluded variables) 。

- 以牛肉需求模型为例。其中： Y_t 表示牛肉需求量； X_{2t} 表示牛肉价格； X_{3t} 表示消费者收入； X_{4t} 表示猪肉价格。

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$$

- 如果我们把上述模型做成如下情形，在猪肉价格影响牛肉消费的情形下，新模型的残差 v_t 将表现出某种系统的模式。也即模型出现了错误设定，导致随机干扰项不满足CLRM假设。

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + v_t$$

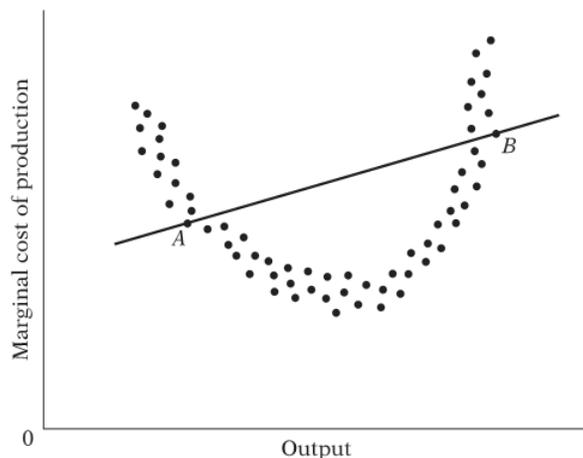


产生自相关的原因2

产生自相关的原因2-2: 模型设定偏误——不正确的函数形式。

- 以企业边际成本-产出模型为例。
如果“真实”的边际成本决定模型为:

$$Cost_t = \beta_1 + \beta_2 Output_t + \beta_3 Output_t^2 + u_t$$



- 如果我们把上述模型做成如下情形（模型错误设定），由于模型函数形式的错误使用，残差 v_t 将反映出自相关性质（因为 $v_t = \beta_3 Output_t^2 + u_t$ ）。

$$Cost_t = \alpha_1 + \alpha_2 Output_t + v_t$$

西北农林科技大学
NORTHWEST A&F UNIVERSITY



产生自相关的原因3

产生自相关的原因3：蛛网现象（Cobweb phenomenon）的存在。

- 以农产品供给模型为例。如果在当期（t期）价格 P_t 的低于前一期（t-1期）价格 P_{t-1} ，那么在**未来1期**（t+1期），农户将会决定生产更少的农产品。如此往复决策，将会使得形成蛛网生产和价格模式。

$$Supply_t = \beta_1 + \beta_2 P_{t-1} + u_t$$



产生自相关的原因4

产生自相关的原因4：滞后效应的普遍存在。

- 以消费支出模型为例。当期消费（t期）还会受到前期（t-1期）消费水平的影响。这种带有因变量的滞后变量 $Consumption_{t-1}$ 的回归也叫自回归（auto-regression）。

$$Consumption_t = \beta_1 + \beta_2 Income_t + \beta_3 Consumption_{t-1} + u_t$$



产生自相关的原因5

产生自相关的原因5：数据的“编造”。

-从月度数据计算得出季度数据，会减小波动，引进匀滑作用，使扰动项出现系统性模式

- 数据的内插（interpolation）：人口普查10年一次
- 数据的外推（extrapolation）



产生自相关的原因6

产生自相关的原因6：数据的“变换”。

- 例如，有时候我们可以构造如下的一阶差分模型，这类模型也被称为动态回归模型（Dynamic regression）。

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_t + u_t \\ Y_{t-1} &= \beta_1 + \beta_2 X_{t-1} + u_{t-1} \\ \Delta Y_t &= \beta_2 \Delta X_t + \Delta u_t \\ \Delta Y_t &= \beta_2 \Delta X_t + v_t \end{aligned}$$

- 如果 Y_t 和 X_t 都是已经经过了对数化处理，则一阶差分变换后的模型中随机干扰项 v_t 将会出现自相关模式。



产生自相关的原因7

产生自相关的原因7：时间序列非平稳性的广泛存在。

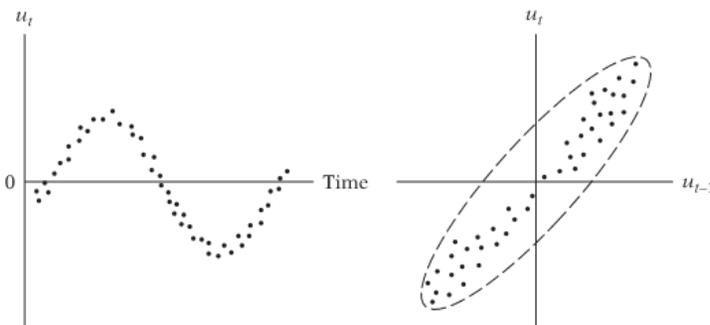
$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

- 因变量 Y_t 和自变量 X_t 很可能都是非平稳的。因此随机干扰项 u_t 也是非平稳的。此时，随机干扰项将表现出自相关。

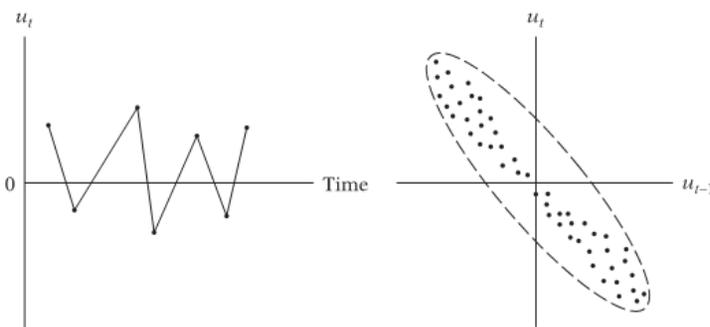


自相关的几种模式：视角1

直接观察视角：我们可以从随机干扰项 u_t 与其滞后变量 $u_{t-1}, u_{t-2}, \dots, u_{t-p}, \dots, u_{t-(T-1)}$ 的散点图来观察自相关模式（比较容易理解）。



(a)



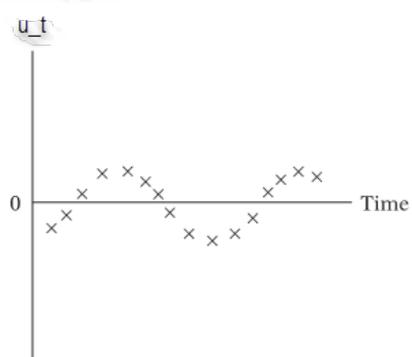
(b)



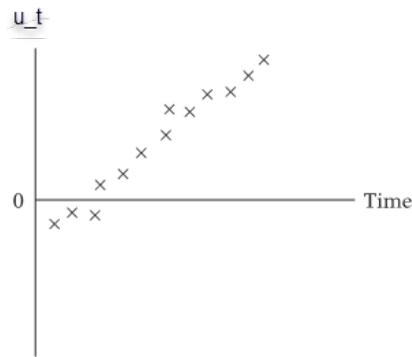


自相关的几种模式：视角2

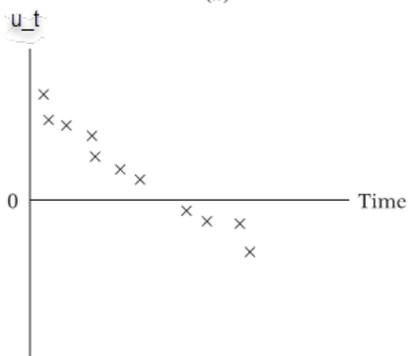
间接观察视角：我们可以从随机干扰项 u_t 随时间 t 的散点图变化来观察自相关模式（不太好理解）。



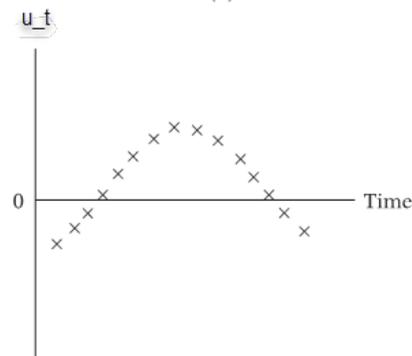
(a)



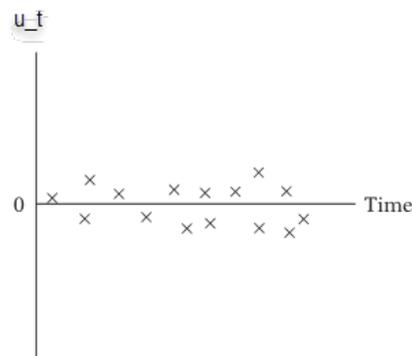
(b)



(c)



(d)



(e)



3.3.2 序列自相关性的影响和后果



马尔可夫阶自回归

下面给出自相关情形为马尔可夫1阶自回归模式:

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \quad \leftarrow [-1 < \rho < 1] \end{aligned}$$

- ρ 被称为自协方差系数(coefficient of autocovariance)

$$\rho = \frac{E[(u_t - E(u_t))(u_{t-1} - E(u_{t-1}))]}{\sqrt{\text{var}(u_t)}\sqrt{\text{var}(u_{t-1})}} = \frac{E(u_t u_{t-1})}{\text{var}(u_{t-1})}$$

- ε_t 是满足以下标准CLRM假定的随机干扰项。此时 ε_t 也被成为白噪声误差项(white noise error Term)。

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ \text{var}(\varepsilon_t) &= \sigma^2 \\ \text{cov}(\varepsilon_t, \varepsilon_{t+s}) &= 0; \quad s \neq 0 \end{aligned}$$



出现自相关性时的态度I

态度1：忽视自相关性，坚持错误地使用CLRM假设下OLS方法的各种公式。

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \quad \leftarrow [-1 < \rho < 1] \end{aligned}$$

在自相关性存在的情形下（马尔科夫1阶自相关），却坚持使用OLS方法下的方差公式：

$$\hat{\beta}_2 \Big|_{OLS}^{CLRM} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{n \sum X_t Y_t - \sum X_t \sum Y_t}{n \sum X_t^2 - (\sum X_t)^2}$$

$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM} = \frac{\sigma^2}{\sum x_t^2}$$

- 坚持使用的方差公式 $\text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM}$ 是有偏的，可能高估或低估其真实方差 $\text{var}(\hat{\beta}_2)$ 。



出现自相关性时的态度?

态度2: 承认“自相关性”这一事实, 但仍旧使用OLS方法。

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \quad \leftarrow [-1 < \rho < 1] \end{aligned}$$

在自相关性存在的情形下 (马尔科夫1阶自相关), 仍旧使用OLS方法, 得到估计量及其方差的“新公式” (细节见下一页slide):

$$\hat{\beta}_2 \Big|_{OLS}^{AR1} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{n \sum X_t Y_t - \sum X_t \sum Y_t}{n \sum X_t^2 - (\sum X_t)^2}$$

$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{AR1} = \frac{\sigma^2}{\sum x_t^2} \frac{1+r\rho}{1-r\rho} = \text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM} \cdot \frac{1+r\rho}{1-r\rho}$$

- 系数估计量仍是一致的; 方差公式是有偏的, 可能高估或低估其真实方差 $\text{var}(\hat{\beta}_2)$ 。
- 请验证 $r = 0.6, \rho = 0.8$ 时, 二者的大小关系?



附录：方差计算细节

给定马尔科夫1阶自相关情形 (AR(1)) :

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \quad \leftarrow [-1 < \rho < 1] \end{aligned}$$

$$\text{var}(u_t) = E(u_t^2) = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

$$\text{cov}(u_t, u_{t-s}) = \text{cov}(u_t, u_{t+s}) = E(u_t u_{t-s}) = \rho^s \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

$$\text{cor}(u_t, u_{t-s}) = \text{cor}(u_t, u_{t+s}) = \rho^s$$



附录：方差计算细节

$$\begin{aligned}\text{var} \left(\hat{\beta}_2 \right)_{OLS}^{AR1} &= \frac{\sigma^2}{\sum x_t^2} \left[1 + 2\rho \frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2 \frac{\sum x_t x_{t-2}}{\sum x_t^2} + \dots + 2\rho^{n-1} \frac{x_1 x_n}{\sum x_t^2} \right] \\ &= \frac{\sigma^2}{\sum x_t^2} \frac{1 + r\rho}{1 - r\rho} \quad \leftarrow [if \quad X_t = rX_{t-1} + v_t]\end{aligned}$$

- 若 $\rho = 1$ ，上述方差和协方差将不能定义；若 $|\rho| < 1$ ， u_t 的均值、方差和协方差都不随时间而变化。AR(1)过程被称为是**平稳的**。此时，协方差的值将随着两个误差的时间间隔越远而越小。



出现自相关性时的态度3

态度3：在自相关性存在情形下，首先想办法消除自相关性，再使用OLS方法。广义最小二乘法（GLS）下（如广义差分方程法），估计量及其方差公式可写成：

$$\hat{\beta}_2 \Big|_{\text{GLS}}^{\text{AR1}} = \frac{\sum_{t=2}^n (x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^n (x_t - \rho x_{t-1})^2} + C \quad \leftarrow [C \text{ is a correction factor}]$$

$$\text{var}(\hat{\beta}_2)_{\text{GLS}}^{\text{AR1}} = \frac{\sigma^2}{\sum_{t=2}^n (x_t - \rho x_{t-1})^2} + D \quad \leftarrow [D \text{ is a correction factor}]$$

- 系数估计量是一致的
- 方差公式是无偏的，其期望将等于真实方差 $\text{var}(\hat{\beta}_2)$ 。
- 此时，GLS得到的才是**BLUE**！（证明过程略）

提示：GLS中我们通过变量变换把额外的信息（异方差性或自相关性）包括到估计程序中去，而在OLS中

我们并不直接考虑这种附加信息。



出现自相关性时的态度：总结

- 态度1：“把头埋进沙堆的鸵鸟”

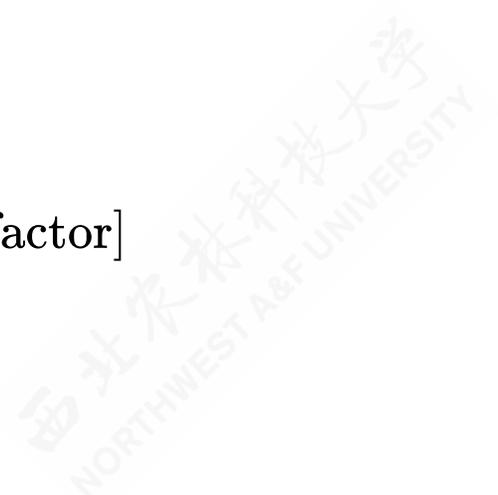
$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM} = \frac{\sigma^2}{\sum x_t^2}$$

- 态度2：“将错就错地走下去”

$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{AR1} = \frac{\sigma^2}{\sum x_t^2} \frac{1+r\rho}{1-r\rho} = \text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM} \cdot \frac{1+r\rho}{1-r\rho}$$

- 态度3：“直面困难找出路”

$$\text{var}(\hat{\beta}_2)_{GLS}^{AR1} = \frac{\sigma^2}{\sum_{t=2}^n (x_t - \rho x_{t-1})^2} + D \quad \leftarrow [D \text{ is a correction factor}]$$





出现自相关时使用OLS的后果

在自相关出现时，OLS 估计量仍是线性的、无偏的和一致性的，但不再是有效的（亦即最小方差）。那么，如果我们继续使用OLS估计量，我们平常的假设检验程序会遇到什么问题呢？主要有：

- 参数估计不再是有效估计量(亦即不再是方差最小)
- 参数的显著性检验失去意义
- 模型的预测失效

下面分两种情形来讨论其后果：

- **态度1**：忽视自相关，采用OLS估计
- **态度2**：考虑自相关，采用OLS估计



出现自相关时使用OLS的后果 (态度1和态度2)

对于态度1和态度2, 执意采用OLS估计, 将会: 残差方差可能低估真实方差; 高估可绝系数; 或低估考虑一阶自回归的方差; 检验无效, 得出错误的结论。

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad \leftarrow [-1 < \rho < 1]$$

CLRM条件下采用OLS估计:

$$\hat{\beta}_2 \Big|_{OLS}^{CLRM} = \frac{\sum x_t y_t}{\sum x_t^2}$$

$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM} = \frac{\sigma^2}{\sum x_t^2}$$

$$\hat{\sigma}^2 \Big|_{OLS}^{CLRM} = \frac{\sum e_i^2}{n-2}$$

$$E(\hat{\sigma}^2 \Big|_{OLS}^{CLRM}) = \sigma^2$$

AR(1)情形下使用OLS估计:

$$\hat{\beta}_2 \Big|_{OLS}^{AR1} = \frac{\sum x_t y_t}{\sum x_t^2}$$

$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{AR1} = \frac{\sigma^2}{\sum x_t^2} \frac{1+r\rho}{1-r\rho}$$

$$\hat{\sigma}^2 \Big|_{OLS}^{AR1} = \frac{\sigma^2 \left\{ n - \left[\frac{2}{(1-\rho)} - 2\rho r \right] \right\}}{n-2}$$

$$E(\hat{\sigma}^2 \Big|_{OLS}^{AR1}) < \sigma^2$$



出现自相关时使用OLS的后果

实际经济分析中，正自相关性一般更为普遍，也即 $(0 < \rho < 1; 0 < r < 1)$ 。此时：

$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{AR1} = \frac{\sigma^2}{\sum x_t^2} \frac{1 + r\rho}{1 - r\rho} = \text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM} \cdot \frac{1 + r\rho}{1 - r\rho}$$

$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{AR1} > \text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM} \quad \leftarrow [\text{if } 0 < \rho < 1; 0 < r < 1]$$

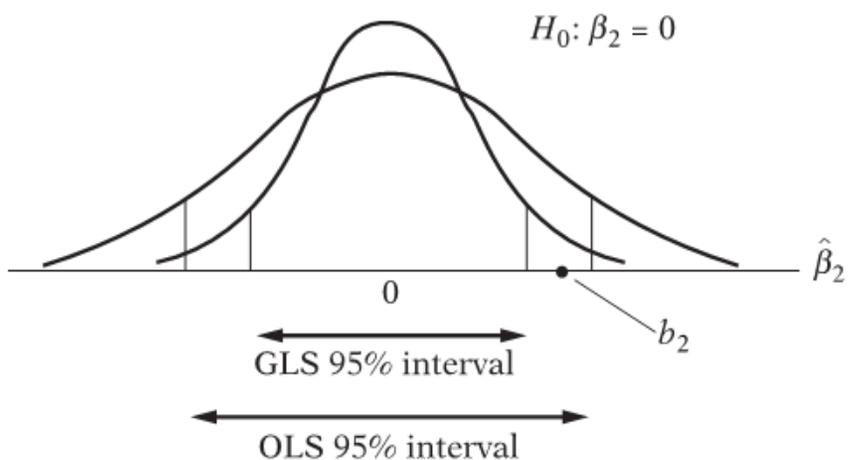


出现自相关时使用OLS的后果

事实上，在AR(1)情形下采用广义最小二乘法（GLS）才能得到BLUE，意味着：

$$\text{var}(\hat{\beta}_2) \Big|_{OLS}^{AR1} > \text{var}(\hat{\beta}_2) \Big|_{OLS}^{CLRM} > \text{var}(\hat{\beta}_2) \Big|_{GLS}^{AR1}$$

同时，也将意味着GLS方法下斜率系数的置信区间也将更窄：



- 如果一个对斜率参数 β_2 的点估计值为 b_2 ，并给定显著性水平为 $\alpha = 0.05$
- 那么则有可能出现如下结果：在OLS方法下，95%置信区间检验**不显著**（接受原假设 H_0 ）；但是在GLS方法下，95%置信区间检验却可以是**显著**的（拒绝原假设 H_0 ，接受备择假设 H_1 ）



蒙特卡洛模拟 (数据生成I)

给定如下存在自相关情形AR(1)的模拟数据表 (观测数 $T = 9$) :

$$Y_t = 1 + 0.8X_t + u_t$$

$$u_t = 0.7u_{t-1} + \varepsilon_t \quad \leftarrow [u_0 = 5]$$

$$\varepsilon_t \sim N(0, 1) \quad \leftarrow [\varepsilon_0 = 0]$$

ε_t	u_t	X_t	Y_t	$E(Y X_t)$	Y_t^*
0.0000	5.0000				
-0.5605	2.9395	1	4.7395	0.8000	0.2395
-0.2302	1.8275	2	4.4275	0.6400	1.3698
1.5587	2.8380	3	6.2380	0.5120	3.9587
0.0705	2.0571	4	6.2571	0.4096	3.2705
0.1293	1.5692	5	6.5692	0.3277	4.1293

Showing 1 to 6 of 10 entries

Previous

1

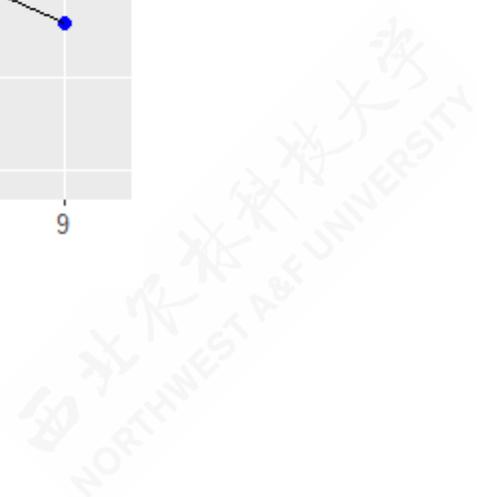
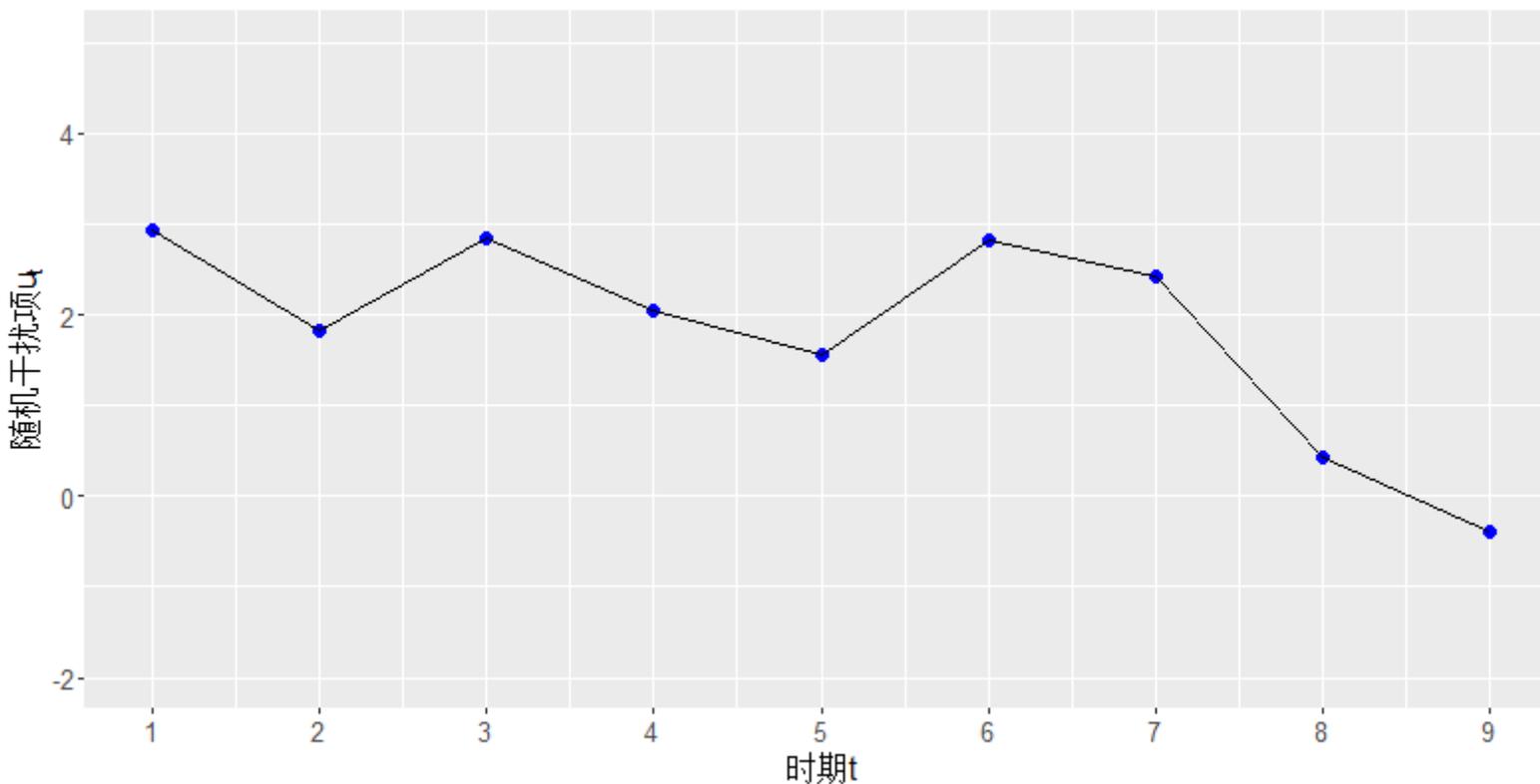
2

Next



蒙特卡洛模拟 (随机干扰项模式1)

存在自相关情形AR(1)的模拟数据，随机干扰项 u_t 的分布模式如下：





蒙特卡洛模拟 (回归分析I)

对以上存在AR(1)的模拟数据，我们构建如下的回归模型，并坚持采用OLS方法得到估计结果：

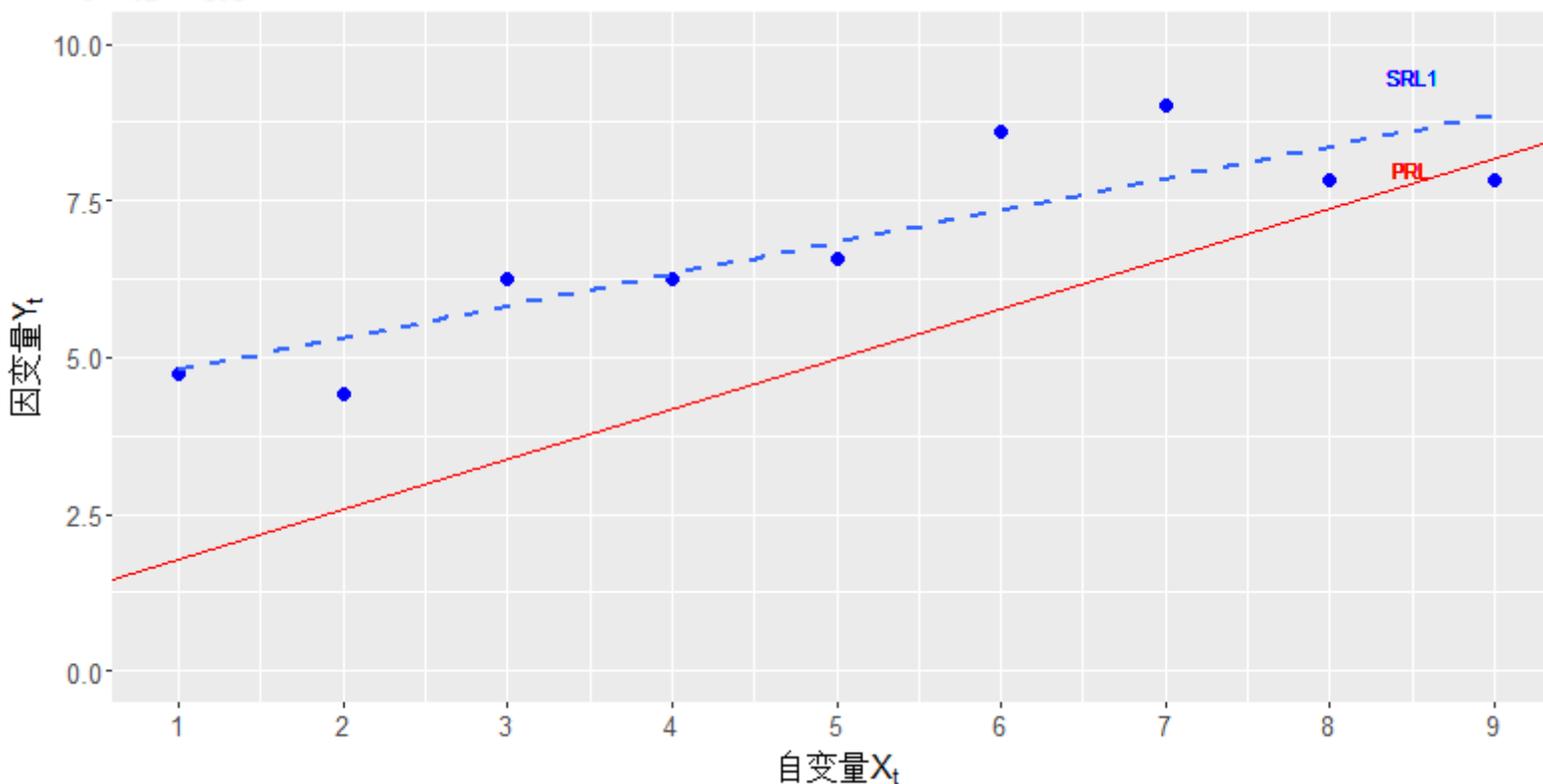
$$Y_t = + \beta_1 + \beta_2 X_t + u_t$$

$$\begin{aligned} \widehat{Y}_t &= + 4.30 && + 0.51 X_t \\ (t) & (6.7249) && (4.4750) \\ (se) & (0.6389) && (0.1135) \\ (\text{fitness}) & R^2 = 0.7410; \bar{R}^2 = 0.7040 \\ & F^* = 20.03; p = 0.0029 \end{aligned}$$



蒙特卡洛模拟 (回归线I)

对OLS方法下的样本回归线 (SRL1) 绘图, 并与真实的总体回归线 (PRL) 进行对比:



$$E(Y|X_t) = 1 + 0.8X_t \quad (PRL/PRF)$$



蒙特卡洛模拟 (数据生成2)

如果我们使用符合CLRM假设的模拟数据 (Y_t^*, X_t) , 其中:

$$Y_t^* = 1 + 0.8X_t + \varepsilon_t$$
$$\varepsilon_t \sim N(0, 1) \quad \leftarrow [\varepsilon_0 = 0]$$

ε_t	u_t	X_t	Y_t	$E(Y X_t)$	Y_t^*
0.0000	5.0000				
-0.5605	2.9395	1	4.7395	0.8000	0.2395
-0.2302	1.8275	2	4.4275	0.6400	1.3698
1.5587	2.8380	3	6.2380	0.5120	3.9587
0.0705	2.0571	4	6.2571	0.4096	3.2705
0.1293	1.5692	5	6.5692	0.3277	4.1293

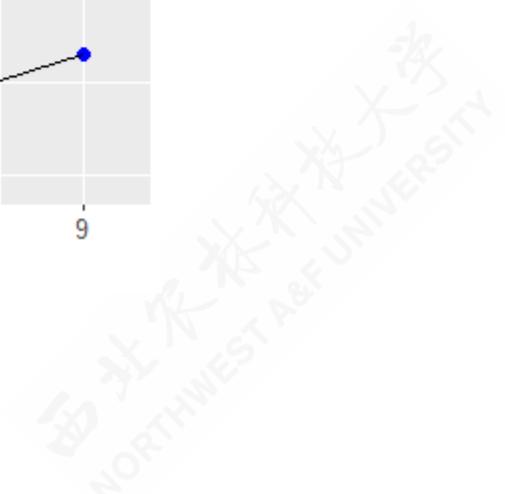
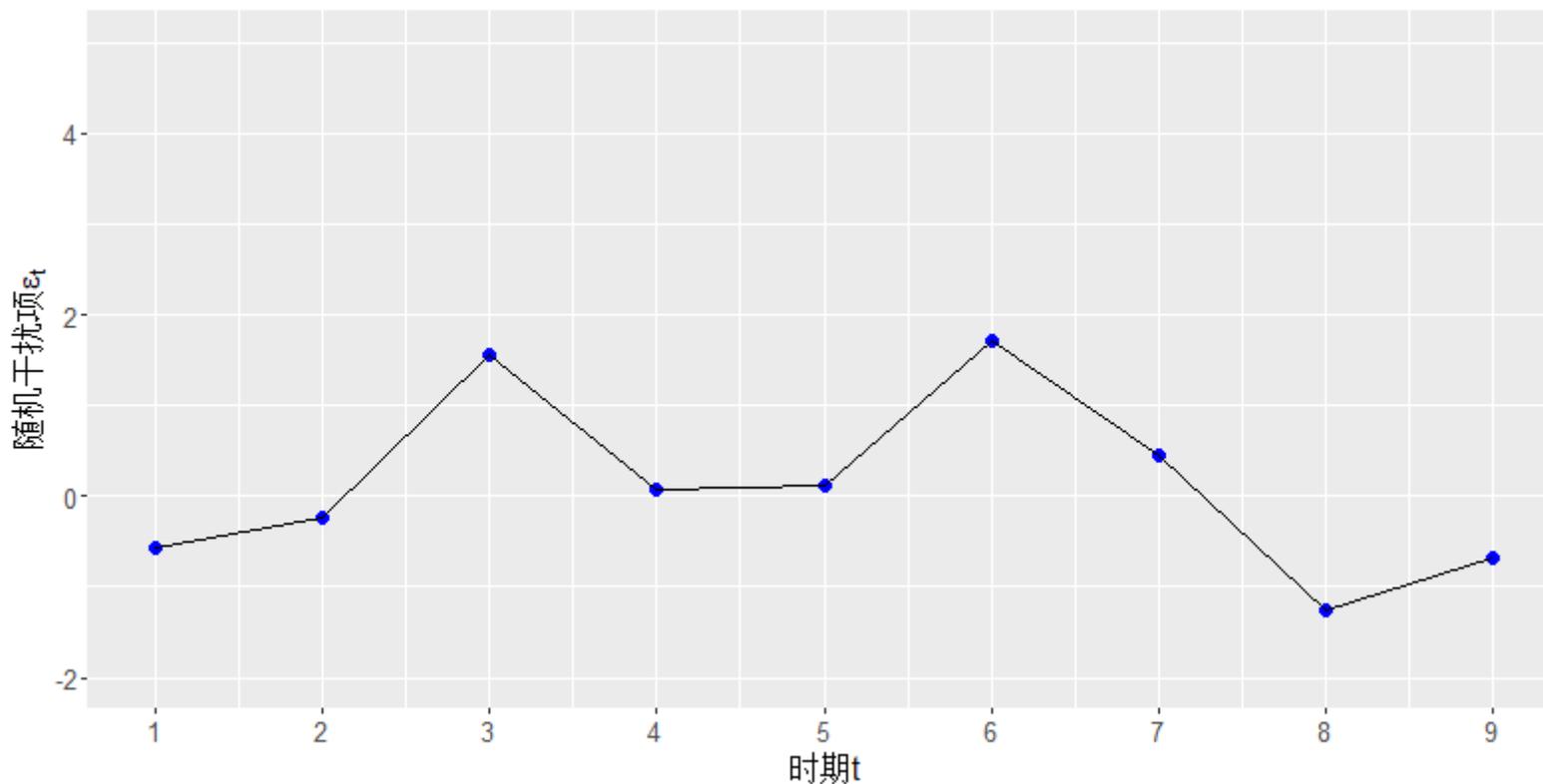
Showing 1 to 6 of 10 entries

Previous 1 2 Next



蒙特卡洛模拟 (随机干扰项模式2)

符合CLRM假设的模拟数据，随机干扰项 ε_t 的分布模式如下：





蒙特卡洛模拟 (回归分析2)

对以上符合CLRM假设模拟数据，我们构建如下的回归模型，并采用OLS方法得到估计结果：

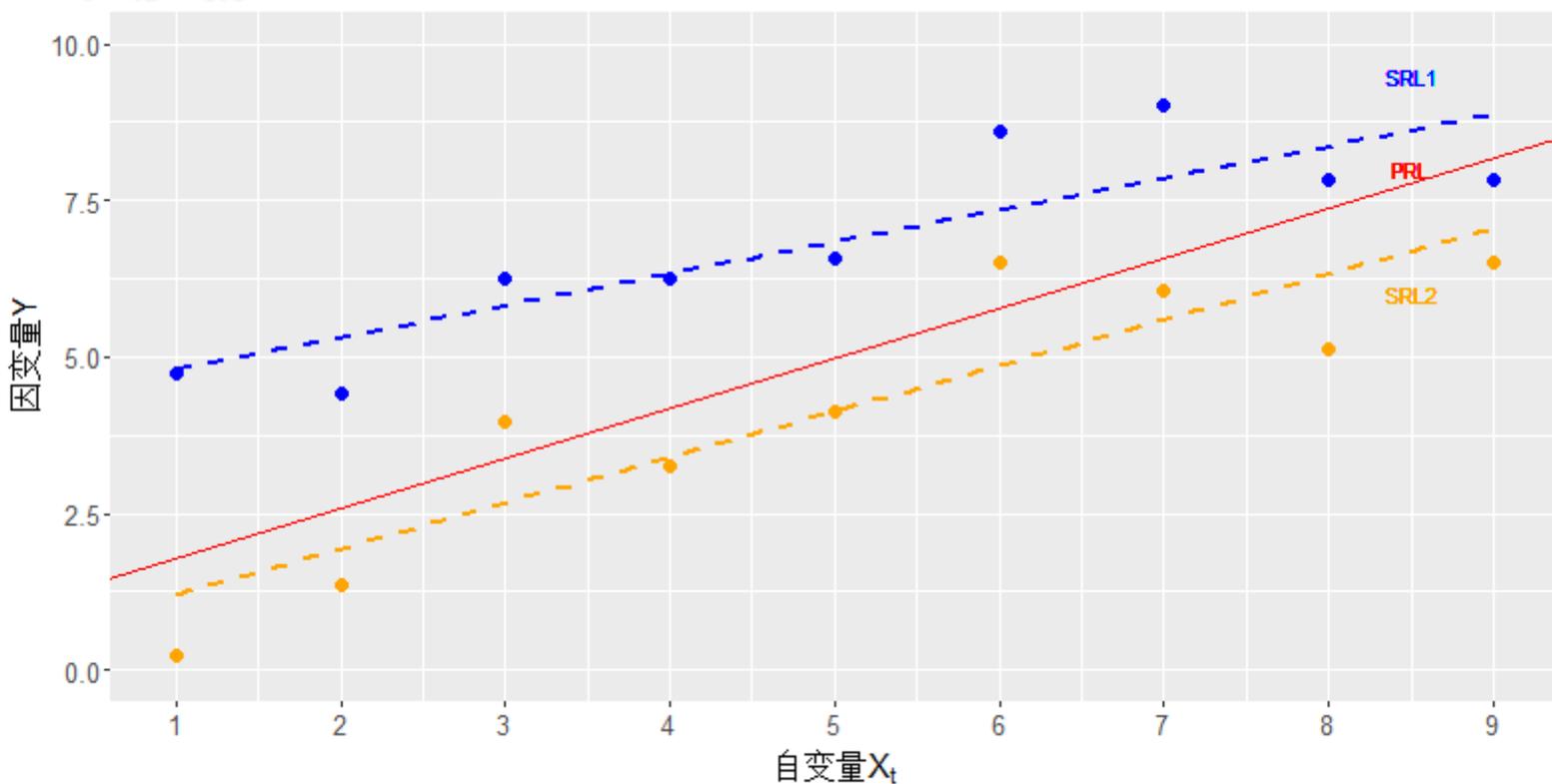
$$\begin{aligned} \widehat{NewY}_t &= + 0.48 && + 0.73X_t \\ (t) & (0.6331) && (5.4323) \\ (se) & (0.7569) && (0.1345) \\ (fitness) & R^2 = 0.8083; \bar{R}^2 = 0.7809 \\ & F^* = 29.51; p = 0.0010 \end{aligned}$$

注意：此处的 \widehat{NewY}_t 表示 \hat{Y}_t^*



蒙特卡洛模拟 (回归线2)

对OLS方法下的样本回归线 (SRL2) 绘图, 并与真实的总体回归线 (PRL) 进行对比:



$$E(Y|X_t) = 1 + 0.8X_t \quad (PRL/PRF)$$



工资与生产率案例

Y_t	X_t
60.8	48.9
62.5	50.6
64.6	52.9
66.1	55
67.7	56.8
69.1	58.8
71.7	61.2
73.5	62.5

Showing 1 to 8 of 46 entries

Previous 2 3 4 5 6 Next

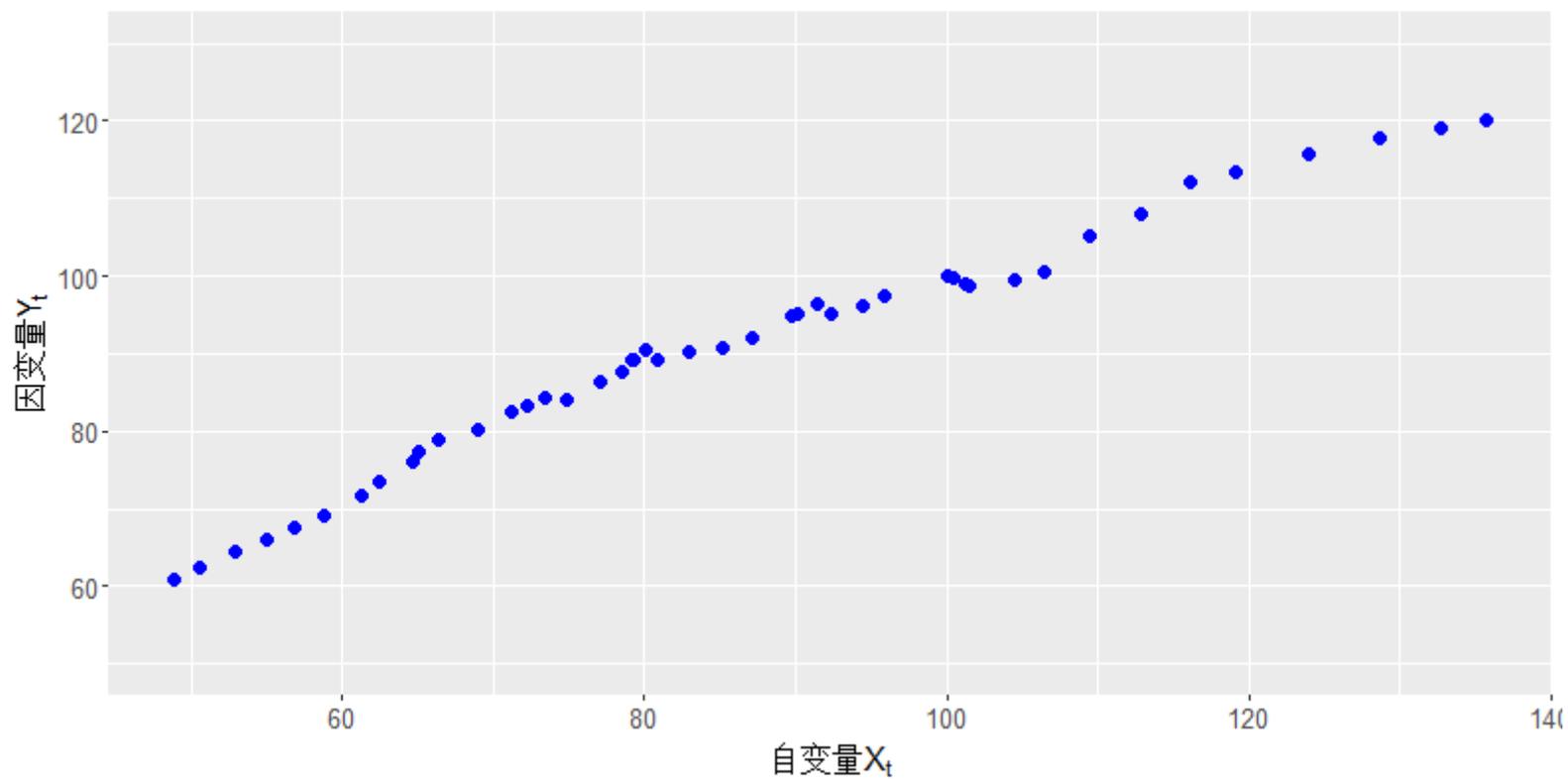
1960-2005年间美国商业部门工资与生产率数据($T=46$)

- Y_t 表示时均真实工资指数;
- X_t 表示生产效率指数





散点图





回归分析

对以上数据，我们可以分别构建如下的经典线性回归模型和双对数模型，并坚持采用OLS方法得到估计结果：

$$Y_t = + \beta_1 + \beta_2 X_t + u_t$$

$$\begin{aligned} \widehat{Y}_t &= + 4.30 && + 0.51X_t \\ (t) & (6.7249) && (4.4750) \\ (se) & (0.6389) && (0.1135) \\ (\text{fitness}) & R^2 = 0.7410; \bar{R}^2 = 0.7040 \\ & F^* = 20.03; p = 0.0029 \end{aligned}$$

$$\log(Y_t) = + \beta_1 + \beta_2 \log(X_t) + u_t$$

$$\begin{aligned} \widehat{\log(Y_t)} &= + 1.45 && + 0.31 \log(X_t) \\ (t) & (15.6171) && (5.2716) \\ (se) & (0.0930) && (0.0590) \\ (\text{fitness}) & R^2 = 0.7988; \bar{R}^2 = 0.7700 \\ & F^* = 27.79; p = 0.0012 \end{aligned}$$

思考：

- 我们的数据是否受到自相关问题的困扰？
- 我们如何发现自相关问题并加以纠正？

3.3.3 序列自相关性问题的诊断



图示法

图示法重点关注模型残差序列 (e_t) 是否存在某种系统化模式。总体回归模型PRM中随机干扰项 u_t 是不能直接观测到, 所以可通过观察样本回归模型SRM中残差的行为模式, 间接诊断随机干扰项是否存在自相关性问题。

- 图形1: 残差序列 e_t 时序图 (serial plot, 也即残差 e_t 随时期 t 的变化图)。
- 图形2: 残差序列 e_t 与残差1阶滞后序列 e_{t-1} 的散点图 (scatter plot)。
- 图形3: 残差序列 e_t 与残差 p 阶滞后序列 e_{t-p} , $p \in (2, 3, \dots, T-1)$ 的散点图 (scatter plot)。



图示法：案例演示（主回归）

首先构建如下双对数主回归模型：

$$\log(Y_t) = + \hat{\beta}_1 + \hat{\beta}_2 \log(X_t) + e_i$$

双对数主回归模型回归结果为：

$$\begin{aligned} \widehat{\log(Y_t)} &= + 1.61 && + 0.65 \log(X_t) \\ (t) & (29.3680) && (52.7996) \\ (se) & (0.0547) && (0.0124) \\ (\text{fitness}) & R^2 = 0.9845; \bar{R}^2 = 0.9841 \\ & F^* = 2787.80; p = 0.0000 \end{aligned}$$



图示法：案例演示（主回归EViews报告）

作为对照，下面给出的是主模型的EViews报告：

Equation: EQ_M0 Workfile: WAGE::corr\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(Y)
Method: Least Squares
Date: Time:
Sample: 1960 2005
Included observations: 46

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.606680	0.054709	29.36802	0.0000
LOG(X)	0.652216	0.012353	52.79963	0.0000

R-squared	0.984462	Mean dependent var	4.490151
Adjusted R-squared	0.984109	S.D. dependent var	0.175142
S.E. of regression	0.022078	Akaike info criterion	-4.745943
Sum squared resid	0.021448	Schwarz criterion	-4.666437
Log likelihood	111.1567	Hannan-Quinn criter.	-4.716159
F-statistic	2787.801	Durbin-Watson stat	0.217558
Prob(F-statistic)	0.000000		



图示法：案例演示（残差数据）

得到主回归模型的残差序列 e_t 、残差1阶滞后序列 e_{t-1} ，以及标准化变换的残差序列 $e_t^* = \frac{e_t}{S_{e_t}}$ ：

Y_t	X_t	\hat{Y}_t	e_t	e_{t-1}	e_t^*	$e_t^*/100$
60.8	48.9	4.1437	-0.0361		-1.6521	-0.0165
62.5	50.6	4.1659	-0.0308	-0.0361	-1.4099	-0.0141
64.6	52.9	4.1949	-0.0267	-0.0308	-1.2241	-0.0122
66.1	55	4.2203	-0.0292	-0.0267	-1.3357	-0.0134
67.7	56.8	4.2413	-0.0262	-0.0292	-1.2022	-0.0120
69.1	58.8	4.2639	-0.0283	-0.0262	-1.2985	-0.0130

Showing 1 to 6 of 46 entries

Previous

1

2

3

4

5

...

8

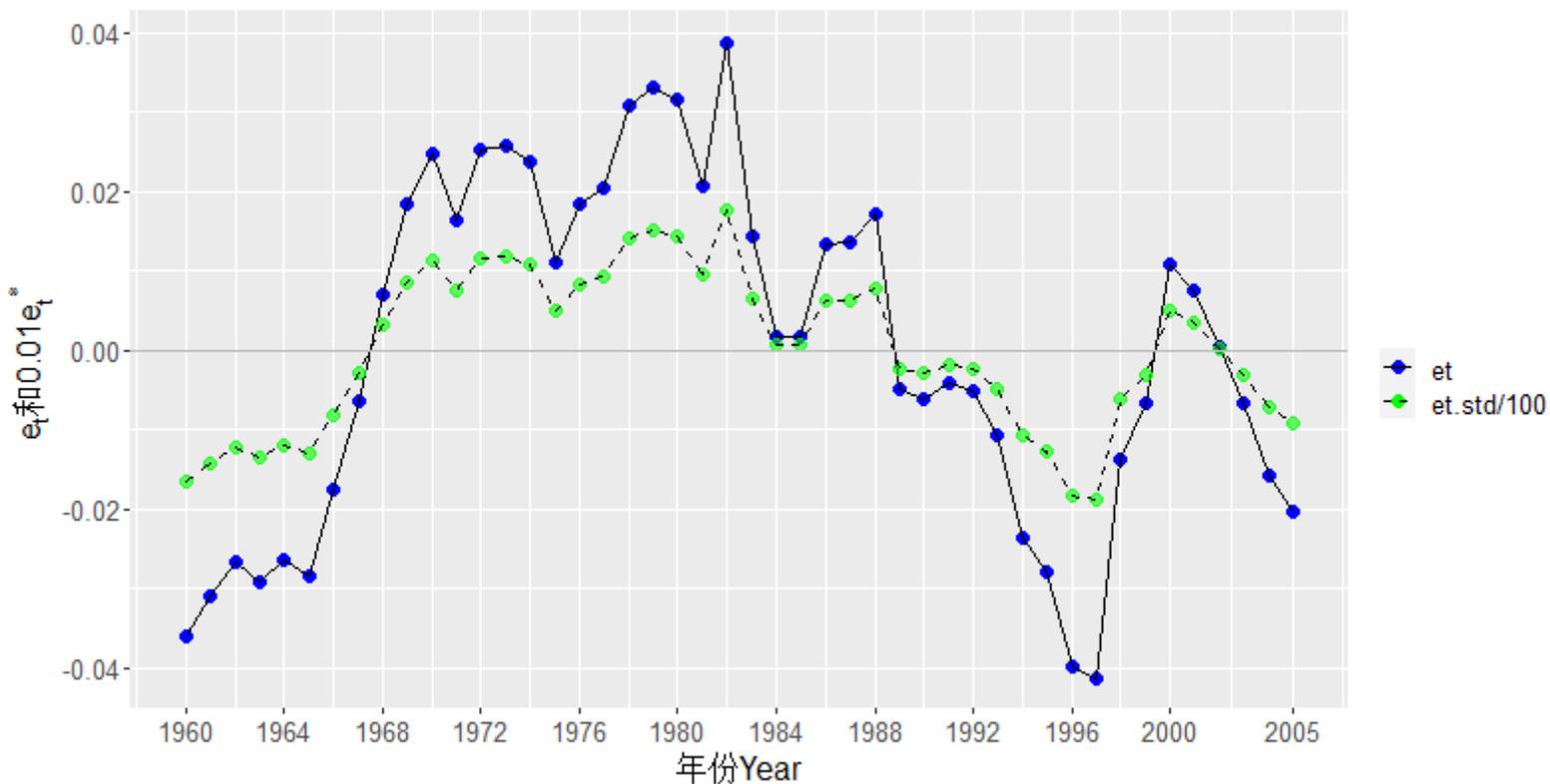
Next

提示： Y_t 表示时均真实工资指数； X_t 表示生产效率指数； $e_t^* = \frac{e_t}{S_{e_t}} = \frac{e_t}{\sum e_t^2 / (T-1)}$



图示法：案例演示（残差模式1）

残差序列（ e_t ）和1/100标准化残差（ $0.01e_t^*$ ）时序图（serial plot）如下：



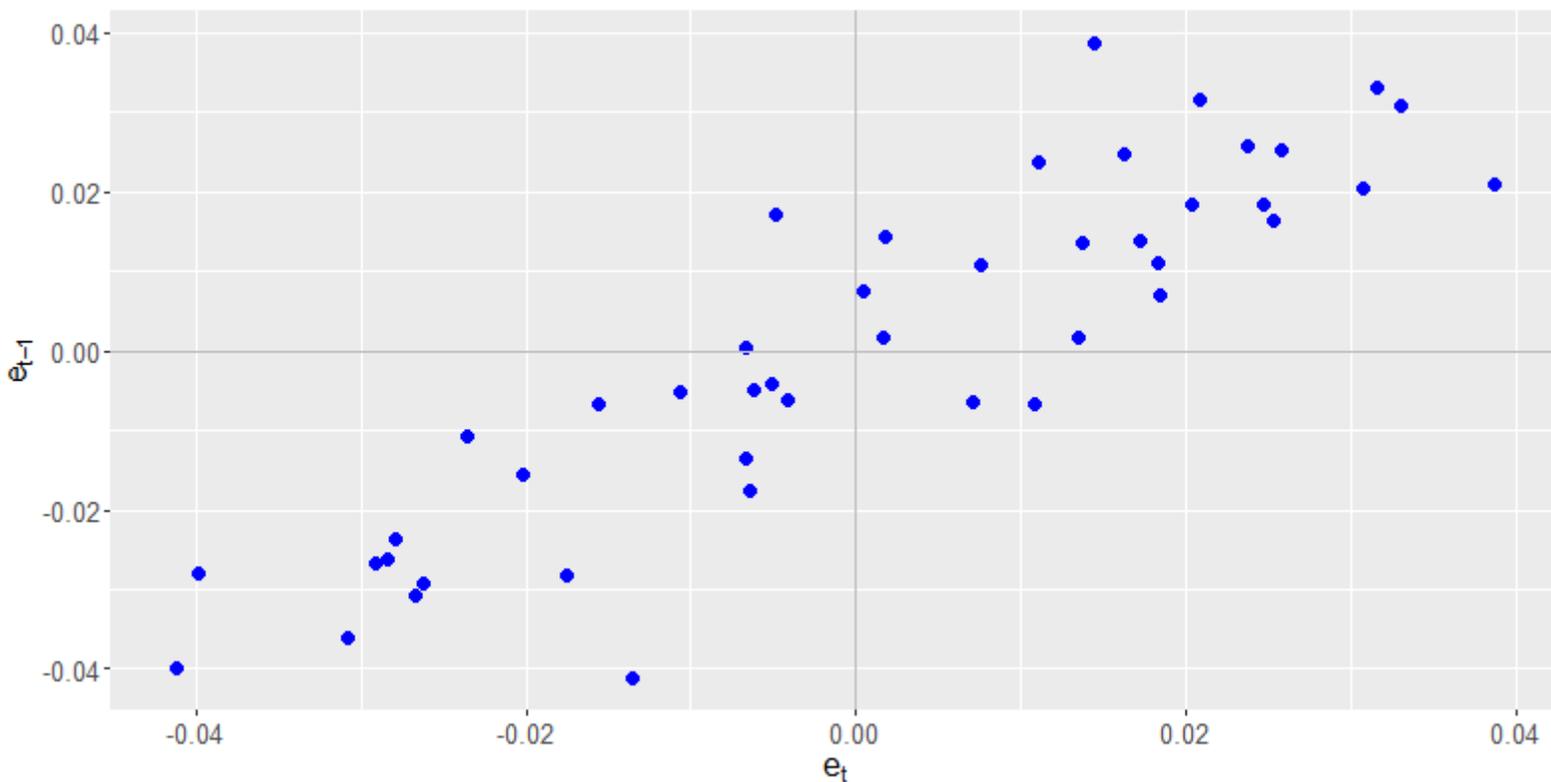
提示：et表示残差序列（ e_t ）；et.std/100表示1/100标准化残差序列（ $e_t^*/100$ ）。





图示法：案例演示（残差模式2）

残差序列 e_t 与残差1阶滞后序列 e_{t-1} 的散点图（scatter plot）为：



提示：因为残差滞后1期序列 e_{t-1} 的观测数只有45，所以此处只显示45个观测点。



辅助回归法

思路：利用样本数据，构建并分析残差 e_t 序列对 e_{t-1}, e_{t-2}, \dots 序列的**辅助回归方程**，从而间接判断随机干扰项 u_t 的自相关性模式。

考虑如下的对应关系：

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} \quad p \in (1, 2, \dots, T-1)$$

$$e_t = \hat{\rho}_1 e_{t-1} + \hat{\rho}_2 e_{t-2} + \dots + \hat{\rho}_p e_{t-p} + v_i \quad p \in (1, 2, \dots, T-1)$$



辅助回归法

步骤:

- 构建主回归模型，并进行OLS估计，得到残差序列（及其滞后 p 阶序列）；
- 根据残差图模式，构建相应的残差辅助回归方程（无截距模型），根据回归报告，得到主模型是否存在自相关性问题的初步结论。

诊断依据:

- 如果残差辅助回归方程的F检验显著，则表明主模型存在残差辅助回归方程所示的自相关性问题的。
- 如果残差辅助回归方程的F检验不显著，则表明主模型不存在残差辅助回归方程所示的自相关性问题的。



辅助回归法：OLS估计

利用主回归模型的残差序列 e_t ，我们可以构建如下的辅助回归模型：

$$e_t = \hat{\rho}_1 e_{t-1} + v_i$$

OLS估计的简要报告如下：

$$\begin{aligned} \hat{e}t &= + 0.87et.l1 \\ (t) & (12.7360) \\ (se) & (0.0681) \\ (fitness) R^2 &= 0.7866; \bar{R}^2 = 0.7818 \\ F^* &= 162.21; p = 0.0000 \end{aligned}$$

提示： `et` 表示残差序列 e_t ，`et.l1` 表示残差序列的1阶滞后变量 e_{t-1} 。



自相关和偏相关分析法

思路：分析残差 e_t 序列对 e_{t-1}, e_{t-2}, \dots 序列的自相关和偏相关统计图表报告（注意滞后阶数的选择）。

步骤：

- 对主模型进行OLS估计，得到残差 e_t 序列
- 利用统计软件（EViews或R等）绘制残差序列的自相关和偏相关图表
- 观察和比对残差序列的自相关和偏相关图表，得到主模型是否存在自相关性问题的初步结论。

诊断依据：

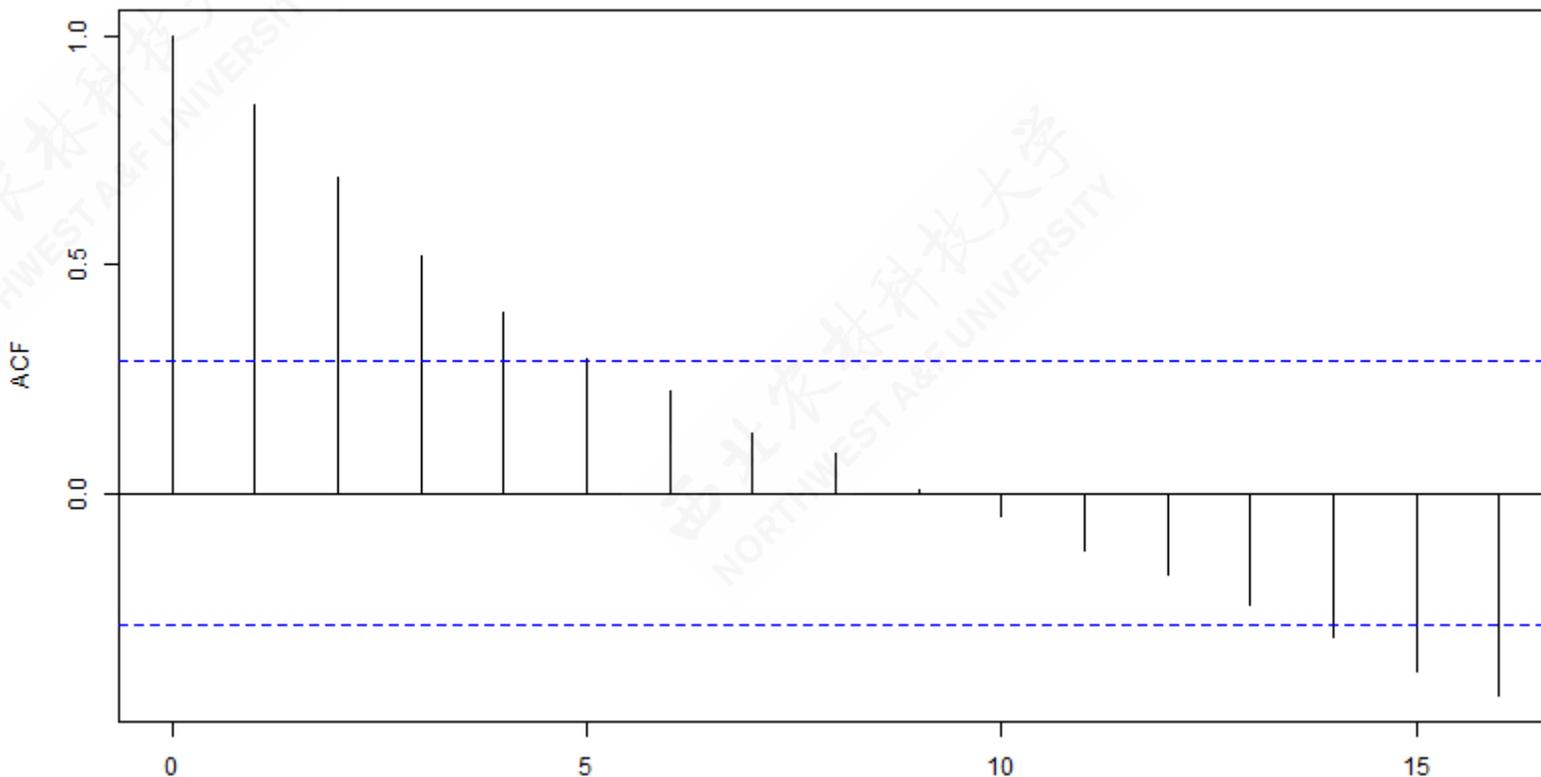
- 观察自相关图和偏相关图的组合关系，判断残差序列的自相关性模式。

$$e_t = \hat{\rho}_1 e_{t-1} + \hat{\rho}_2 e_{t-2} + \dots + \hat{\rho}_p e_{t-p} + v_i \quad p \in (1, 2, \dots, T-1)$$



案例：残差的自相关分析 (ACF) (R软件)

对于工资和劳动率案例，残差 e_t 序列的自相关图 (ACF) 为：

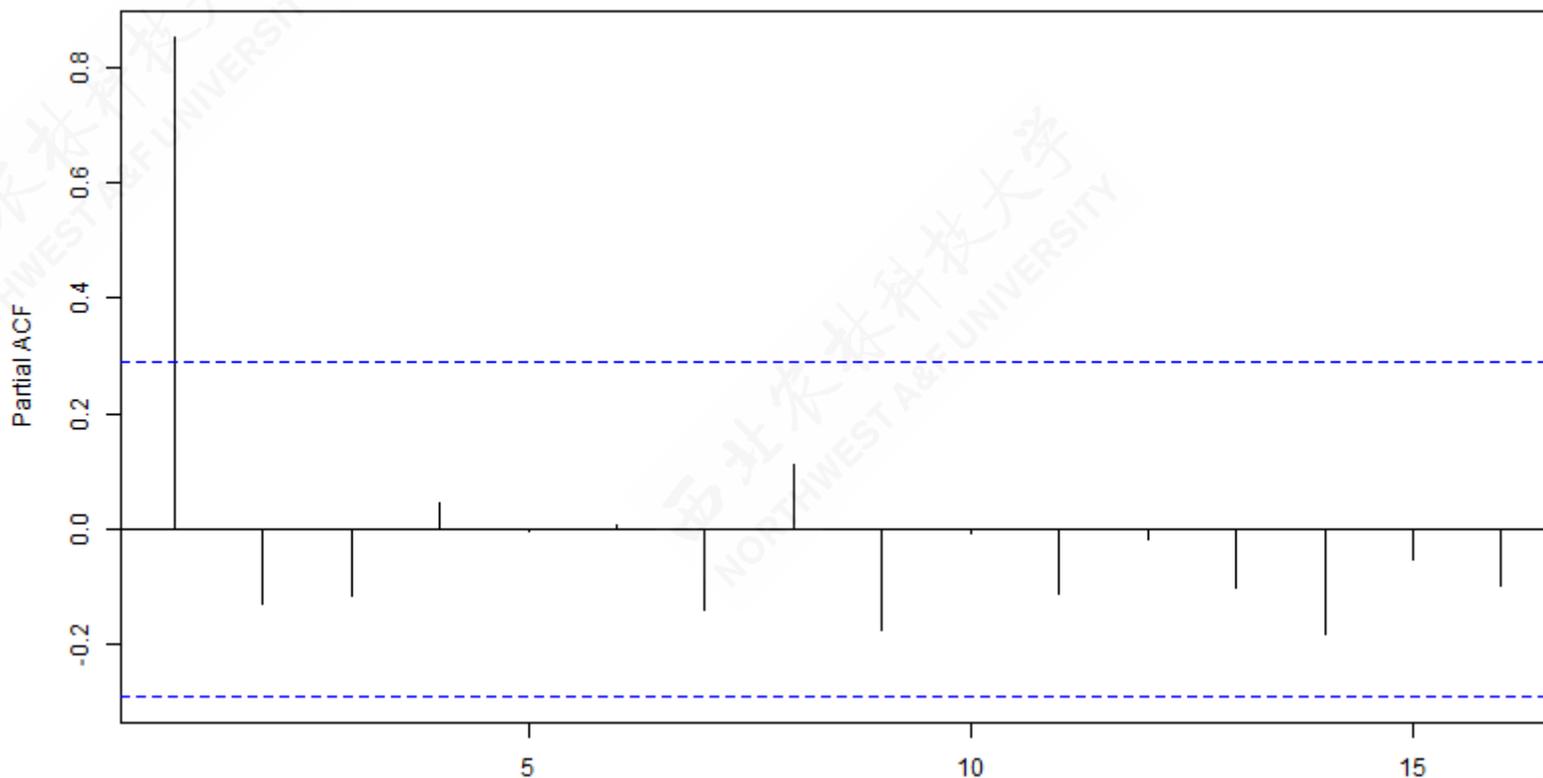


残差序列 e_t 的自相关图 (ACF)



案例：残差序列的偏自相关分析 (PACF) (R软件)

对于工资和劳动率案例，残差 e_t 序列的偏自相关图 (PACF) 为：



残差序列 e_t 的偏自相关图



案例：残差序列的自相关和偏自相关分析 (EViews软件)

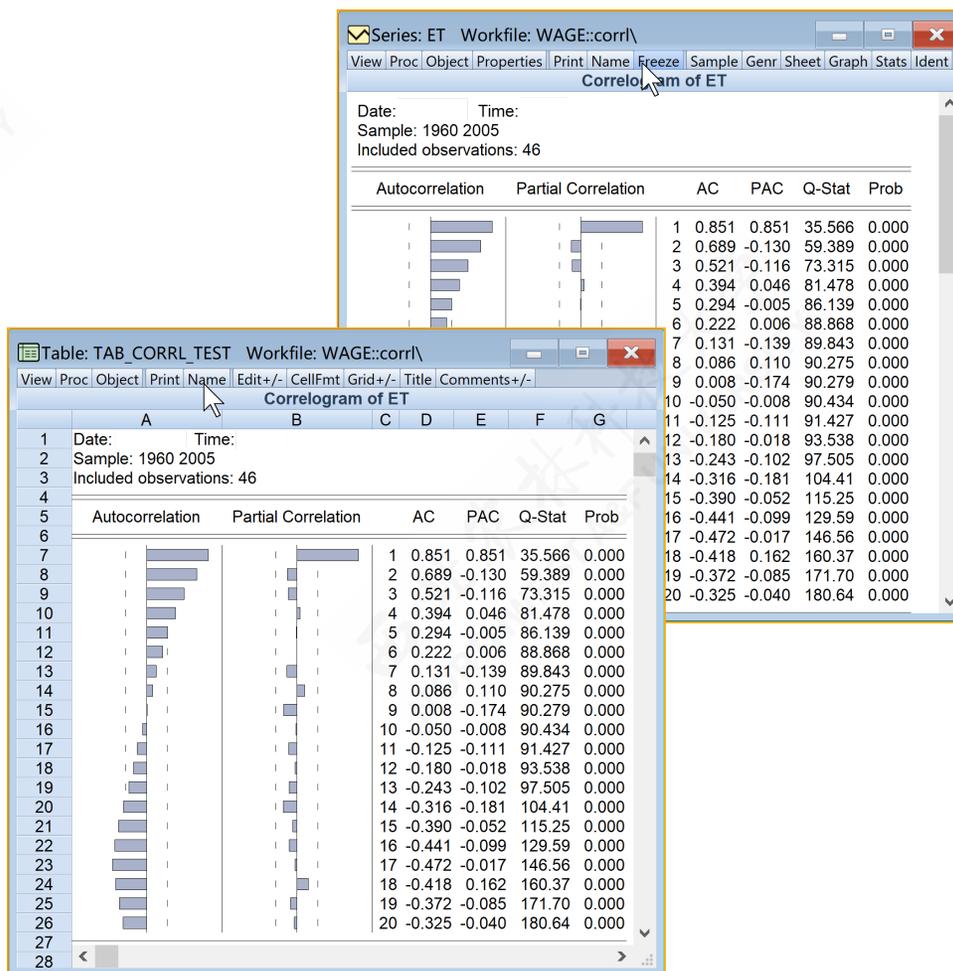
The screenshot shows the EViews software interface. The main window displays a list of objects, including 'et' (residuals). A context menu is open over the 'et' object, with 'Correlogram...' selected. A 'Correlogram Specification' dialog box is open, showing options for 'Level', '1st difference', and '2nd difference', with 'Level' selected. The 'Lags to include' is set to 20. The 'OK' button is highlighted.

Year	Value
1974	0.023744
1975	0.011131
1976	0.018359
1977	0.020416
1978	0.030781
1979	

残差自相关和偏相关分析的Eviews操作



案例：残差序列的自相关和偏自相关分析 (EViews软件)



残差自相关和偏相关的Eviews报告



德宾-沃森检验法 (Durbin-Watson test)

假设总体回归模型PRM存在如下1阶自相关,则可以认为存在: :

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t \quad \leftarrow [-1 < \rho < 1]$$

$$Y_t = \hat{\beta}_1 + \hat{\beta}_2 X_t + e_t$$
$$e_t = \hat{\rho} e_{t-1} + v_i$$

那么就可以构造得到如下的样本Durbin-Watson的d统计量:

$$d = \frac{\sum_2^T (e_t - e_{t-1})^2}{\sum_1^T e_t^2}$$

- 它其实是用相继残差的差异的平方和与“残差平方和RSS”之比。
- 由于取相继差异时损失1个观测值, 德宾-沃森d统计量的分子只有 $T - 1$ 次观测值。



德宾-沃森检验法 (Durbin-Watson test)

$$\begin{aligned}d &= \frac{\sum_2^T (e_t - e_{t-1})^2}{\sum_2^T e_t^2} \\&= \frac{\sum e_t^2 + \sum e_{t-1}^2 - 2 \sum e_t e_{t-1}}{\sum e_t^2} \quad \leftarrow \left[\sum e_t^2 \approx \sum e_{t-1}^2 \right] \\&\doteq 2 \left(1 - \frac{\sum e_t e_{t-1}}{\sum e_t^2} \right) \quad \leftarrow \left[\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_t^2} \right] \\&= 2(1 - \hat{\rho})\end{aligned}$$

因为 $0 \leq \hat{\rho} \leq 1$ ，所以德宾-沃森d统计量 $0 \leq d \leq 4$ ，并具有如下特征：

- 如果 $\hat{\rho} = +1$ ，则 $d = 0$ ，此时残差序列存在完全1阶正自相关性。
- 如果 $\hat{\rho} = 0$ ，则 $d = 2$ ，此时残差序列不存在自相关性。
- 如果 $\hat{\rho} = -1$ ，则 $d = 4$ ，此时残差序列存在完全1阶负自相关性。



德宾-沃森检验法 (Durbin-Watson test)

德宾-沃森检验法的步骤:

- 对主模型进行OLS估计, 得到残差 e_t 序列
- 得到主回归方程分析报告中的德宾-沃森d统计量 (Durbin-Watson)
- 查找德宾-沃森统计量 (Durbin-Watson) 理论表, 找到理论下界值 d_L 和理论上界值 d_U 。
- 比较德宾-沃森d统计量与查表值之间的关系, 根据诊断依据, 得到**主模型**是否存在自相关性问题的初步结论。

Durbin-Watson统计量服从 $\chi^2(n, k, \alpha)$ 分布, [具体可以参看Eviews在线帮助文档](#)

下界值 d_L 和上界值 d_U 的理论值使用bootstrap方法仿真计算得到, 与 (n, k, α) 有关

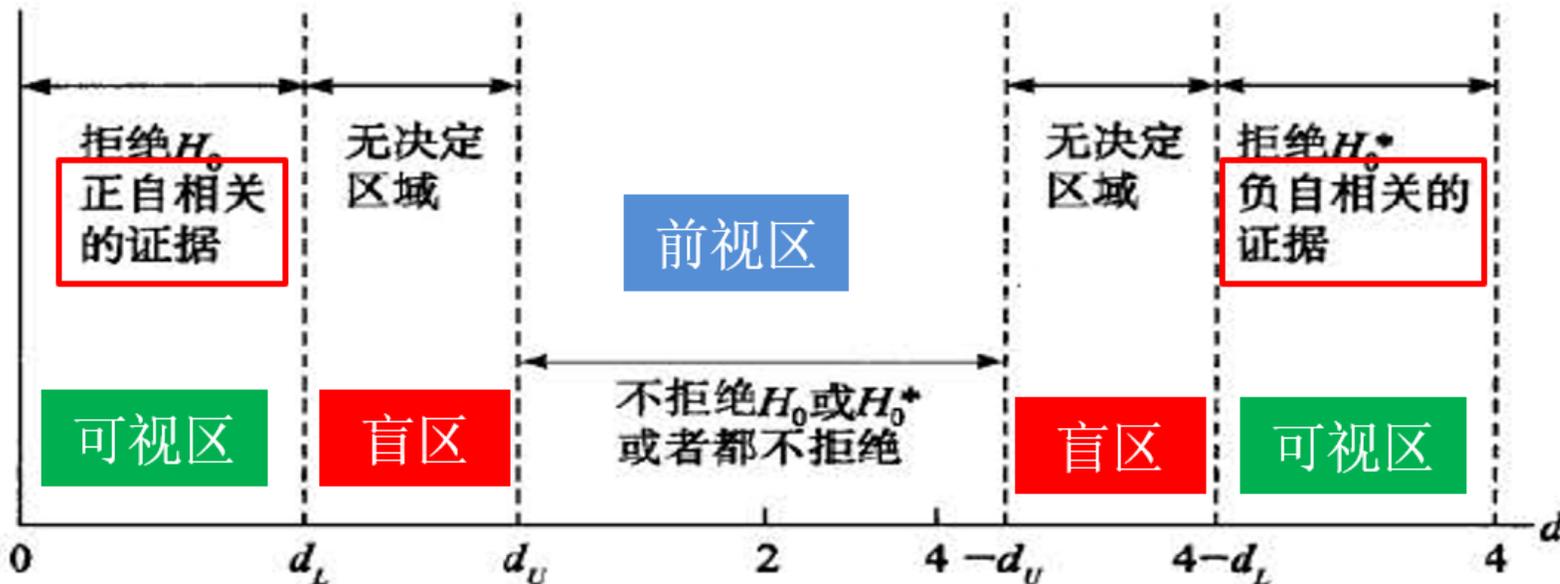
下界值 d_L 和上界值 d_U 的理论查表值可以参考[在线文档](#)



德宾-沃森检验法 (Durbin-Watson test)

德宾-沃森检验法的诊断依据（后视镜法则）：

- 如果 $0 < d < d_L$ ，则表明主模可能存在的1阶正自相关问题。
- 如果 $4 - d_L < d < 4$ ，则表明主模型可能存在的1阶负自相关问题。



虚拟假设：
 H_0 ：无正自相关
 H_0^* ：无负自相关

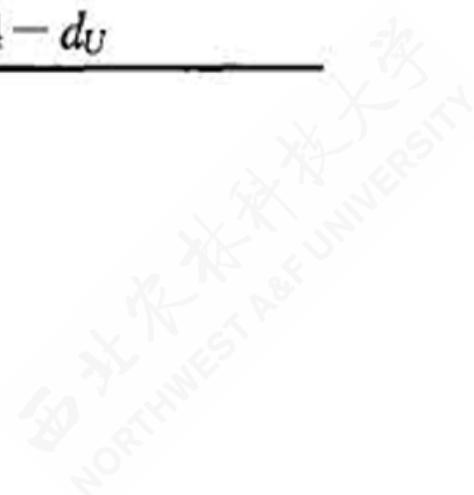


德宾-沃森检验法 (Durbin-Watson test)

德宾-沃森检验法的诊断依据（后视镜法则）：

德宾-沃森 d 检验：决策规则

虚拟假设	决策	如果
无正自相关	拒绝	$0 < d < d_L$
无正自相关	无决定	$d_L \leq d \leq d_U$
无负自相关	拒绝	$4 - d_L < d < 4$
无负自相关	无决定	$4 - d_U \leq d \leq 4 - d_L$
无正或负自相关	不拒绝	$d_U < d < 4 - d_U$





德宾-沃森检验法 (Durbin-Watson test)

德宾-沃森检验法的适用条件:

- 回归模型含有截距项, 如果没有截距项, 就必须重新做带有截距的回归
- 解释变量 $X_{2t}, X_{3t}, \dots, X_{kt}$ 是非随机的, 或者说, 在反复抽样中是被固定的.
- 随机干扰项 u_t 是按1阶自回归模式产生的:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t \quad \leftarrow [-1 < \rho < 1]$$

- 因变量 Y_t 的滞后变量 $Y_{t-1}, Y_{t-2}, \dots, Y_{t-q}$ 不能当作解释变量之一。如模型:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t$$

- 没有数据缺损, 如果数据缺失, d统计量无法补偿。





案例：德宾-沃森检验 (R软件)

工资-劳动率案例中，双对数主回归模型及其OLS回归结果为：

$$\begin{aligned} \widehat{\log(Y_t)} &= + 1.61 && + 0.65\log(X_t) \\ (t) & (29.3680) && (52.7996) \\ (se) & (0.0547) && (0.0124) \\ (\text{fitness}) & R^2 = 0.9845; \bar{R}^2 = 0.9841 \\ & F^* = 2787.80; p = 0.0000 \end{aligned}$$

我们可以利用残差序列，采用R软件计算得到德宾-沃森d统计量：

```
Durbin-Watson test

data: lm.main
DW = 0.21756, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```

- 根据主回归报告的计算结果，Durbin-Watson的d统计量为 $d = 0.2175583$ 。查表可知在当 $(n = 'rn', k = 3, \alpha = 0.05)$ 时， $d_L = 1.475, d_U = 1.566$ ，表明 $0 < d < d_L$ ，因此认为主模型可能存在的1阶正自相关性问题。



案例：德宾-沃森检验 (EViews报告)

The screenshot shows the EViews interface. The 'Equation Estimation' dialog box is open, showing the equation specification $\log(y) = c + \log(x)$ and the estimation method 'LS - Least Squares (NLS and ARMA)'. The 'Equation: EQ_M0' window displays the following regression results:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.606680	0.054709	29.36802	0.0000
LOG(X)	0.652216	0.012353	52.79963	0.0000

R-squared	0.984462	Mean dependent var	4.490151
Adjusted R-squared	0.984109	S.D. dependent var	0.175142
S.E. of regression	0.022078	Akaike info criterion	-4.745943
Sum squared resid	0.021448	Schwarz criterion	-4.666437
Log likelihood	111.1567	Hannan-Quinn criter.	-4.716159
F-statistic	2787.801	Durbin-Watson stat	0.217558
Prob(F-statistic)	0.000000		

主回归模型Eviews操作





案例：德宾-沃森检验 (EViews报告)

Equation: EQ_M0 Workfile: WAGE::corr\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(Y)
Method: Least Squares
Date: Time:
Sample: 1960 2005
Included observations: 46

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.606680	0.054709	29.36802	0.0000
LOG(X)	0.652216	0.012353	52.79963	0.0000

R-squared	0.984462	Mean dependent var	4.490151
Adjusted R-squared	0.984109	S.D. dependent var	0.175142
S.E. of regression	0.022078	Akaike info criterion	-4.745943
Sum squared resid	0.021448	Schwarz criterion	-4.666437
Log likelihood	111.1567	Hannan-Quinn criter.	-4.716159
F-statistic	2787.801	Durbin-Watson stat	0.217558
Prob(F-statistic)	0.000000		

主回归模型Eviews报告



拉格朗日乘数检验法 (LM test)

拉格朗日乘数检验 (LM test)，也称为布罗施-戈弗雷检验 (Breusch-Goldfrey, BG test)。

拉格朗日乘数检验假定随机干扰项 u_t 服从如下的 p 阶自回归模式 AR(p):

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t$$

拉格朗日乘数检验的原假设为： $H_0 : \rho_1 = \rho_2 = \cdots = \rho_p = 0$



拉格朗日乘数检验法 (LM test)

拉格朗日乘数检验的适用条件:

- 可以对高阶自相关模式[AR(p)]进行检验
- 允许非随机回归元
- 允许随机干扰项为自回归异动平均ARMA(p,q)模式。也即:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \cdots + \lambda_q \varepsilon_{t-q}$$

$$\varepsilon_t \sim iidN(0, 1)$$





拉格朗日乘数检验法 (LM test)

拉格朗日乘数检验的步骤:

- 对主模型进行OLS估计, 得到残差 e_t 序列
- 再利用主回归模型的残差序列, 做如下的**LM辅助回归**:

$$resid_t = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \cdots + \hat{\alpha}_k X_{kt} + \hat{\rho}_1 e_{t-1} + \hat{\rho}_2 e_{t-2} + \cdots + \hat{\rho}_p e_{t-p} + v_t$$

- 计算**LM辅助回归**方程的判定系数 R^2 , 并得到如下**LM统计量** (卡方统计量):

$$LM \equiv \chi^{2*} = (n - p)R^2 \sim \chi^2(p)$$





拉格朗日乘数检验法 (LM test)

拉格朗日乘数检验的诊断依据:

- 如果LM辅助回归方程的卡方检验显著, 也即 $LM \equiv \chi^{2*} > \chi_{1-\alpha}^2(p)$ (对应的概率值 $P < 0.1$), 则表明主模型是存在LM辅助回归方程形式的自相关性问题。
- 如果LM辅助回归方程的卡方检验不显著, 也即 $LM \equiv \chi^{2*} < \chi_{1-\alpha}^2(p)$ (对应的概率值 $P > 0.1$), 则表明主模型是不存在LM辅助回归方程形式的自相关性问题。



案例：拉格朗日乘数检验

工资-劳动率案例中，双对数主回归模型及其OLS回归结果为：

$$\begin{aligned} \widehat{\log(Y_t)} &= + 1.61 && + 0.65\log(X_t) \\ (t) & (29.3680) && (52.7996) \\ (se) & (0.0547) && (0.0124) \\ (\text{fitness}) & R^2 = 0.9845; \bar{R}^2 = 0.9841 \\ & F^* = 2787.80; p = 0.0000 \end{aligned}$$

我们可以利用残差序列构建 $p = 1$ 的LM辅助回归方程：

$$e_t = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2t} + \hat{\rho}_1 e_{t-1} + v_t$$



案例：拉格朗日乘数检验 (R软件)

采用R软件的拉格朗日乘数检验结果为：

```
Breusch-Godfrey test for serial correlation of order  
up to 1
```

```
data: lm.main  
LM test = 34.02, df = 1, p-value = 5.456e-09
```

$$resid_t = +0.0063 - 0.0015log(X_t) + 0.8687lag(resid)_1$$

- 根据LM辅助方程的计算结果，LM统计量为 $LM = \chi^{2*} = 34.0196$ ，其对应的概率值p为0.0000。
- 查表也可知在当 $n = 46, k = 2, \alpha = 0.05$ 时， $\chi_{1-\alpha}^2(p) = \chi_{0.95}^2(1) = 3.841459$ 要远远小于样本LM统计量值。
- 因此认为主模型可能存在自相关性问题。

提示： $resid_t$ 表示 e_t ； $lag(resid)_1$ 表示 e_{t-1} 。



案例：拉格朗日乘数检验 (EViews软件)

The screenshot shows the EViews software interface. The main window displays the command window with the following text:

```
Command  
View Proc Object View Proc Quick Options Add-ins Window Help  
Capture Command  
View Proc Object Save Snaps  
Range: 1960 2005 -- 46  
Sample: 1960 2005 -- 46  
c  
dot_resid  
eq_ar1_test  
eq_m0  
et  
et_100  
et_11  
et_sd  
resid  
scatter_et  
tab_corrl_test  
tab_lm_test  
x  
y
```

The 'Equation: EQ_M0' window is open, showing the 'Residual Diagnostics' menu. The 'Serial Correlation LM Test...' option is selected. The 'Lag Specification' dialog box is open, showing 'Lags to include' set to 1.

	Std. Error	t-Statistic	Prob.
Adjusted R-squared	0.984108		
S.E. of regression	0.022078		
Sum squared resid	0.021448		
Log likelihood	111.1567		
F-statistic	2787.801		
Prob(F-statistic)	0.000000		

The 'Lag Specification' dialog box shows:

Lags to include: 1

OK Cancel

拉格朗日自相关检验的Eviews操作



案例：拉格朗日乘数检验 (EViews软件)

Equation: EQ_M0 Workfile: WAGE::corr\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	122.1032	Prob. F(1,43)	0.0000
Obs*R-squared	34.01962	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: Time:
Sample: 1960 2005
Included observations: 46
Presample missing value lagged resid

Variable	Coefficient
C	0.006258
LOG(X)	-0.001502
RESID(-1)	0.868652

R-squared 0.739557
Adjusted R-squared 0.727443
S.E. of regression 0.011398
Sum squared resid 0.005586
Log likelihood 142.1002
F-statistic 61.05162
Prob(F-statistic) 0.000000

Table: TAB_LM_TEST Workfile: WAGE::corr\

View Proc Object Print Name Edit+/- CellFmt Grid+/- Title Comments+/-

	A	B	C	D	E
1	Breusch-Godfrey Serial Correlation LM Test:				
2					
3	F-statistic	122.1032	Prob. F(1,43)	0.0000	
4	Obs*R-squared	34.01962	Prob. Chi-Square(1)	0.0000	
5					
6					
7	Test Equation:				
8	Dependent Variable: RESID				
9	Method: Least Squares				
10	Date: 11.050(Time:).0000				
11	Sample: 1960 2005				
12	Included observations: 46				
13	Presample missing value lagged residuals set to zero.				
14					
15	Variable	Coefficient	Std. Error	t-Statistic	Prob.
16					
17	C	0.006258	0.028248	0.221551	0.8257
18	LOG(X)	-0.001502	0.006378	-0.235464	0.8150
19	RESID(-1)	0.868652	0.078611	11.05003	0.0000
20					
21	R-squared	0.739557	Mean dependent var	5.71E-16	
22	Adjusted R-squared	0.727443	S.D. dependent var	0.021832	
23	S.E. of regression	0.011398	Akaike info criterion	-6.047835	
24	Sum squared resid	0.005586	Schwarz criterion	-5.928576	
25	Log likelihood	142.1002	Hannan-Quinn criter.	-6.003160	
26	F-statistic	61.05162	Durbin-Watson stat	1.518451	
27	Prob(F-statistic)	0.000000			
28					
29					

拉格朗日自相关检验的Eviews报告

3.3.4 序列自相关性问题的矫正



广义差分方程法（自相关系数已知）

广义差分方程法：对主回归方程进行合适的广义差分变换，使得变换以后的新模型不再有自相关问题，然后再使用OLS方法估计，从而得到参数估计的**BLUE**。

广义差分法变换法属于广义最小二乘法（GLS）的一种，专门用来处理随机干扰项出现自相关性问题的一种常用方法。



广义差分方程法 (自相关系数已知)

广义差分方程法的原理：如果主模型随机干扰项的自相关系数 ρ 已知，则可以直接用差分变换得到新模型，容易证明新模型将不再有自相关问题。

下面将说明1阶自相关情形AR(1)下的广义差分变换的理论过程：

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \quad (\text{PRM})$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (\text{AR}(1))$$

$$\rho Y_{t-1} = \rho \beta_1 + \beta_2 \rho X_{2t-1} + \rho u_{t-1} \quad (\text{Lag 1 Model})$$

$$(Y_t - \rho Y_{t-1}) = \beta_1(1 - \rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + (u_t - \rho u_{t-1}) \quad (\Delta \text{ Model})$$

$$Y_t^* = \beta_1^* + \beta_2^* X_{2t}^* + \varepsilon_t \quad (\text{Adjusted Model})$$

其中，AR(1)模型中的 $\varepsilon_t \sim i.i.d N(0, \sigma_\varepsilon^2)$ 。



广义差分方程法：基于残差辅助方程

矫正思路：如果主模型随机干扰项的自相关系数未知，则可以直接用基于残差辅助方程估计得到 $\hat{\rho}$ ；再根据 $\rho \simeq \hat{\rho}$ 用广义差分变换得到新模型，容易证明新模型将不再有自相关问题。

如下将展示1阶自相关AR(1)情形下的广义差分变换的理论过程：

$Y_t = \beta_1 + \beta_2 X_{2t} + u_t$	PRM
$u_t = \rho u_{t-1} + \varepsilon_t$	AR(1)
$Y_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + e_t$	SRM
$e_t = \hat{\rho} e_{t-1} + v_t$	Auxiliary Model
$\rho \simeq \hat{\rho}$	
$\rho Y_{t-1} = \rho \beta_1 + \beta_2 \rho X_{2t-1} + \rho u_{t-1}$	Lag 1 Model
$(Y_t - \rho Y_{t-1}) = \beta_1(1 - \rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + (u_t - \rho u_{t-1})$	$\Delta 1$ Model
$Y_t^* = \beta_1^* + \beta_2^* X_{2t}^* + \varepsilon_t$	Adjusted Model

其中，AR(1)模型中的 $\varepsilon_t \sim i.i.d N(0, \sigma_\varepsilon^2)$ 。



案例矫正：基于残差辅助方程 (EViews软件)

The screenshot shows the EViews software interface. The Command window contains the following command:

```
scalar rho_ar1=@round(10000*eq_ar1_test.@coefs(1))/10000
```

The Object window shows a list of objects, with 'rho_ar1' selected. A pop-up window titled 'Scalar: RHO_AR1' displays the value of the scalar:

	Value
RHO_AR1	0.867900

基于残差辅助方程近似计算自相关系数



案例矫正：基于残差辅助方程 (EViews软件)

Equation Estimation

Equation specification
Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like $Y=c(1)+c(2)*X$.

$(\log(y)-\rho_{ar1}\log(y(-1))) c (\log(x)-\rho_{ar1}\log(x(-1)))$

Equation: EQ_ADJ_AR1 Workfile: WAGE::corr\

Dependent Variable: $\text{LOG}(Y)-0.8679*\text{LOG}(Y(-1))$
Method: Least Squares
Date: Time:
Sample (adjusted): 1961 2005
Included observations: 45 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.257951	0.025225	10.22606	0.0000
$\text{LOG}(X)-0.8679*\text{LOG}(X(-1))$	0.577378	0.041608	13.87650	0.0000

R-squared 0.817454 Mean dependent var 0.607417
Adjusted R-squared 0.813209 S.D. dependent var 0.022291
S.E. of regression 0.009634 Akaike info criterion -6.403601
Sum squared resid 0.003991 Schwarz criterion -6.323305
Log likelihood 146.0810 Hannan-Quinn criter. -6.373667
F-statistic 192.5573 Durbin-Watson stat 1.702603
Prob(F-statistic) 0.000000

基于残差辅助方程的广义差分矫正操作



案例矫正：基于残差辅助方程 (EViews软件)

Equation: EQ_ADJ_AR1 Workfile: WAGE::corr\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: $\text{LOG}(Y)-0.8679*\text{LOG}(Y(-1))$
Method: Least Squares
Date: Time:
Sample (adjusted): 1961 2005
Included observations: 45 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.257951	0.025225	10.22606	0.0000
$\text{LOG}(X)-0.8679*\text{LOG}(X(-1))$	0.577378	0.041608	13.87650	0.0000

R-squared	0.817454	Mean dependent var	0.607417
Adjusted R-squared	0.813209	S.D. dependent var	0.022291
S.E. of regression	0.009634	Akaike info criterion	-6.403601
Sum squared resid	0.003991	Schwarz criterion	-6.323305
Log likelihood	146.0810	Hannan-Quinn criter.	-6.373667
F-statistic	192.5573	Durbin-Watson stat	1.702603
Prob(F-statistic)	0.000000		

基于残差辅助方程的广义差分矫正报告



广义差分方程法：基于D-W统计量

矫正思路：如果主模型随机干扰项的自相关系数未知，则可以基于Durbin-Waston检验的d统计量计算得到 $\hat{\rho}$ ，再根据 $\rho \simeq \hat{\rho}$ 用广义差分变换得到新模型，容易证明新模型将不再有自相关问题。

如下将展示一阶自相情形下的广义差分变换的理论过程

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$d \simeq 2(1 - \hat{\rho})$$

$$\hat{\rho} \simeq 1 - d/2$$

$$\rho \simeq \hat{\rho}$$

$$\rho Y_{t-1} = \rho \beta_1 + \beta_2 \rho X_{2t-1} + \rho u_{t-1}$$

$$(Y_t - \rho Y_{t-1}) = \beta_1(1 - \rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + (u_t - \rho u_{t-1})$$

$$Y_t^* = \beta_1^* + \beta_2^* X_{2t}^* + \varepsilon_t$$

PRM

AR(1)

Durbin-Waston

Lag 1 Model

$\Delta 1$ Model

Adjusted Model

其中，AR(1)模型中的 $\varepsilon_t \sim i.i.d N(0, \sigma_\varepsilon^2)$ 。



案例矫正：基于D-W统计量

The screenshot shows the EViews interface with the following command window content:

```
Command  
scalar dw=eq_m0.@dw  
scalar rho_dw=@round(10000*(1-dw/2))/10000
```

Below the command window, two scalar result windows are displayed:

Scalar: DW Workfile: WAGE::corr\

	Value
DW	0.217558

Scalar: RHO_DW Workfile: WAGE::corr\

	Value
RHO_DW	0.891200

基于Durbin-Waston统计量近似计算自相关系数



案例矫正：基于D-W统计量

The screenshot shows the EViews interface with the 'Quick' menu open and 'Estimate Equation...' selected. The 'Equation Estimation' dialog box is open, showing the equation specification: $(\log(y) - \rho_{dw} \log(y-1)) - c(\log(x) - \rho_{dw} \log(x-1))$. The estimation settings are set to 'Method: LS - Least Squares (NLS and ARMA)' and 'Sample: 1960 2005'.

基于Durbin-Watson统计量的广义差分模型矫正操作



案例矫正：基于D-W统计量

Equation: EQ_ADJ_DW Workfile: WAGE::corr\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(Y)-0.8912*LOG(Y(-1))
Method: Least Squares
Date: Time:
Sample (adjusted): 1961 2005
Included observations: 45 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.219629	0.024824	8.847427	0.0000
LOG(X)-0.8912*LOG(X(-...	0.563819	0.049317	11.43245	0.0000
R-squared	0.752448	Mean dependent var	0.502951	
Adjusted R-squared	0.746691	S.D. dependent var	0.019173	
S.E. of regression	0.009650	Akaike info criterion	-6.400369	
Sum squared resid	0.004004	Schwarz criterion	-6.320073	
Log likelihood	146.0083	Hannan-Quinn criter.	-6.370435	
F-statistic	130.7009	Durbin-Watson stat	1.723983	
Prob(F-statistic)	0.000000			

基于Durbin-Watson统计量的广义差分模型矫正报告





可行广义最小二乘法(FGLS)：基迭代法

矫正思路：

- 如果主模型随机干扰项的自相关系数未知，而且存在高阶自相关情形，则可以使用基于迭代的可行广义最小二乘法(FGLS)计算得到 $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_p$ $p \in (1, 2, \dots)$ ，再根据 $\rho \simeq \hat{\rho}$ 用广义差分变换得到新模型，容易证明新模型将不再有自相关问题。
- 这些迭代方法主要包括：
 - 科克伦-奥克特迭代法(Cochrane-Orcutt iterative procedure)；
 - 科克伦-奥克特两步法(Cochrane-Orcutt two-step procedure)；
 - 德宾两步法(Durbin two-step procedure)；
 - 希尔德雷思-卢扫描或搜寻程序(Hildreth-Lu scanning or search procedure) 等



可行广义最小二乘法(FGLS)：基于迭代法

如下将展示科克伦-奥克特迭代法下对二阶自相关 ($AR(p), p = 2$) 情形下的广义差分变换的理论过程

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t$$

PRM

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

AR(2)

...

Cochrane-Orcutt iterative

$$\rho_p \simeq \hat{\rho}_p$$

$$\rho_1 Y_{t-1} = \rho_1 \beta_1 + \beta_2 \rho_1 X_{2t-1} + \rho_1 u_{t-1}$$

Lag 1 Model

$$\rho_2 Y_{t-2} = \rho_2 \beta_1 + \beta_2 \rho_2 X_{2t-2} + \rho_2 u_{t-2}$$

Lag 2 Model

$$(Y_t - \rho_1 Y_{t-1} - \rho_2 Y_{t-2}) = \beta_1(1 - \rho_1 - \rho_2) + \beta_2(X_{2t} - \rho_1 X_{2t-1} - \rho_2 X_{2t-2}) + (u_t - \rho_1 u_{t-1} - \rho_2 u_{t-2})$$

$\Delta 2$ Model

$$Y_t^* = \beta_1^* + \beta_2^* X_{2t}^* + \varepsilon_t$$

Adjusted Model



矫正案例：基于迭代法 (2GLS)

Equation Estimation

Specification Options

Equation specification
Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like $Y=c(1)+c(2)*X$.

log(y) c log(x) AR(1)

Equation: EQ_ADJ_CO Workfile: WAGE::corr1\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(Y)
Method: ARMA Maximum Likelihood (BFGS)
Date: Time:
Sample: 1960 2005
Included observations: 46
Convergence achieved after 6 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.563926	0.129071	12.11683	0.0000
LOG(X)	0.660116	0.029140	22.65321	0.0000
AR(1)	0.899443	0.063526	14.15868	0.0000
SIGMASQ	9.83E-05	2.52E-05	3.905427	0.0003

R-squared 0.996724 Mean dependent var 4.490151
Adjusted R-squared 0.996490 S.D. dependent var 0.175142
S.E. of regression 0.010376 Akaike info criterion -6.179775
Sum squared resid 0.004521 Schwarz criterion -6.020762
Log likelihood 146.1348 Hannan-Quinn criter. -6.120208
F-statistic 4260.033 Durbin-Watson stat 1.675329
Prob(F-statistic) 0.000000

Inverted AR Roots .90

基于科克伦-奥克特迭代法的2GLS模型矫正Eviews操作



矫正案例：基于迭代法 (2GLS)

Equation: EQ_ADJ_CO Workfile: WAGE::corr\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(Y)
Method: ARMA Maximum Likelihood (BFGS)
Date: Time:
Sample: 1960 2005
Included observations: 46
Convergence achieved after 6 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.563926	0.129071	12.11683	0.0000
LOG(X)	0.660116	0.029140	22.65321	0.0000
AR(1)	0.899443	0.063526	14.15868	0.0000
SIGMASQ	9.83E-05	2.52E-05	3.905427	0.0003

R-squared	0.996724	Mean dependent var	4.490151
Adjusted R-squared	0.996490	S.D. dependent var	0.175142
S.E. of regression	0.010376	Akaike info criterion	-6.179775
Sum squared resid	0.004521	Schwarz criterion	-6.020762
Log likelihood	146.1348	Hannan-Quinn criter.	-6.120208
F-statistic	4260.033	Durbin-Watson stat	1.675329
Prob(F-statistic)	0.000000		

Inverted AR Roots	.90
-------------------	-----

基于科克伦-奥克特迭代法的2GLS模型矫正Eviews报告



矫正案例：基于迭代法 (JGLS)

The screenshot shows the EViews interface. The Command window contains the command: `scalar rho_co=eq_adj_co.@coefs(3)`. Below the command window, a list of objects is displayed, with `rho_co` selected. A pop-up window titled "Scalar: RHO_CO" is open, showing the value 0.8994429974194881. Below the value, a table displays the scalar value:

	Value
RHO_CO	0.899443

基于科克伦-奥克特迭代法近似计算的自相关系数



一致性标准误校正法 (HAC) : 尼威-威斯特(Newey-West)

目标: 直接用尼威-威斯特(Newey-West)一致性标准误校正流程方法, 构建回归分析模型, 此时模型的自相关问题将会有所缓解。

矫正思路: 利用统计软件 (Eviews或R等), 进行基于尼威-威斯特(Newey-West)一致性标准误校正程序的建模分析。

理论提示: (数学表达和证明过程略)

- 异方差-自相关一致性标准误 (heteroscedasticity-autocorrelation consistent standard errors, HAC) 也被简称为尼威-威斯特一致性标准误(Newey-West consistent standard errors)
- 尼威-威斯特(Newey-West)一致性标准误校正程序或菜单, 在主流的统计软件里都会配置
- 尼威-威斯特(Newey-West)一致性标准误校正程序, 严格意义上对于大样本数据是有效的, 因此不太适合于小样本数据的情形。



矫正案例：一致性标准误校正法 (HAC)

Equation Estimation

Specification Options

Equation specification

Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like $Y=c(1)+c(2)*X$.

$\log(y) = \log(x)$

Equation Estimation

Specification Options

Coefficient covariance

Covariance method: HAC (Newey-West)

Info matrix: HAC (Newey-West)

d.f. Adjusted HAC (various)

Cluster-robust

Weights

Type: None

Weight series:

Scaling: EViews default

Equation: EQ_ADJ_NW Workfile: WAGE::corr1

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(Y)

Method: Least Squares

Date: Time:

Sample: 1960 2005

Included observations: 46

HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 4.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.606680	0.101930	15.76266	0.0000
LOG(X)	0.652216	0.022564	28.90473	0.0000

R-squared 0.984462 Mean dependent var 4.490151

Adjusted R-squared 0.984109 S.D. dependent var 0.175142

S.E. of regression 0.022078 Akaike info criterion -4.745943

Sum squared resid 0.021448 Schwarz criterion -4.666437

Log likelihood 111.1567 Hannan-Quinn criter. -4.716159

F-statistic 2787.801 Durbin-Watson stat 0.217558

Prob(F-statistic) 0.000000 Wald F-statistic 835.4835

Prob(Wald F-statistic) 0.000000

厄威-威斯特(Newey-West)矫正法的操作



矫正案例：一致性标准误校正法 (HAC)

Equation: EQ_ADJ_NW Workfile: WAGE::corr\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(Y)
Method: Least Squares
Date: Time:
Sample: 1960 2005
Included observations: 46
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 4.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.606680	0.101930	15.76266	0.0000
LOG(X)	0.652216	0.022564	28.90473	0.0000

R-squared	0.984462	Mean dependent var	4.490151
Adjusted R-squared	0.984109	S.D. dependent var	0.175142
S.E. of regression	0.022078	Akaike info criterion	-4.745943
Sum squared resid	0.021448	Schwarz criterion	-4.666437
Log likelihood	111.1567	Hannan-Quinn criter.	-4.716159
F-statistic	2787.801	Durbin-Watson stat	0.217558
Prob(F-statistic)	0.000000	Wald F-statistic	835.4835
Prob(Wald F-statistic)	0.000000		

尼威-威斯特(Newey-West)矫正法的Eviews报告

本章結束

