

## § 2.7 序列相关性

### Serial Correlation

- 一、序列相关性
- 二、序列相关性的后果
- 三、序列相关性的检验
- 四、具有序列相关性模型的估计
- 五、案例

普通最小二乘法（OLS）要求计量模型的随机误差项相互独立或序列不相关。

如果模型的随机误差项违背了互相独立的基本假设的情况，称为序列相关性。

# 一、序列相关性

# 1、序列相关的概念

对于模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i \quad i=1,2,\cdots,n$$

随机误差项互不相关的基本假设表现为：

$$Cov(\mu_i, \mu_j) = 0 \quad i \neq j, \quad i, j=1,2,\cdots,n$$

如果对于不同的样本点，随机误差项之间不再是不相关的，而是存在某种相关性，则认为出现了**序列相关性**。

在其他假设仍成立的条件下，**序列相关**即意味着

$$E(\mu_i \mu_j) \neq 0$$

或

$$\begin{aligned} E(NN^T) &= E\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} (\mu_1 \quad \cdots \quad \mu_n)\right) = E\begin{pmatrix} \mu_1^2 & \cdots & \mu_1 \mu_n \\ \vdots & \ddots & \vdots \\ \mu_n \mu_1 & \cdots & \mu_n^2 \end{pmatrix} \\ &= \begin{pmatrix} E(\mu_1^2) & \cdots & E(\mu_1 \mu_n) \\ \vdots & \ddots & \vdots \\ E(\mu_n \mu_1) & \cdots & E(\mu_n^2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \cdots & E(\mu_1 \mu_n) \\ \vdots & \ddots & \vdots \\ E(\mu_n \mu_1) & \cdots & \sigma_n^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma^2 & \cdots & E(\mu_1 \mu_n) \\ \vdots & \ddots & \vdots \\ E(\mu_n \mu_1) & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \Omega \neq \sigma^2 \mathbf{I} \end{aligned} \quad (2.5.1)$$

如果仅存在

$$E(\mu_i \mu_{i+1}) \neq 0 \quad i=1,2,\cdots,n-1 \quad (2.5.2)$$

称为**一阶序列相关**，或**自相关**（autocorrelation）。这是最常见的一种序列相关问题。

**自相关**往往可写成如下形式：

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t \quad -1 < \rho < 1 \quad (2.5.3)$$

其中： $\rho$ 被称为**自协方差系数**（coefficient of autocovariance）或**一阶自相关系数**（first-order coefficient of autocorrelation）。

**非自相关假定：**  $\text{Cov}(u_i, u_j) = \text{E}(u_i u_j) = 0, (i, j \in T, i \neq j)$

如果  $\text{Cov}(u_i, u_j) \neq 0, (i, j \in T, i \neq j)$  则称误差项  $u_t$  存在自相关。

自相关又称序列相关。也是相关关系的一种。

自相关按形式可**分为两类**：

(1) 一阶自回归形式。  $u_t = f(u_{t-1})$

(2) 高阶自回归形式。  $u_t = f(u_{t-1}, u_{t-2}, \dots)$

经济计量模型中自相关的最常见形式是**一阶线性自回归形式**。

$$u_t = \alpha_1 u_{t-1} + v_t$$

$$\text{E}(v_t) = 0, t = 1, 2, \dots, T$$

$$\text{Var}(v_t) = \sigma_v^2, t = 1, 2, \dots, T$$

$$\text{Cov}(v_i, v_j) = 0, i \neq j, i, j = 1, 2, \dots, T$$

$$\text{Cov}(u_{t-1}, v_t) = 0, t = 1, 2, \dots, T$$

依据普通最小二乘法公式，模型  $u_t = \alpha_1 u_{t-1} + v_t$  中  $\alpha_1$  的估计公式是，

$$\hat{\alpha}_1 = \frac{\sum_{t=2}^T u_t u_{t-1}}{\sum_{t=2}^T u_{t-1}^2} \text{。若把 } u_t, u_{t-1} \text{ 看作两个变量，则它们的相关系数是}$$

$$\hat{\rho} = \frac{\sum_{t=2}^T u_t u_{t-1}}{\sqrt{\sum_{t=2}^T u_t^2} \sqrt{\sum_{t=2}^T u_{t-1}^2}} \text{。对于充分大的样本显然有 } \sum_{t=2}^T u_t^2 \approx \sum_{t=2}^T u_{t-1}^2 \text{。}$$

把这种关系代入上式得

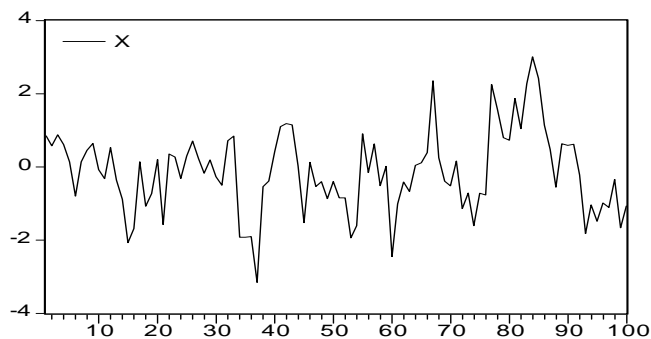
$$\hat{\rho} \approx \frac{\sum_{t=2}^T u_t u_{t-1}}{\sum_{t=2}^T u_{t-1}^2} = \hat{\alpha}_1$$

对于总体参数有  $\rho = \alpha_1$ ，回归模型中误差项  $u_t$  的

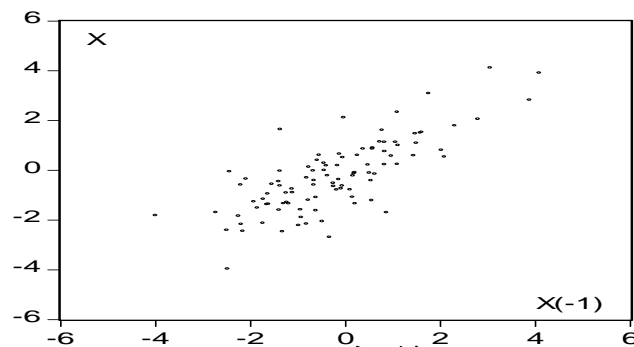
一阶自回归形式可表示为，  $u_t = \rho u_{t-1} + v_t$



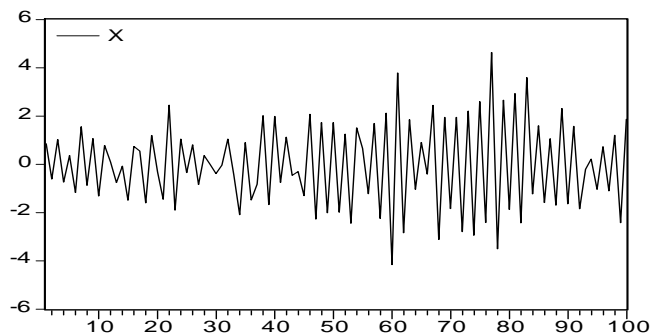
序列的自相关特征分析。给出具有正自相关，负自相关和非自相关三个序列。



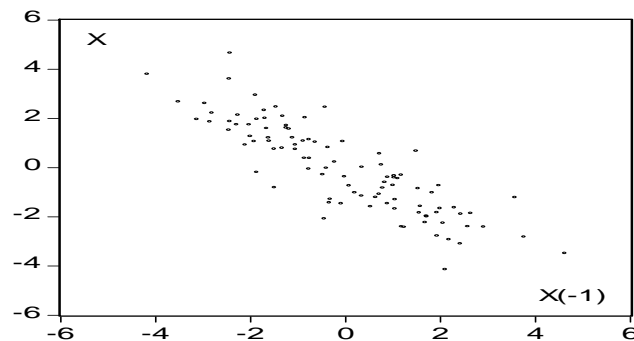
**a. 正相关序列**



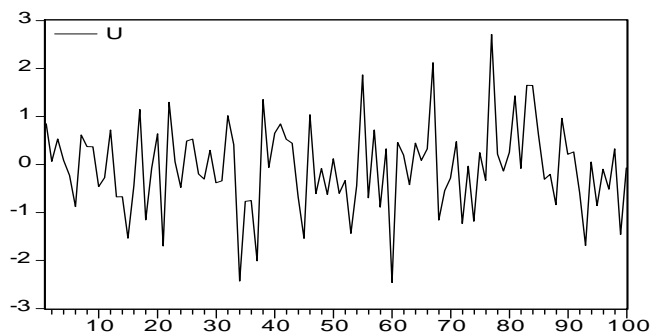
**b. 正相关**



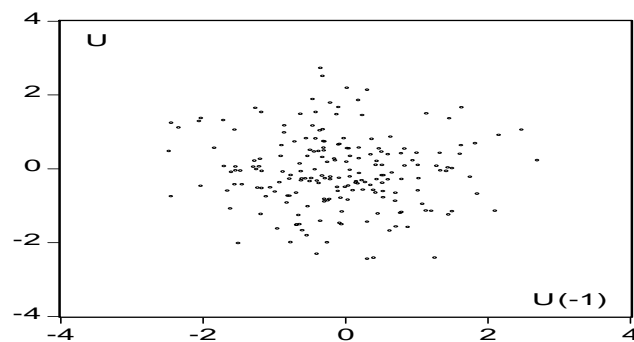
**c. 负相关序列**



**d. 负相关**



**e. 非自相关序列**

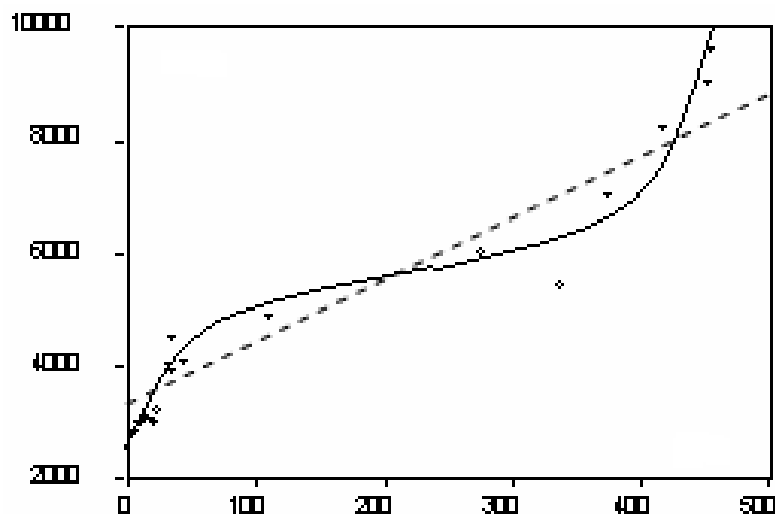


**f. 非自相关**

# 自相关的来源与后果

自相关的来源：

1. 模型的数学形式不妥。



- 2. 惯性。大多数经济时间序列都存在自相关。
- 3. 回归模型中略去了带有自相关的重要解释变量。

## 2、序列相关产生的原因

### (1) 惯性

大多数经济时间数据都有一个明显的特点，就是它的惯性。

GDP、价格指数、生产、就业与失业等时间序列都呈周期性，如周期中的复苏阶段，大多数经济序列均呈上升势，序列在每一时刻的值都高于前一时刻的值，似乎有一种内在的动力驱使这一势头继续下去，直至某些情况（如利率或课税的升高）出现才把它拖慢下来。

## (2) 设定偏误：模型中遗漏了显著的变量

例如：如果对牛肉需求的正确模型应为

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \mu_t$$

其中：Y=牛肉需求量， $X_1$ =牛肉价格， $X_2$ =消费者收入， $X_3$ =猪肉价格。

如果模型设定为：

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + v_t$$

那么该式中的随机误差项实际上是： $v_t = \beta_3 X_{3t} + \mu_t$ ,

于是在猪肉价格影响牛肉消费量的情况下，这种模型设定的偏误往往导致随机项中有一个重要的系统性影响因素，使其呈序列相关性。

### (3) 设定偏误：不正确的函数形式

例如：如果边际成本模型应为：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \mu_t$$

其中：Y=边际成本，X=产出。

但建模时设立了如下模型：

$$Y_t = \beta_0 + \beta_1 X_t + v_t$$

因此，由于  $v_t = \beta_2 X_t^2 + \mu_t$ ，包含了产出的平方对随机项的系统性影响，随机项也呈现序列相关性。

## (4) 蛛网现象

例如，农产品供给对价格的反映本身存在一个滞后期：

$$\text{供给}_t = \beta_0 + \beta_1 \text{价格}_{t-1} + \mu_t$$

意味着，农民由于在年度 $t$ 的过量生产（使该期价格下降）很可能导致在年度 $t+1$ 时削减产量，因此不能期望随机干扰项是随机的，往往产生一种蛛网模式。

## (5) 数据的“编造”

例如，季度数据来自月度数据的简单平均，这种平均的计算减弱了每月数据的波动而引进了数据中的匀滑性，这种匀滑性本身就能使干扰项中出现系统性的因素，从而出现序列相关。

还有就是两个时间点之间的“内插”技术往往导致随机项的序列相关性。

## 二、序列相关性的后果



## 1、参数估计量非有效

- **OLS**参数估计量仍具无偏性
- **OLS**估计量不具有有效性
- 在大样本情况下，参数估计量仍然不具有渐近有效性，这就是说参数估计量不具有 consistency

## 2、变量的显著性检验失去意义

在关于变量的显著性检验中，当存在序列相关时，参数的OLS估计量的方差增大，标准差也增大，因此实际的  $t$  统计量变小，从而接受原假设  $\beta_i=0$  的可能性增大，检验就失去意义。

采用其它检验也是如此。

### 3、模型的预测失效

区间预测与参数估计量的方差有关，在方差有偏误的情况下，使得预测估计不准确，预测精度降低。所以，当模型出现序列相关性时，它的预测功能失效。

### 三、序列相关性的检验

## 1、基本思路

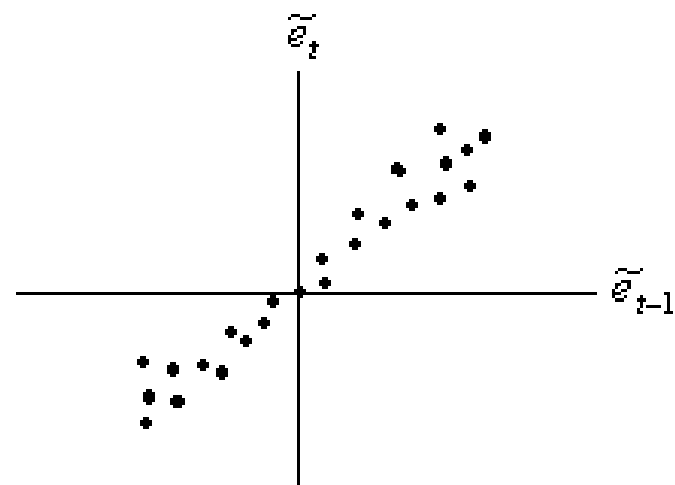
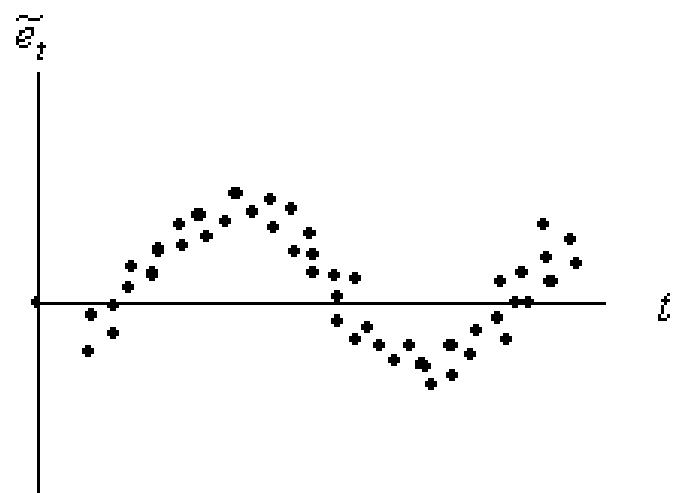
- 序列相关性检验方法有多种，但基本思路是相同的。
- 首先采用普通最小二乘法估计模型，以求得随机误差项的“近似估计量”：

$$\tilde{e}_i = y_i - (\hat{y}_i)_{ols}$$

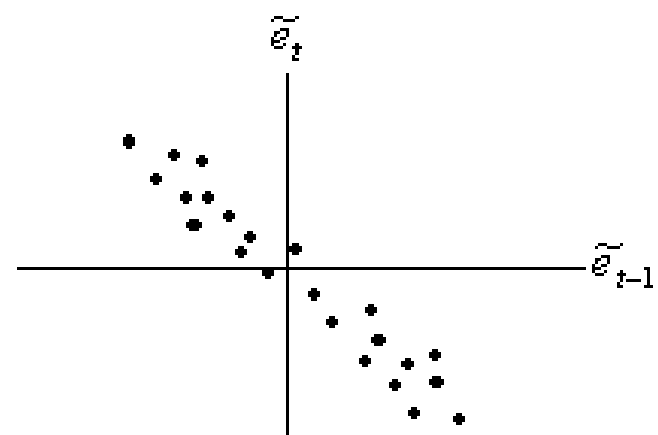
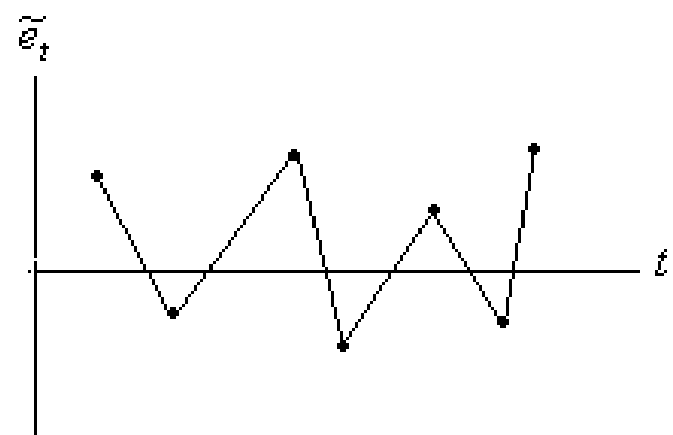
- 然后，通过分析这些“近似估计量”之间的相关性，以达到判断随机误差项是否具有序列相关性的目的。

## 2、图示法

由于残差  $\tilde{e}_i$  可以作为  $\mu_i$  的估计，因此如果  $\mu_i$  存在序列相关，必然会由残差项  $\tilde{e}_i$  反映出来，因此可利用  $\tilde{e}_i$  的变化图形来判断随机项的序列相关性。



正序列相关（正自相关）



负序列相关（负自相关）

## 2、解析法

### (1) 回归检验法

以  $\tilde{e}_i$  为被解释变量，以各种可能的相关量，诸如以  $\tilde{e}_{i-1}$ 、 $\tilde{e}_{i-2}$ 、 $\tilde{e}_i^2$  等为解释变量，建立各种方程：

$$\tilde{e}_i = \rho \tilde{e}_{i-1} + \varepsilon_i \quad i=2, \cdots, n$$

$$\tilde{e}_i = \rho_1 \tilde{e}_{i-1} + \rho_2 \tilde{e}_{i-2} + \varepsilon_i \quad i=3, \cdots, n$$

...



对方程估计并进行显著性检验，如果存在某一种函数形式，使得方程显著成立，则说明原模型存在序列相关性。

- 具体应用时需要反复试算。
- 回归检验法的优点是：

一旦确定了模型存在序列相关性，也就同时知道了相关的形式；

它适用于任何类型的序列相关性问题的检验。

## (2) 杜宾-瓦森 (Durbin-Watson) 检验法

- **D-W**检验是杜宾 (J.Durbin) 和瓦森 (G. S. Watson) 于1951年提出的一种检验序列自相关的方法。

- 该方法的假定条件是:

- (1) 解释变量  $X$  非随机;

- (2) 随机误差项  $\mu_i$  为一阶自回归形式:

$$\mu_i = \rho\mu_{i-1} + \varepsilon_i$$

- (3) 回归模型中不应含有滞后应变量作为解释变量, 即不应出现下列形式:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \gamma Y_{i-1} + \mu_i$$

- (4) 回归含有截距项;

- (5) 没有缺失数据。

- **D.W.统计量**

Durbin 和 Watson 假设:

$H_0 : \rho = 0$  , 即  $\mu_i$  不存在一阶自回归;

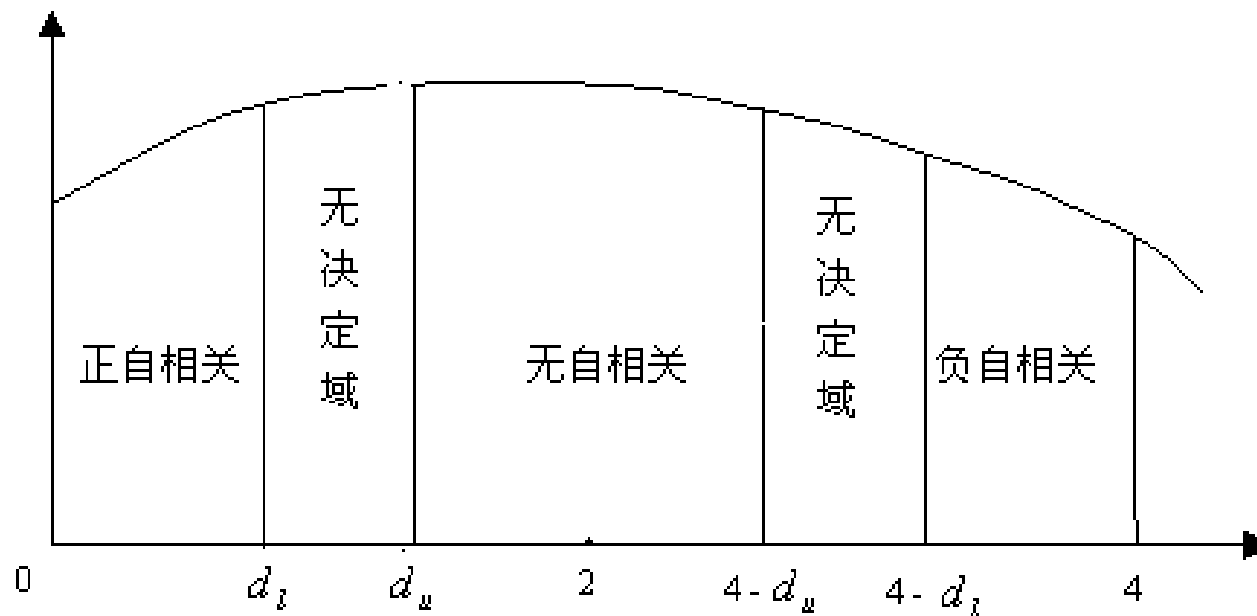
$H_1 : \rho \neq 0$  , 即  $\mu_i$  存在一阶自回归

并构造如下统计量:

$$D.W. = \frac{\sum_{i=2}^n (\tilde{e}_i - \tilde{e}_{i-1})^2}{\sum_{i=1}^n \tilde{e}_i^2} \quad (2.5.5)$$

- 该统计量的分布与出现在给定样本中的 $X$ 值有复杂的关系，因此其精确的分布很难得到。
- 但是，Durbin和Watson成功地导出了临界值的下限 $d_L$ 和上限 $d_U$ ，且这些上下限只与样本的容量 $n$ 和解释变量的个数 $k$ 有关，而与解释变量 $X$ 的取值无关。
- 检验步骤
  - ①计算该统计量的值，
  - ②根据样本容量 $n$ 和解释变量数目 $k$ 查D.W.分布表，得到临界值 $d_L$ 和 $d_U$ ，
  - ③按照下列准则考察计算得到的D.W.值，以判断模型的自相关状态。

若	$0 < D.W. < d_l$	则存在正自相关
	$d_l < D.W. < d_u$	不能确定
	$d_u < D.W. < 4 - d_u$	无自相关
	$4 - d_u < D.W. < 4 - d_l$	不能确定
	$4 - d_l < D.W. < 4$	存在负自相关



- 可以看出，当**D.W.**值在**2**左右时，模型不存在一阶自相关。

**证明：**展开 D.W.统计量：

$$D.W. = \frac{\sum_{i=2}^n \tilde{e}_i^2 + \sum_{i=2}^n \tilde{e}_{i-1}^2 - 2 \sum_{i=2}^n \tilde{e}_i \tilde{e}_{i-1}}{\sum_{i=1}^n \tilde{e}_i^2} \quad (2.5.6)$$

当 n 较大时， $\sum_{i=2}^n \tilde{e}_i^2$ ,  $\sum_{i=2}^n \tilde{e}_{i-1}^2$ ,  $\sum_{i=1}^n \tilde{e}_i^2$  大致相等，则(2.5.6)可以化简为：

$$D.W. \approx 2 \left( 1 - \frac{\sum_{i=2}^n \tilde{e}_i \tilde{e}_{i-1}}{\sum_{i=1}^n \tilde{e}_i^2} \right) \approx 2(1 - \rho)$$

式中,

$$\sum_{i=2}^n \tilde{e}_i \tilde{e}_{i-1} / \sum_{i=1}^n \tilde{e}_i^2 \approx \sum_{i=2}^n \tilde{e}_i \tilde{e}_{i-1} / \sum_{i=2}^n \tilde{e}_i^2 = \rho$$

为一阶自相关模型

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t \quad -1 < \rho < 1$$

的参数估计,

如果存在完全一阶正相关, 即

$\rho=1$ , 则 **D.W. $\approx 0$**

如果存在完全一阶负相关, 即

$\rho=-1$ , 则 **D.W. $\approx 4$**

如果完全不相相关, 即

$\rho=0$ , 则 **D.W. $\approx 2$**

•注意:

(1) 从判断准则看到, 存在一个不能确定的D.W.值区域, 这是这种检验方法的一大缺陷。

(2) D.W.检验虽然只能检验一阶自相关, 但在实际计量经济学问题中, 一阶自相关是出现最多的一类序列相关;

(3) 经验表明, 如果不存在一阶自相关, 一般也不存在高阶序列相关。

所以在实际应用中, 对于序列相关问题一般只进行D.W.检验。



# 自相关检验

## (3) *LM*检验（亦称*BG*检验）法

*LM* 统计量既可检验一阶自相关，也可检验高阶自相关。

*LM* 检验是通过一个辅助回归式完成的，具体步骤如下。

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t$$

考虑误差项为  $n$  阶自回归形式  $u_t = \rho_1 u_{t-1} + \dots + \rho_n u_{t-n} + v_t$

$$H_0: \rho_1 = \rho_2 = \dots = \rho_n = 0$$

用多元回归式得到的残差建立辅助回归式，

$$e_t = \hat{\rho}_1 e_{t-1} + \dots + \hat{\rho}_n e_{t-n} + \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + v_t$$

估计并计算确定系数  $R^2$ 。构造 *LM* 统计量， $LM = TR^2$

若  $LM = TR^2 \leq \chi^2(n)$ ，接受  $H_0$ ；若  $LM = TR^2 > \chi^2(n)$ ，拒绝  $H_0$ 。

## 四、具有序列相关性模型的估计

- 如果模型被检验证明存在序列相关性，则需要发展新的方法估计模型。
- 最常用的方法是广义最小二乘法（GLS: Generalized least squares）、一阶差分法（First-Order Difference）和广义差分法 (Generalized Difference)。

# 自相关的解决方法

1. 如果自相关是由于错误地设定模型的数学形式所致，那么就应当修改模型的数学形式。方法是用残差 $e_t$ 对解释变量的较高次幂进行回归。

2. 如果自相关是由于模型中省略了重要解释变量造成的，那么解决办法就是找出略去的解释变量，把它做为重要解释变量列入模型。

怎样查明自相关是由于略去重要解释变量引起的？一种方法是用残差 $e_t$ 对那些可能影响被解释变量，但又未单列入模型的解释变量回归，并作显著性检验。

只有当以上两种引起自相关的原因都排除后，才能认为误差项 $u_t$ 真正存在自相关。

在这种情况下，解决办法是变换原回归模型，使变换后模型的随机误差项消除自相关。这种估计方法称作广义最小二乘法。

# 自相关的解决方法

注意：

(1) 上式中的  $\beta_1, \dots, \beta_k$  就是原模型中的  $\beta_1, \dots, \beta_k$ 。

$\beta_0^*$  与原模型中的  $\beta_0$  有如下关系，  $\beta_0 = \beta_0^* / (1 - \rho)$

上述方法得到的  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  称作回归系数的

广义最小二乘估计量。

(2) 当误差项  $u_t$  的自相关具有高阶自回归形式时，仍可用  
与上述相类似的方法进行广义差分变换。

(3) 当用广义差分变量回归的结果中仍存在自相关时，  
可以对广义差分变量继续进行广义差分，直至回归  
模型中不存在自相关为止。

# 1、广义最小二乘法

- 对于模型

$$\mathbf{Y}=\mathbf{XB}+\mathbf{N} \quad (2.5.7)$$

如果存在序列相关，同时存在异方差，即有

$$E(\mathbf{N})=0$$

$$Cov(\mathbf{NN}') = E(\mathbf{NN}') = \sigma^2 \mathbf{\Omega}$$

$$\mathbf{\Omega} = \begin{bmatrix} w_1 & w_{12} & \cdots & w_{1n} \\ w_{21} & w_2 & \cdots & w_{2n} \\ & & \ddots & \\ w_{n1} & w_{n2} & \cdots & w_n \end{bmatrix}$$

- 设  $\Omega = \mathbf{D}\mathbf{D}'$

用  $\mathbf{D}^{-1}$  左乘(2.5.7)两边，得到一个新的模型：

$$\mathbf{D}^{-1} \mathbf{Y} = \mathbf{D}^{-1} \mathbf{X}\mathbf{B} + \mathbf{D}^{-1} \mathbf{N} \quad (2.5.8)$$

即  $\mathbf{Y}^* = \mathbf{X}^*\mathbf{B} + \mathbf{N}^*$

该模型具有同方差性和随机误差项互相独立性。

$$\begin{aligned} E(\mathbf{N}^* \mathbf{N}^{*'}) &= E(\mathbf{D}^{-1} \mathbf{N} \mathbf{N}' \mathbf{D}^{-1'}) \\ &= \mathbf{D}^{-1} E(\mathbf{N} \mathbf{N}') \mathbf{D}^{-1'} \\ &= \mathbf{D}^{-1} \sigma^2 \mathbf{W} \mathbf{D}^{-1'} \\ &= \mathbf{D}^{-1} \sigma^2 \mathbf{D} \mathbf{D}' \mathbf{D}^{-1'} \\ &= \sigma^2 \mathbf{I} \end{aligned}$$

- 于是，可以用**OLS**法估计模型(2.5.8)，得

$$\begin{aligned}\hat{\mathbf{B}} &= (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{Y}^* \\ &= (\mathbf{X}' \mathbf{D}^{-1'} \mathbf{D}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{D}^{-1'} \mathbf{D}^{-1} \mathbf{Y} \\ &= (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}' \Omega^{-1} \mathbf{Y}\end{aligned}\tag{2.5.9}$$

- 这就是原模型(2.5.7)的广义最小二乘估计量(**GLS estimators**)，是无偏的、有效的估计量。



- 如何得到矩阵 $\Omega$ ?

仍然是对原模型(2.5.7)首先采用普通最小二乘法，得到随机误差项的近似估计量，以此构成矩阵的估计量 $\hat{\Omega}$ ，即

$$\hat{\Omega} = \begin{bmatrix} \tilde{e}_1^2 & \tilde{e}_1\tilde{e}_2 & \cdots & \tilde{e}_1\tilde{e}_n \\ \tilde{e}_2\tilde{e}_1 & \tilde{e}_2^2 & \cdots & \tilde{e}_2\tilde{e}_n \\ & & \ddots & \\ \tilde{e}_n\tilde{e}_1 & \tilde{e}_n\tilde{e}_2 & \cdots & \tilde{e}_n^2 \end{bmatrix}$$

- 可行的广义最小二乘法 (**FGLS**, **Feasible Generalized Least Squares**)

文献中常见的术语

如果能够找到一种方法，求得到  $\Omega$  的估计量，使得**GLS**能够实现，都称为**FGLS**

前面提出的方法，就是**FGLS**

## 2、一阶差分法

一阶差分法是将原模型

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \quad i=1,2,\cdots,n$$

变换为

$$\Delta Y_i = \beta_1 \Delta X_i + \mu_i - \mu_{i-1} \quad i=2,\cdots,n \quad (2.5.10)$$

其中

$$\Delta Y_i = Y_i - Y_{i-1}$$

...

- 如果原模型存在完全一阶正自相关，即在

$$\mu_i = \rho\mu_{i-1} + \varepsilon_i$$

中， $\rho=1$ 。(2.5.10)可变换为：

$$\Delta Y_i = \beta_1 \Delta X_i + \varepsilon_i$$

由于 $\varepsilon_i$ 不存在序列相关，该差分模型满足应用**OLS**法的基本假设，用**OLS**法估计可得到原模型参数的无偏的、有效的估计量。

- 即使对于非完全一阶正相关的情况，只要存在一定程度的一阶正相关，差分模型就可以有效地加以克服。

### 3、广义差分法

如果原模型存在：

$$\mu_i = \rho_1 \mu_{i-1} + \rho_2 \mu_{i-2} + \cdots + \rho_l \mu_{i-l} + \varepsilon_i \quad (2.5.11)$$

可以将原模型变换为：

$$Y_i - \rho_1 Y_{i-1} - \cdots - \rho_l Y_{i-l} = \beta_0 (1 - \rho_1 - \cdots - \rho_l) + \beta_1 (X_i - \rho_1 X_{i-1} - \cdots - \rho_l X_{i-l}) + \varepsilon_i$$
$$i = 1+l, 2+l, \cdots, n$$

(2.5.12)

模型(2.5.12)为**广义差分模型**，该模型不存在序列相关问题。采用OLS法估计可以得到原模型参数的无偏、有效的估计量。

**广义差分法**可以克服所有类型的序列相关带来的问题，一阶差分法是它的一个特例。

## 4、随机误差项相关系数 $\rho$ 的估计

- 应用广义差分法，必须已知不同样本点之间随机误差项的相关系数 $\rho_1, \rho_2, \dots, \rho_l$ 。实际上，人们并不知道它们的具体数值，所以必须首先对它们进行估计。
- 常用的方法有：
  - (1) 科克伦-奥科特 (Cochrane-Orcutt) 迭代法。
  - (2) 杜宾 (durbin) 两步法

## (1) 科克伦-奥科特迭代法

首先，采用OLS法估计原模型

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

得到的随机误差项的“近似估计值”，并以之作为观测值采用OLS法估计下式

$$\mu_i = \rho_1 \mu_{i-1} + \rho_2 \mu_{i-2} + \dots + \rho_L \mu_{i-L} + \varepsilon_i$$

得到  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_l$ ，作为随机误差项的相关系数  $\rho_1, \rho_2, \dots, \rho_l$  的**第一次估计值**。

其次，将上述  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_l$  代入广义差分模型

$$Y_i - \rho_1 Y_{i-1} - \dots - \rho_l Y_{i-l} = \beta_0 (1 - \rho_1 - \dots - \rho_l) + \beta_1 (X_i - \rho_1 X_{i-1} - \dots - \rho_l X_{i-l}) + \varepsilon_i$$

$$i = 1 + l, 2 + l, \dots, n$$

并对之进行 OLS 估计，得到  $\hat{\beta}_0$ 、 $\hat{\beta}_1$ 。

再次，将  $\hat{\beta}_0$ 、 $\hat{\beta}_1$  代回原模型，计算出原模型随机误差项的新的“近似估计值”，并以之作为模型

$$\mu_i = \rho_1 \mu_{i-1} + \rho_2 \mu_{i-2} + \dots + \rho_l \mu_{i-l} + \varepsilon_i$$

的样本观测值，采用 OLS 法估计该方程，得到  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_l$ ，作为相关系数  $\rho_1, \rho_2, \dots, \rho_l$  的第二次估计值。



类似地，可进行第三次、第四次迭代。

- 关于迭代的次数，可根据具体的问题来定。
- 一般是事先给出一个精度，当相邻两次 $\rho_1, \rho_2, \dots, \rho_L$ 的估计值之差小于这一精度时，迭代终止。
- 实践中，有时只要迭代两次，就可得到较满意的结果。两次迭代过程也被称为科克伦-奥科特两步法。

## (2) 杜宾 (durbin) 两步法

该方法仍是先估计  $\rho_1, \rho_2, \dots, \rho_L$ , 再对差分模型进行估计。

第一步, 变换差分模型为下列形式:

$$Y_i = \rho_1 Y_{i-1} + \dots + \rho_l Y_{i-l} + \beta_0 (1 - \rho_1 - \dots - \rho_l) + \beta_1 (X_i - \rho_1 X_{i-1} - \dots - \rho_l X_{i-l}) + \varepsilon_i$$
$$i = 1+l, 2+l, \dots, n$$

(2.5.13)

采用 OLS 法估计该方程, 得各  $Y_j (j = i-1, i-2, \dots, i-l)$  前的系数  $\rho_1, \rho_2, \dots, \rho_l$  的估计值  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_l$ 。

**第二步**，将估计的  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_l$  代入差分模型

$$Y_i - \rho_1 Y_{i-1} - \dots - \rho_l Y_{i-l} = \beta_0 (1 - \rho_1 - \dots - \rho_l) + \beta_1 (X_i - \rho_1 X_{i-1} - \dots - \rho_l X_{i-l}) + \varepsilon_i$$

$$i = 1 + l, 2 + l, \dots, n$$

采用 OLS 法估计，得到参数  $\beta_0 (1 - \hat{\rho}_1 - \dots - \hat{\rho}_l), \beta_1$  的估计量，记为  $\hat{\beta}_0^*, \hat{\beta}_1^*$ 。

于是：

$$\hat{\beta}_0 = \hat{\beta}_0^* / (1 - \hat{\rho}_1 - \dots - \hat{\rho}_l), \quad \hat{\beta}_1 = \hat{\beta}_1^*$$

## 5、应用软件中的广义差分法

- 在Eview/TSP软件包下，广义差分采用了科克伦-奥科特（Cochrane-Orcutt）迭代法估计 $\rho$ 。
- 在解释变量中引入AR(1)、AR(2)、...，即可得到参数和  $\rho_1$ 、 $\rho_2$ 、...的估计值。其中AR(m)表示随机误差项的m阶自回归。在估计过程中自动完成了  $\rho_1$ 、 $\rho_2$ 、...的迭代。

## 6、虚假序列相关问题

- 由于随机项的序列相关往往是在模型设定中遗漏了重要的解释变量或对模型的函数形式设定有误，这种情形可称为**虚假序列相关**，应在模型设定中排除。
- 避免产生虚假序列相关性的措施是在开始时建立一个“一般”的模型，然后逐渐剔除确实不显著的变量。

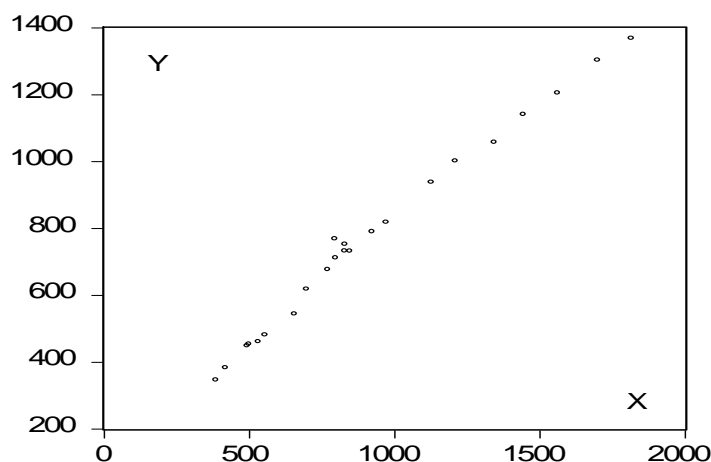
# 案例分析

例 天津市城镇居民人均消费与人均可支配收入的关系。

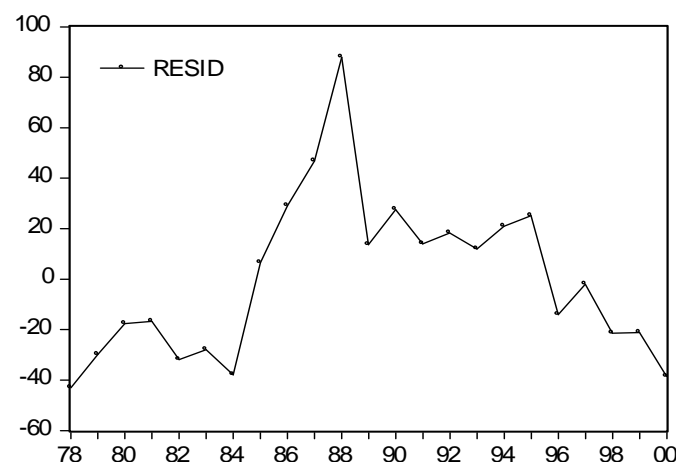
改革开放以来，天津市城镇居民人均消费性支出（**CONSUM**），人均可支配收入（**INCOME**）以及消费价格定基指数（**PRICE**）数据（1978~2000年）见表6.2。现在研究人均消费与人均可支配收入的关系。

先定义不变价格（1978=1）的人均消费性支出（ $Y_t$ ）和人均可支配收入（ $X_t$ ）。令  $Y_t = \text{CONSUM} / \text{PRICE}$ ,  $X_t = \text{INCOME} / \text{PRICE}$

假定所建立的回归模型形式是  $Y_t = \beta_0 + \beta_1 X_t + u_t$



$Y_t$  和  $X_t$  散点图



残差图

## 例 天津市城镇居民人均消费与人均可支配收入的关系。

(1) 估计线性回归模型并计算残差。

$$\hat{Y}_t = 111.44 + 0.7118 X_t$$

(6.5)      (42.1)       $R^2 = 0.9883, \text{s.e.} = 32.8, DW = 0.60, T = 23$

(2) 分别用DW、LM统计量检验误差项  $u_t$  是否存在自相关。

已知  $DW = 0.60$ ，若给定  $\alpha = 0.05$ ，查附表4，得DW检验临界值  $d_L = 1.26$ ， $d_U = 1.44$ 。因为  $DW = 0.60 < 1.26$ ，认为误差项  $u_t$  存在严重的正自相关。

LM (BG) 自相关检验辅助回归式估计结果是

$$e_t = 0.6790 e_{t-1} + 3.1710 - 0.0047 X_t + v_t$$

(3.9)      (0.2)      (- 0.4)       $R^2 = 0.43, DW = 2.00$

$LM = T R^2 = 23 \times 0.43 = 9.89$ 。因为  $\chi^2_{0.05(1)} = 3.84$ ， $LM = 9.89 > 3.84$ ，所以LM检验结果也说明误差项存在一阶正自相关。

**EViews的LM自相关检验操作：** 点击最小二乘回归窗口中的**View**键，选**Residual Tests/Serial Correlation LM Test...**，在随后弹出的滞后期对话框中给出最大滞后期。点击**OK**键。

## 例 天津市城镇居民人均消费与人均可支配收入的关系。

(3) 用广义最小二乘法估计回归参数。

首先估计自相关系数  $\hat{\rho}$ 。

$$\hat{\rho} = 1 - \frac{DW}{2} = 1 - \frac{0.60}{2} = 0.70$$

对原变量做广义差分变换。令

$$GDY_t = Y_t - 0.70 Y_{t-1}$$

$$GDX_t = X_t - 0.70 X_{t-1}$$

以  $GDY_t, GDX_t$ , (1979 ~ 2000 年), 为样本再次回归, 得

$$GDY_t = 45.2489 + 0.6782 GDX_t$$

$$(3.7) \quad (20.0)$$

$$R^2 = 0.95, DW = 2.31, T = 22, (1979 \sim 2000)$$

$DW = 2.31$ 。查附表 4,  $d_L = 1.26, d_U = 1.43$ 。

因为  $DW = 2.31 < (4 - 1.43) = 2.57$ ,

依据判别规则, 误差项已消除自相关。



例 天津市城镇居民人均消费与人均可支配收入的关系。



$$\hat{\beta}_0^* = 45.2489。$$

$$\hat{\beta}_0 = \hat{\beta}_0^* / (1 - \hat{\rho}) = 45.2489 / (1 - 0.70) = 150.8297$$

则原模型的广义最小二乘估计结果是

$$\hat{Y}_t = 150.8297 + 0.6782 X_t$$

经济含义是天津市城镇居民人均消费性支出

平均占人均可支配收入的 **67.82%**。

## 例 天津市城镇居民人均消费与人均可支配收入的关系。

注意：

(1)  $R^2$ 值有所下降。不应该不相信估计结果。原因是两个回归式所用变量不同，所以不可以直接比较确定系数 $R^2$ 的值。

(2) 两种估计方法的回归系数有差别。计量经济理论认为回归系数广义最小二乘估计量优于误差项存在自相关的OLS估计量。所以0.6782应该比0.7118更可信。特别是最近几年，天津市城镇居民人均收入的人均消费边际系数为0.6782更可信。

(3) 用EViews生成新变量的方法：

从工作文件主菜单中点击Quick键，选择Generate Series ...功能。打开生成序列（Generate Series by Equation）对话框。在对话框中输入如下命令（每次只能输入一个命令），

**$Y = \text{CONSUM} / \text{PRICE}$**

**$X = \text{INCOME} / \text{PRICE}$**

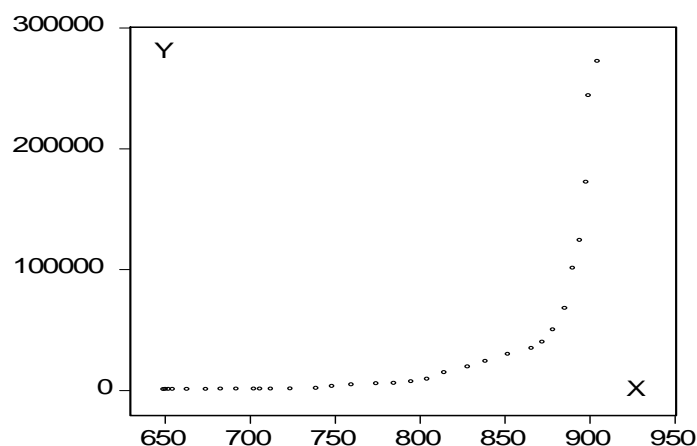
按OK键。变量Y和X将自动显示在工作文件中。

## 例 天津市保费收入和人口的回归关系

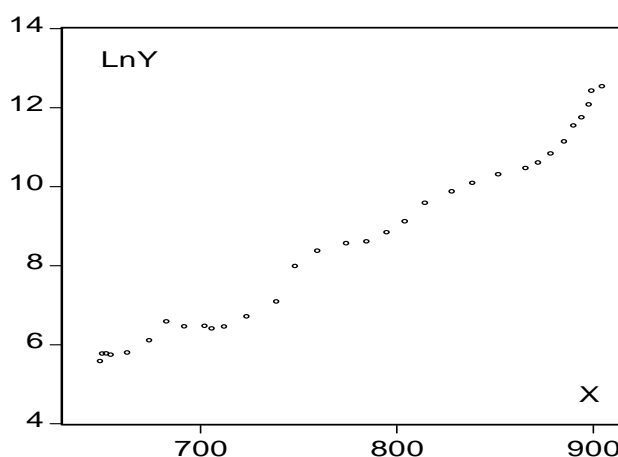
本案例主要用来展示当模型误差项存在2阶自回归形式的自相关时，怎样用广义差分法估计模型参数。

1967~1998年天津市的保险费收入（ $Y_t$ ，万元）和人口（ $X_t$ ，万人）数据散点图见图。 $Y_t$ 与 $X_t$ 的变化呈指数关系。对 $Y_t$ 取自然对数。 $\ln Y_t$ 与 $X_t$ 的散点图见图。

可以在 $\ln Y_t$ 与 $X_t$ 之间建立线性回归模型。 $\ln Y_t = \beta_0 + \beta_1 X_t + u_t$



$Y_t$ 和 $X_t$ 散点图



$\ln Y_t$ 和 $X_t$ 散点图

（第2版181页）

（第3版155页）

## 例 天津市保费收入和人口的回归关系

(1) 估计线性回归模型并计算残差。

$$\hat{\ln Y_t} = -11.18 + 0.0254 X_t$$

(-20.9) (37.2)

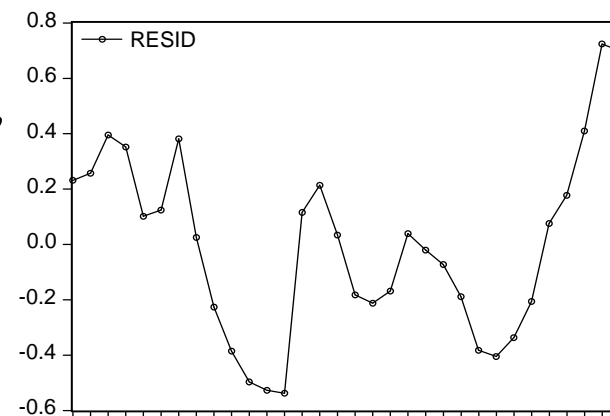
$$R^2 = 0.9788, \text{s.e.} = 0.34, DW = 0.36, T = 32, (1967-1998)$$

(2) 检验误差项  $u_t$  是否存在自相关。

已知  $DW = 0.36$ ，若给定  $\alpha = 0.05$ ，

查附表 4， $d_L = 1.37$ ， $d_U = 1.50$ 。因为  $DW = 0.36 < 1.37$ ，

依据判别规则，认为误差项  $u_t$  存在严重的正自相关。



## 例 天津市保费收入和人口的回归关系

对残差序列的拟合发现， $u_t$ 存在二阶自相关。回归式如下。

$$e_t = 1.186 e_{t-1} - 0.467 e_{t-2} + v_t$$

$$(6.9) \quad (-2.5) \quad R^2 = 0.71, \text{ s.e.} = 0.19, DW = 1.97 \quad (1969-1998)$$

误差项具有二阶自回归形式的自相关。

(3) 用广义差分法消除自相关。

首先推导二阶自相关 $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + v_t$ 条件下的广义差分变换式。设模型为

$$\ln Y_t = \beta_0 + \beta_1 X_t + u_t$$

写出上式的滞后1期、2期表达式并分别乘以 $\phi_1$ 、 $\phi_2$ ,

$$\phi_1 \ln Y_{t-1} = \phi_1 \beta_0 + \phi_1 \beta_1 X_{t-1} + \phi_1 u_{t-1}$$

$$\phi_2 \ln Y_{t-2} = \phi_2 \beta_0 + \phi_2 \beta_1 X_{t-2} + \phi_2 u_{t-2}$$

用以上3式做如下运算，

$$\ln Y_t - \phi_1 \ln Y_{t-1} - \phi_2 \ln Y_{t-2} = \beta_0 - \phi_1 \beta_0 - \phi_2 \beta_0 + \beta_1 X_t - \phi_1 \beta_1 X_{t-1} - \phi_2 \beta_1 X_{t-2} + u_t - \phi_1 u_{t-1} - \phi_2 u_{t-2}$$

将2阶自相关关系式， $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + v_t$ ，代入上式并整理，得

$$(\ln Y_t - \phi_1 \ln Y_{t-1} - \phi_2 \ln Y_{t-2}) = \beta_0 (1 - \phi_1 - \phi_2) + \beta_1 (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2}) + v_t$$

## 例 天津市保费收入和人口的回归关系

二阶广义差分变换应该是

$$GDLnY_t = LnY_t - \phi_1 LnY_{t-1} - \phi_2 LnY_{t-2}$$

$$GDX_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2}$$

$LnY_t$ 和 $X_t$ 的广义差分变换应该是

$$GDLnY_t = LnY_t - 1.186 LnY_{t-1} + 0.467 LnY_{t-2}$$

$$GDX_t = X_t - 1.186 X_{t-1} + 0.467 X_{t-2}$$

广义最小二乘回归结果是

$$\hat{GDLnY_t} = -3.246 + 0.0259 GDX_t$$

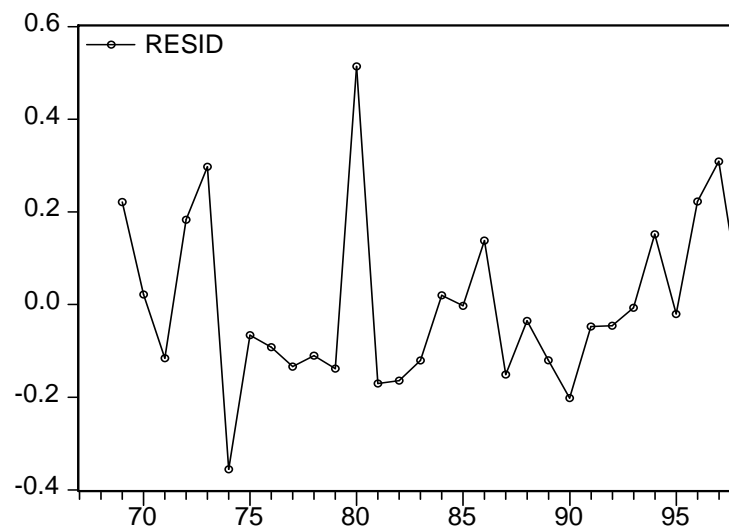
$$(-10.0) \quad (17.9) \quad R^2 = 0.92, DW = 1.99, (1969-1998)$$

$$\beta_0 = -3.246 / (1 - \phi_1 - \phi_2) = -3.246 / (1 - 1.186 + 0.467) = -11.55$$

原模型的广义最小二乘估计结果是  $LnY_t = -11.55 + 0.0259 X_t$

广义最小二乘估计值0.0259比最小二乘估计值0.0254值可信。

经济含义是每增加1万人， $LnY_t$ 增加0.0259，即保费增加1.0262万元。



## 五、案例:地区商品出口模型

## 某地区商品出口总值与国内生产总值的数据

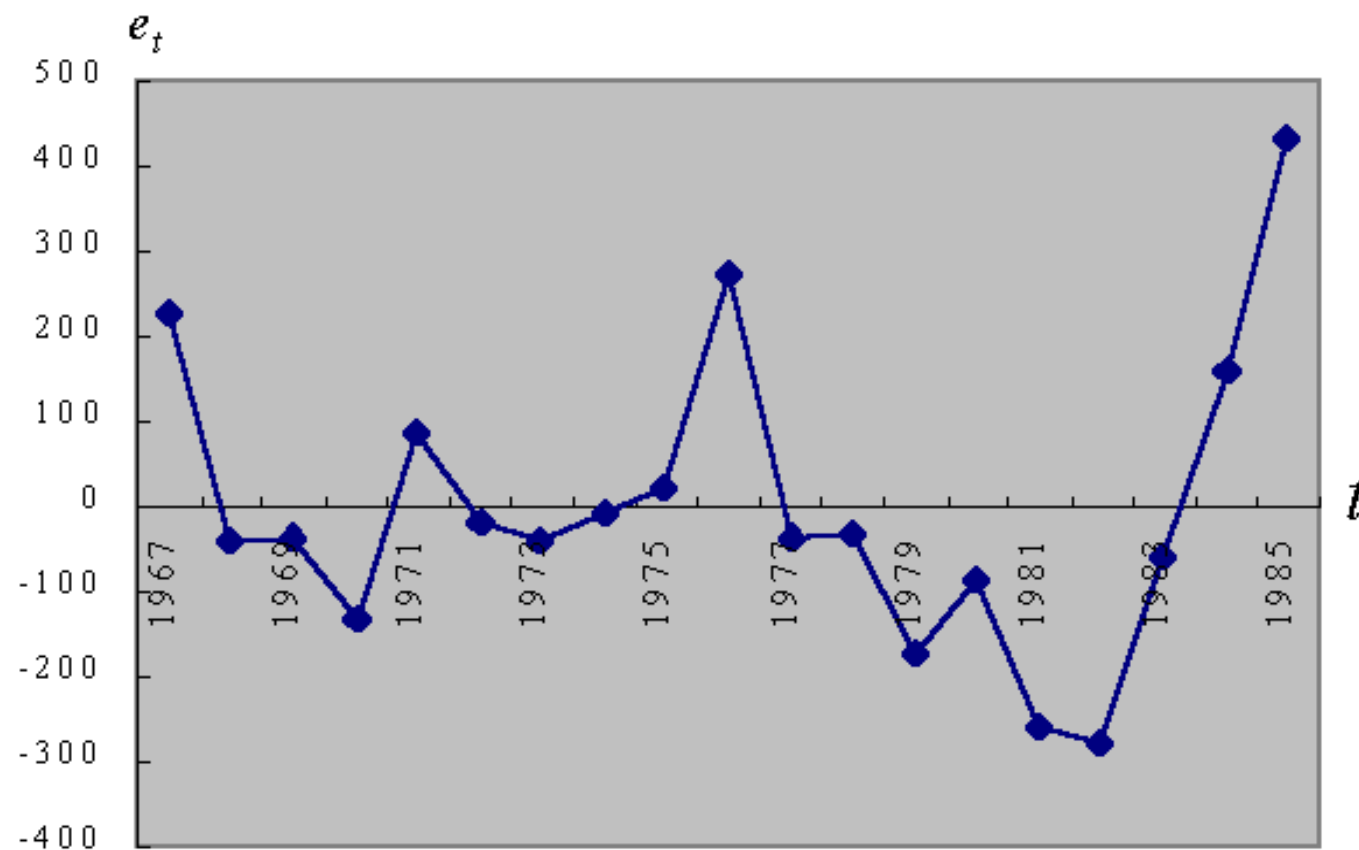
单位：万元

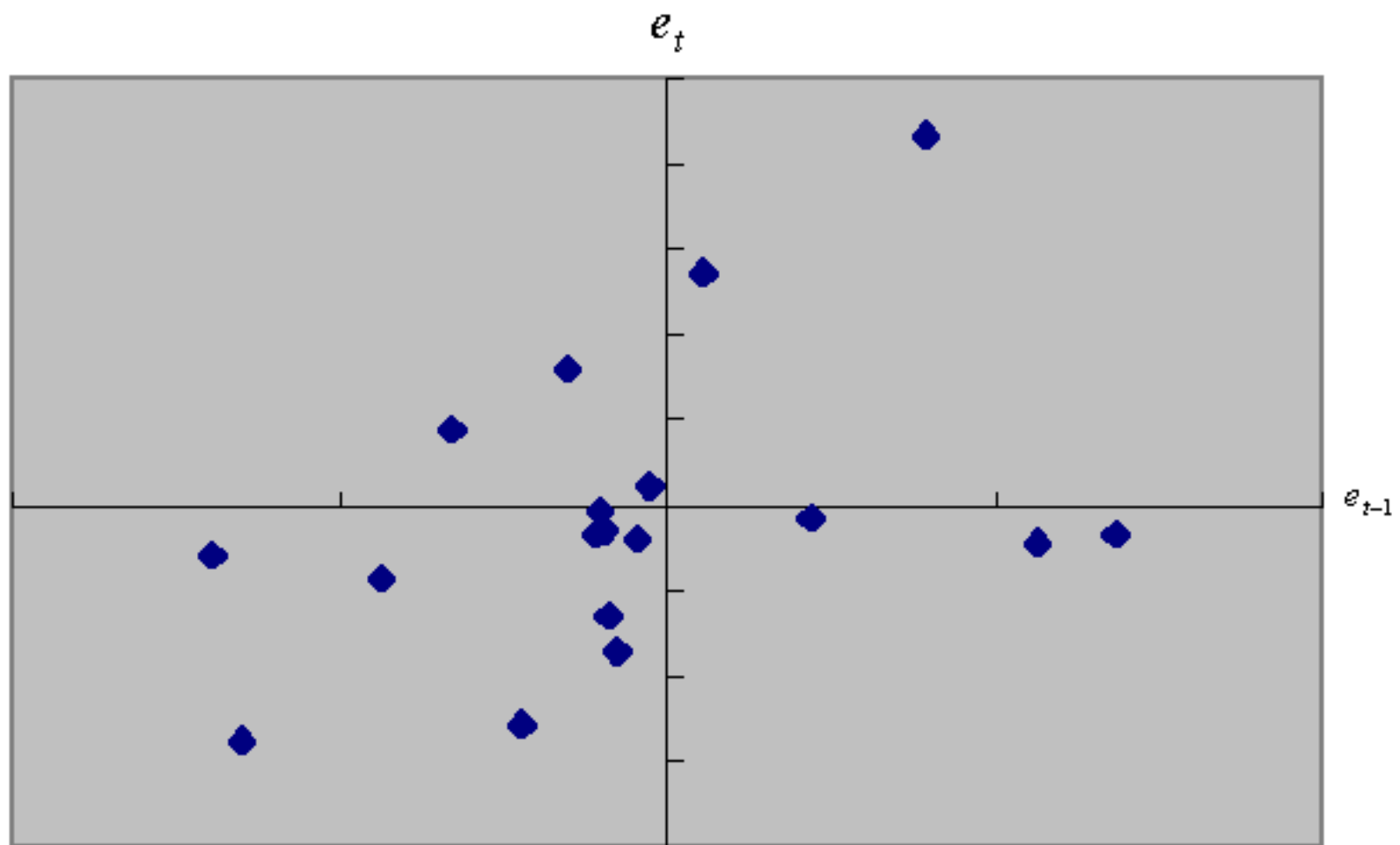
年份	出口 Y	国内生产总值 X	年份	出口 Y	国内生产总值 X
1967	4010	22418	1977	5628	29091
1968	3711	22308	1978	5736	29450
1969	4004	23319	1979	5946	30705
1970	4151	24180	1980	6501	32372
1971	4569	24893	1981	6549	33152
1972	4582	25310	1982	6705	33764
1973	4697	25799	1983	7104	34411
1974	4753	25886	1984	7609	35429
1975	5062	26868	1985	8100	36200
1976	5669	28134			



## 2、序列相关性检验

### (1) 图示法检验





## (2) D.W. 检验

回归结果：

$$\hat{Y}_t = -2531.83 + 0.28 X_t$$

$$(-9.34) \quad (30.11)$$

$$r^2=0.9816, \quad R^2=0.9805 \quad D.W.=0.9505$$

在5%在显著性水平下， $n=19$ ， $k=2$ (包含常数项)，查表得 $d_L=1.18$ ， $d_U=1.40$ ，

由于 $DW=0.9505 < d_L$ ，故存在正自相关。

### 3、自相关的处理

#### (1) 一阶差分法

$$\Delta \hat{Y}_t = 0.3185 \Delta X_t + \varepsilon_t$$

(6.8098)

$$R^2=0.4747, \quad D.W.=1.8623$$

由于DW>du=1.39(注：样本容量为18个)，已不存在自相关。

## (2) 广义差分法

### ①采用杜宾两步法估计 $\rho$

#### 1) 估计模型

$$Y_t = \beta_0^* + \rho Y_{t-1} + \beta_1^* X_t + \beta_2^* X_{t-1} + \varepsilon_{t-1}$$

得：

$$\hat{Y}_t = -1334.79 + 0.5939Y_{t-1} + 0.3348X_t - 0.2109X_{t-1}$$

$$(-1.86) \quad (2.01) \quad (3.41) \quad (-1.53)$$

$$r^2=0.9862, \quad R^2=0.9832, \quad D.W.=1.6282$$

2)将  $\hat{\rho}=0.5939$  代入差分模型

$$Y_t - 0.5939Y_{t-1} = \beta_0^* + \beta_1^*(X_t - 0.5939X_{t-1}) + \varepsilon_t$$

OLS 法估计得：

$$Y_t - 0.5939Y_{t-1} = -1351.01 + 0.3083(X_t - 0.5939X_{t-1}) + \varepsilon_t$$

$$(-5.53) \quad (15.58)$$

$$r^2=0.9382, \quad R^2=0.9343, \quad D.W.=1.6570$$

由于DW>=1.39(注：样本容量为19-1=18个)，已不存在自相关。于是原模型估计式为：

$$\hat{Y}_t = -3326.79 + 0.3083 X_t$$

## ②采用科克伦-奥科特迭代法估计 $\rho$

一阶广义差分的结果：

$$\hat{Y}_t = -3354.72 + 0.3092X_t + 0.6040AR[1]$$

$$(-3.330) \quad (9.417) \quad (2.122)$$

$$r^2=0.9861, \quad R^2=0.9842, \quad D.W.=1.6715$$

由于 $DW > d_u = 1.39$ (注：样本容量为18个)，已不存在自相关。

二阶广义差分的结果：

$$\hat{Y}_t = -3131.94 + 0.3025X_t + 0.6713AR[1] - 0.0542AR[2]$$

$$(-2.41) \quad (7.06) \quad (2.07) \quad (-0.15)$$

$$r^2=0.9844, \quad R^2=0.9808, \quad D.W.=1.8411$$

由于  $DW > d_u = 1.38$  (注：样本容量为  $19-2=17$  个)，已不存在自相关。

但由于  $AR[2]$  前的系数的  $t$  值为  $-0.15$ ，在 5% 的显著性水平下并不显著，说明随机干扰项不存在二阶序列相关性，模型中应去掉  $AR[2]$  项。