计量经济学

第六章　自相关

引子　T检验和F检验一定就可靠吗?

研究居民储蓄存款Y与居民收入X的关系：

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

用普通最小二乘法估计其参数，结果为

$$\hat{Y}_t = 27.9123 + 0.3524 X_t$$

（1.8690）　(0.0055)

t=　(14.9343)　(64.2069)

$R^2 = 0.9966$　　F=4122.531

检验结果：回归系数的标准误差非常小，t统计量较大，说明居民收入X对居民储蓄存款Y的影响非常显著。同时可决系数也非常高，F统计量=4122.531，也表明模型异常的显著。

但此估计结果可能是虚假的，t统计量和F统计量都被虚假地夸大，因此所得结果是不可信的。为什么?

经济系统中的自相关

一、自相关的概念

二、自相关产生的背景与原因

三、自相关性的后果

四、自相关性的检验

五、自相关问题的处理方法

第一节　什么是自相关

一、自相关的概念

自相关（auto correlation），又称序列相关（serial correlation）是指总体回归模型的随机误差项之间存在相关关系。即不同观测点上的误差项彼此相关。

一阶自相关系数

$$\rho = \frac{\sum_{t=2}^{n} u_t u_{t-1}}{\sqrt{\sum_{t=2}^{n} u_t^2}\sqrt{\sum_{t=2}^{n} u_{t-1}^2}}$$

式 (6.1)

自相关系数ρ的定义与普通相关系的公式形式相同，ρ的取值范围为

$-1 \leqslant \rho \leqslant 1$。

式（6.1）中u_{t-1}是u_t滞后一期的随机误差项，因此，将式（6.1）计算的自相关系数ρ称为一阶自相关系数。

二、自相关产生的原因

自相关产生的原因

1、经济系统的惯性

2、经济活动的滞后效应

3、数据处理造成的相关

4、蛛网现象

5、模型设定偏误

原因1—经济系统的惯性

自相关现象大多出现在时间序列数据中，而经济系统的经济行为都具有时间上的惯性。

如GDP、价格、就业等经济指标都会随经济系统的周期而波动。例如，在经济高涨时期，较高的经济增长率会持续一段时间，而在经济衰退期，较高的失业率也会持续一段时间，这种现象就会表现为经济指标的自相关现象。

原因2－ 经济活动的滞后效应

滞后效应是指某一指标对另一指标的影响不仅限于当期而是延续若干期。由此带来变量的自相关。

例如，居民当期可支配收入的增加，不会使居民的消费水平在当期就达到应有水平，而是要经过若干期才能达到。因为人的消费观念的改变客观上存在自适应期。

原因3—数据处理造成的相关

因为某些原因对数据进行了修整和内插处理，在这样的数据序列中就会有自相关。

例如，将月度数据调整为季度数据，由于采用了加合处理，修匀了月度数据的波动，使季度数据具有平滑性，这种平滑性产生自相关。对缺失的历史资料，采用特定统计方法进行内插处理，使得数据前后期相关，产生了自相关。

原因4—蛛网现象

蛛网现象是微观经济学中的一个概念。它表示某种商品的供给量受前一期价格影响而表现出来的某种规律性，即呈蛛网状收敛或发散于供需的均衡点。

许多农产品的供给呈现为蛛网现象，供给对价格的反应要滞后一段时间，因为供给需要经过一定的时间才能实现。如果时期t的价格P_t低于上一期的价格P_{t-1}，农民就会减少时期t+1的生产量。如此则形成蛛网现象，此时的供给模型为:

$$\text{供给}_t=\beta_1+\beta_2\text{价格}_{t-1}+u_t \tag{6.2}$$

原因5—模型设定偏误

如果模型中省略了某些重要的解释变量或者模型函数形式不正确，都会产生系统误差，这种误差存在于随机误差项中，从而带来了自相关。由于该现象是由于设定失误造成的自相关，因此，也称其为虚假自相关。

例如，应该用两个解释变量解释Y，即：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad (6.3)$$

而建立模型时，模型设定为：

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \quad (6.4)$$

则X_{3t}对Y_t的影响在式（6.4）中便归入随机误差项u_t中，由于X_{3t}在不同观测点上是相关的，这就造成了u_t在不同观测点是相关的，呈现出系统模式，此时u_t是自相关的。

模型形式设定偏误也会导致自相关现象。如将U形成本曲线设定为线性成本曲线，则必定会导致自相关。由设定偏误产生的自相关是一种虚假自相关，可通过改变模型设定予以消除。

自相关关系主要存在于时间序列数据中，但是在横截面数据中，也可能会出现自相关,通常称其为空间自相关（Spatial auto correlation）。

例如，在消费行为中，一个家庭、一个地区的消费行为可能会影响另外一些家庭和另外一些地区，就是说不同观测点的随机误差项可能是相关的。

多数经济时间序列在较长时间内都表现为上升或下降的超势，因此大多表现为正自相关。但就自相关本身而言是可以为正相关也可以为负相关。

三、自相关的表现形式

自相关的性质可以用自相关系数ρ的符号判断自相关的方向，即$\rho<0$为负相关，$\rho>0$为正相关。当$|\rho|$接近1时，表示相关的程度很高。但自相关是u_1，u_2，…，u_n序列自身的相关，因n个随机误差项的关联形式不同而具有不同的自相关形式。自相关大多出现在时间序列数据中。

对于样本观测期为n的时间序列数据，可得到总体回归模型(PRF)的随机误差项为u_1，u_2，…，u_n，如果自相关形式为

$$u_t = \rho u_{t-1} + v_t \quad -1 < \rho < 1 \qquad (6.5)$$

其中ρ为自相关系数，v_t为经典误差项，即$E(v_t) = 0$，$Var(v_t) = \sigma^2$，$Cov(v_t, v_{t+s}) = 0$，$s \neq 0$。则式（6.5）称为一阶自回归模式，记为$AR(1)$。因为模型（6.5）中u_{t-1}是u_t滞后一期的值，因此称为一阶。式（6.5）中的ρ也称为一阶自相关系数。

如果式（6.5）中的随机误差项v_t不是经典误差项，即v_t中包含有u_t的成份，如包含有u_{t-2}则需将u_{t-2}显含在回归模型中，其为

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t' \quad (6.6)$$

其中，ρ_1为一阶自相关系数，ρ_2为二阶自相关系数，v_t'是经典误差项。式（6.6）称为二阶自回归模式，记为AR(2)。一般地，如果u_1，u_2，…，u_t之间的关系为

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_m u_{t-m} + v_t \quad (6.7)$$

其中，v_t为经典误差项。则称式(6.7)为m阶自回归模式，记为AR(m)。

在经济计量分析中，通常采用一阶自回归形式，即假定自回归形式为一阶自回归AR(1)。

第二节 自相关的后果

自相关的后果

- 对参数估计的影响
- 对模型检验的影响
- 对模型预测的影响

一、对参数估计的影响

在有自相关的条件下，仍然使用普通最小二乘法将低估估计量 $\hat{\beta}_2$ 的方差 $Var(\hat{\beta}_2)$ 。

$\hat{\sigma}^2 = \dfrac{\Sigma e_i^2}{n-k}$ 将低估真实的 σ^2 。

对于一元线性回归模型:

$$Y = \beta_1 + \beta_2 X + u \qquad (6.8)$$

假定随机误差项*u*存在一阶自相关:

$$u_t = \rho u_{t-1} + v_t \qquad (6.9)$$

其中，u_t 为现期随机误差，u_{t-1} 为前期随机误差。v_t 是经典误差项，满足零均值 $E(v_t)=0$ ，同方差 $Var(v_t)=\sigma_v^2$，无自相关 $E(v_t v_s)=0(t \neq s)$ 的假定。

将随机误差项 u_t 的各期滞后值：

$$u_{t-1} = \rho u_{t-2} + v_{t-1}, u_{t-2} = \rho u_{t-3} + v_{t-2}, \cdots,$$

逐次代入式（**6.9**）可得：

$$u_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \cdots = \sum_{r=0}^{\infty} \rho^r v_{t-r}$$

(6.10)

式（6.10）表明随机误差项 u_t 可表示为独立同分布的随机误差序列 v_t ，v_{t-1}，v_{t-2}，… 的加权和，权数分别为1，ρ ，ρ^2，…，当0< ρ <1时，这些权数是随时间推移而呈几何衰减的；而当-1< ρ <0时，这些权数是随时间推移而交错振荡衰减的。

由式（**6.10**）可推得:

$$E(u_t) = \sum_{r=0}^{\infty} \rho^r E(v_{t-r}) = 0 \qquad (6.11)$$

$$Var(v_t) = \sum_{r=0}^{\infty} \rho^{2n} Var(v_{t-r}) = \frac{\sigma_v^2}{1-\rho^2} = \sigma_u^2 \qquad (6.12)$$

式（**6.11**）、式（**6.12**）表明，在 u_t 为一阶自回归的相关形式时，随机误差 u_t 依然是零均值、同方差的误差项。

由于现期的随机误差项 v_t 并不影响回归模型中随机误差项 u_t 的以前各期值 u_{t-k} **(k>0)**，所以 v_t 与 u_{t-k} 不相关，即有 $E(v_t u_{t-k})=0$ 。因此，由式（**6.9**）可得随机误差项 u_t 与其以前各期 u_{t-k} 的协方差分别为:

$$Cov(u_t, u_{t-1}) = E(u_t u_{t-1})$$

$$= E[(\rho u_{t-1} + v_t) u_{t-1}]$$

$$= \rho E(u_{t-1}^2) + E(v_t u_{t-1}) = \rho\sigma_u^2 = \frac{\rho\sigma_v^2}{1-\rho^2}$$

(6.13)

$$Cov(u_t, u_{t-2}) = E(u_t u_{t-2})$$

$$= E[(\rho u_{t-1} + v_t) u_{t-2}]$$

$$= E[(\rho^2 u_{t-2} + \rho v_{t-1} + v_t) u_{t-2}]$$

$$= \rho^2 E(u_{t-2}^2)$$

$$= \frac{\rho^2 \sigma_v^2}{1 - \rho^2} \qquad (6.14)$$

以此类推，可得：

$$Cov(u_t, u_{t-k}) = \rho^k Var(u_{t-k}) = \frac{\rho^k \sigma_v^2}{1-\rho^2} \qquad (6.15)$$

这些协方差分别称为随机误差项 u_t 的一阶自协方差、二阶自协方差和 k 阶自协方差。

对于一元线性回归模型（**6.8**），当 ***u*** 为经典误差项时，普通最小二乘估计量 $\hat{\beta}_2$ 的方差为：

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum (X-\bar{X})^2} = \frac{\sigma^2}{\sum x^2} \qquad (6.16)$$

当式（**6.8**）的随机误差项***u***有自相关时，$\hat{\beta}_2$ 依然是无偏的，即 $E(\hat{\beta}_2)=\beta_2$，这一点在普通最小二乘法无偏性证明中可以看到。因为，无偏性证明并不需要 u 满足无自相关的假定。那么，最小二乘估计量 $\hat{\beta}_2$ 是否是有效的，即在所有线性无偏估计量中方差最小呢？下面我们将以说明。

式（6.8）中的最小二乘估计量为

$$\hat{\beta}_2 = \frac{\Sigma x_t y_t}{\Sigma x_t^2} = \beta_2 + \frac{\Sigma x_t u_t}{\Sigma x_t^2} \tag{6.17}$$

因此:

$$Var(\hat{\beta}_2) = E(\hat{\beta}_2 - \beta_2)^2 = E\left(\frac{\Sigma x_t u_t}{\Sigma x_t^2}\right)^2$$

$$= \frac{\sigma_u^2}{\sum_{t=1}^{n} x_t^2}\left(1 + 2\rho\frac{\sum_{t=1}^{n-1} x_t x_{t+1}}{\sum_{t=1}^{n} x_t^2} + 2\rho^2\frac{\sum_{t=1}^{n-2} x_t x_{t+2}}{\sum_{t=1}^{n} x_t^2} + \cdots + 2\rho^{n-1}\frac{x_1 x_n}{\sum_{t=1}^{n} x_t^2}\right)$$

(6.18)

当存在自相关时，普通最小二乘估计量不再是最佳线性无估计量，即它在线性无偏估计量中不是方差最小的。在实际经济系统中，通常存在正的自相关，即ρ>0，同时X 序列自身也呈正相关，因此式(6.18)右边括号内的值通常大于0。因此，在有自相关的条件下，仍然使用普通最小二乘法将低估估计量 $\hat{\beta}_2$ 的方差 $Var(\hat{\beta}_2)$。

$\hat{\sigma}^2 = \Sigma e_i^2 / (n-k)$ 将低估真实的 σ^2。

二、对模型检验的影响

考虑自相关时的检验

对模型检验的影响

忽视自相关时的检验

（一）考虑自相关时的检验

由于 $Var(\hat{\beta}_2)$ 并不是所有线性无偏估计量中最小的，使用t检验判断回归系数的显著性时就可能得到错误的结论。

t检验的统计量为 $\longrightarrow$ $t = \dfrac{\text{估计值}}{\text{估计量的标准误}} = \dfrac{\hat{\beta}_2}{Se(\hat{\beta}_2)}$

由于 $SE(\hat{\beta}_2)$ 的错误夸大，得到的t统计量就可能小于临界值 $t_{\alpha/2}$，从而得到参数β不显著的结论。而这一结论可能是不正确的。

（二）忽视自相关时的检验

如果我们忽视自相关问题依然假设经典假定成立，使用 $Var(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_t^2}$ ，将会导致错误结果。

当ρ >0，即有正相关时，对所有的j都有ρ >0。另外回归模型中的解释变量在不同时期通常是正相关的，对于x_t和x_{t+j}来说 $\sum X_t X_{t+j}$ 是大于0的。

因此，由式(6.18)可以看出，普通最小二乘法的方差 $Var(\hat{\beta}_2)=\dfrac{\sigma^2}{\Sigma x_t^2}$ 通常会低估 $\hat{\beta}_2$ 的真实方差。当ρ较大和X_t有较强的正自相关时，普通最小二乘估计量的方差会有很大偏差，我们会夸大估计量的估计精度，即得到较小的标准误。

因为 $\hat{\beta}_2$的标准误是 $\hat{\beta}_2$的标准差的估计值，在有自相关时，普通最小二乘估计 $\hat{\beta}_2$的标准误就不可靠了。

一个被低估了的标准误意味着一个较大的t统计量。因此，当$\rho>0$时，通常t统计量都很 大。这种有偏的t统计量是不能用来判断回归系数的显著性的。

综上所述，在自相关情形下，无论考虑自相关，还是忽视自相关，通常的回归系统显著性的t检验都将是无效的。

类似地，由于自相关的存在，参数的最小二乘估计量是无效的，使得F 检验和R^2检验也是不可靠的。

三、对模型预测的影响

模型预测的精度决定于抽样误差和总体误差项的方差 σ^2 。抽样误差来自于对 $\hat{\beta}_j$ 的估计，在自相关情形下，$\hat{\beta}_j$ 的方差的最小二乘估计变得不可靠，由此必定加大抽样误差。同时，在自相关情形下，对 σ^2 的估计 $\hat{\sigma}^2 = \Sigma e_i^2 / n-k$ 也会不可靠。由此可看出，影响预测精度的两大因素都会因自相关的存在而加大不确定性，使预测的置信区间不可靠，从而降低预测的精度。

第三节　自相关的检验

当随机误差项存在自相关时会给普通最小二乘法的应用带来严重的后果。因此，如何诊断回归模型是否存在自相关就致关重要。

自相关的检验

图示检验法

DW检验法

一、图示检验法

图示法是一种直观的诊断方法，它是把给定的回归模型直接用普通最小二乘法估计参数，求出残差项e_t ,e_t作为u_t随机项的真实估计值，再描绘e_t的散点图，根据散点图来判断e_t的相关性。残差e_t的散点图通常有两种绘制方式 。

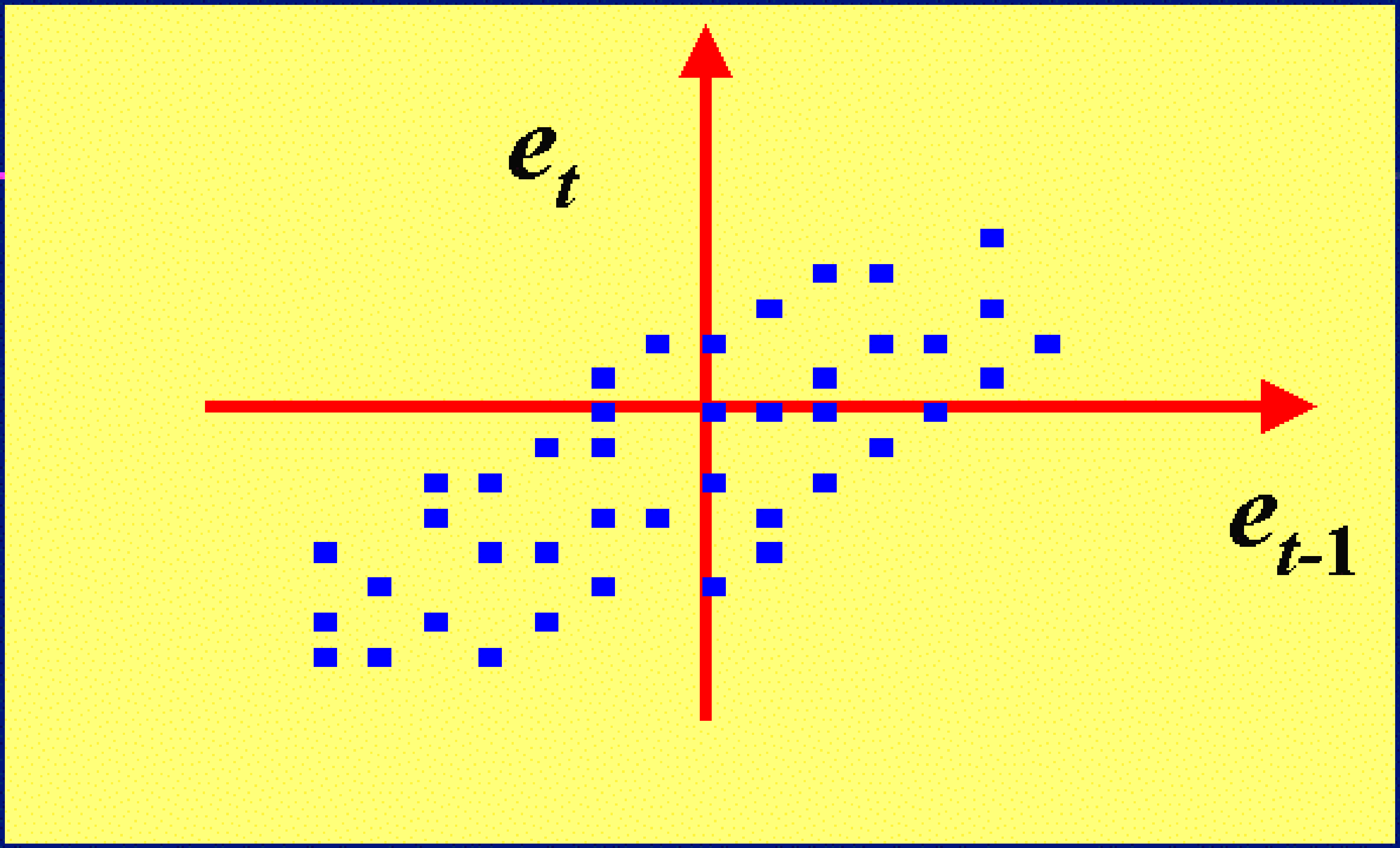


图 6.1　e_t与e_{t-1}的关系

绘制e_{t-1}, e_t的散点图。用(e_{t-1}, e_t)（$t = 1, 2, \cdots n$）作为散布点绘图，如果大部分点落在第Ⅰ、Ⅲ象限，表明随机误差项u_t存在着正自相关，如图6.1所示。

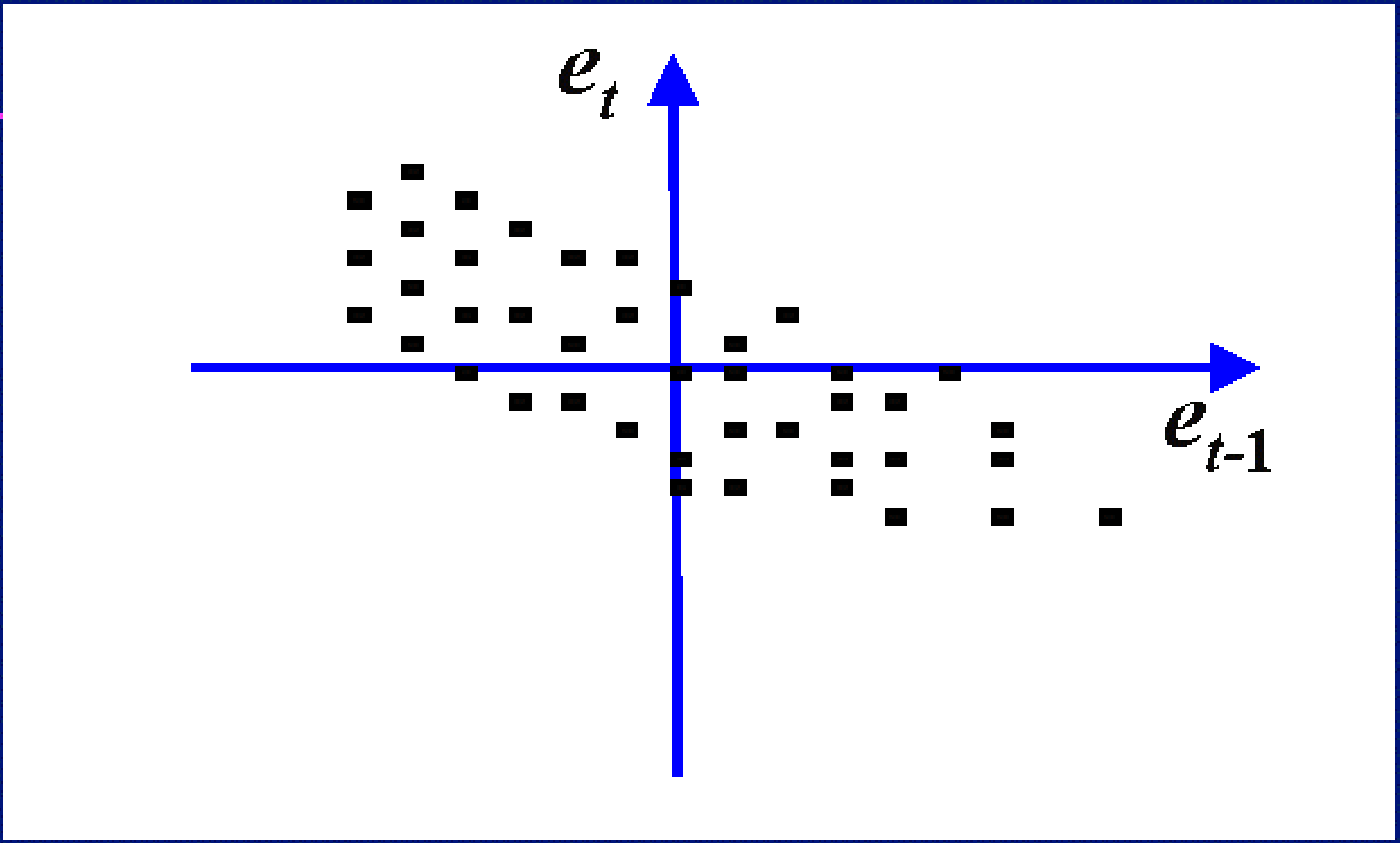


图 6.2　e_t与e_{t-1}的关系

如果大部分点落在第Ⅱ、Ⅳ象限，那么随机误差项u_t存在着负自相关，如图6.2所示。

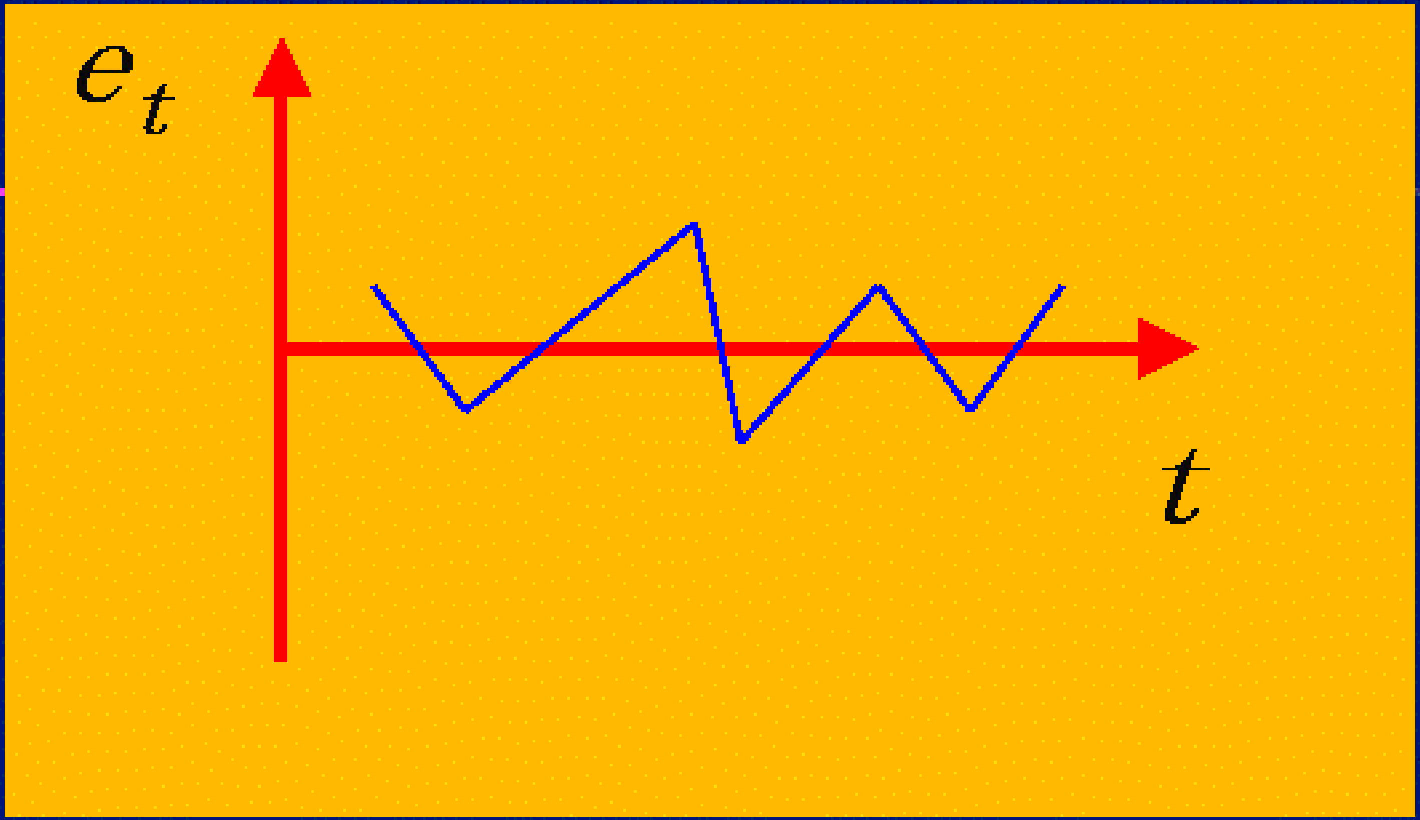


图 6.3　e_t的分布

按照时间顺序绘制回归残差项 e_t 的图形。如果e_t **(t=1,2,…,n)**随着 t 的变化逐次有规律地变化，呈现锯齿形或循环形状的变化，就可断言 e_t存在相关，表明存在着自相关；如果e_t随着 t的变化逐次变化并不断地改变符号，那么随机误差项u_t存在负自相关；如图**6.3**所示。

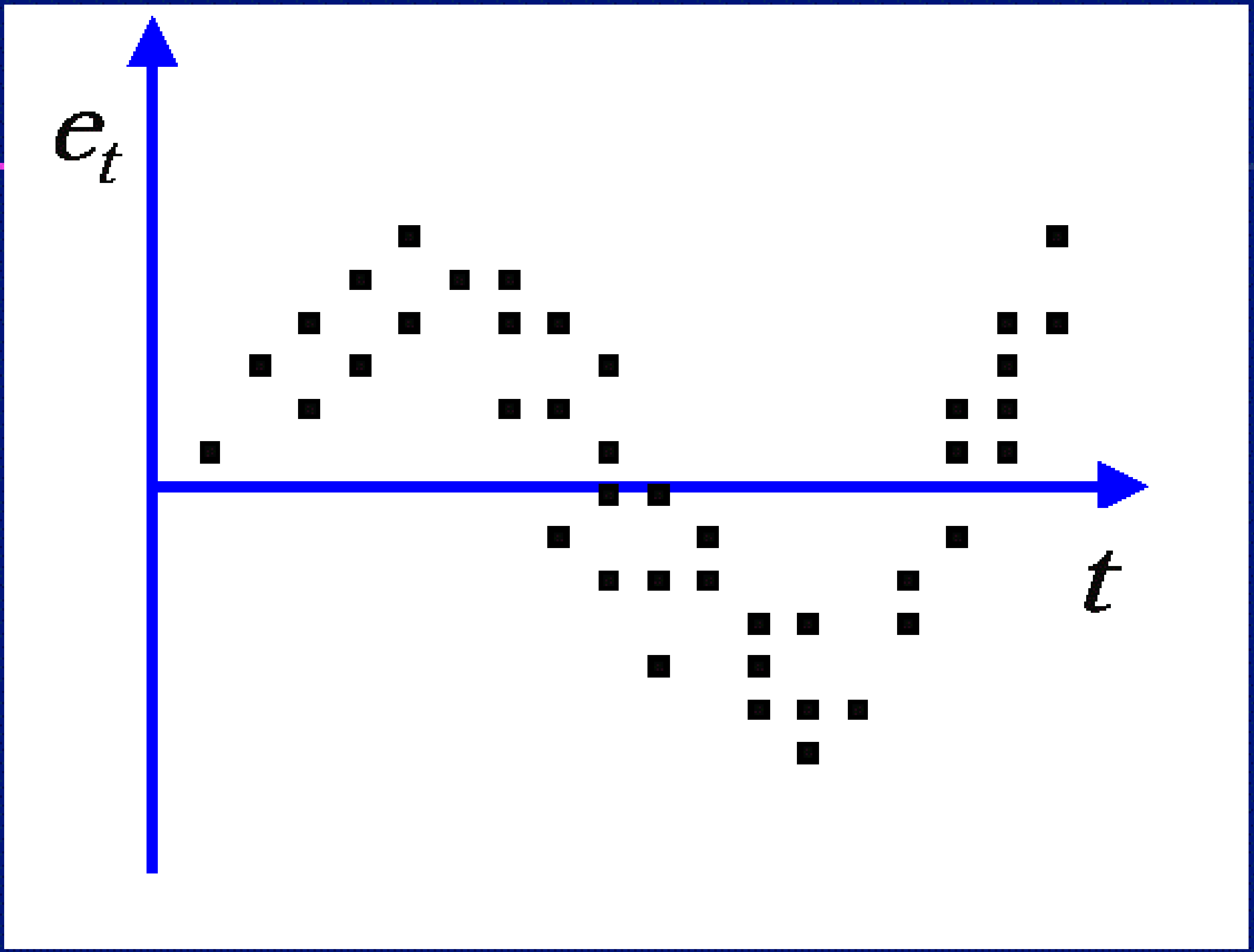


图 6.4　e_t的分布

如果e_t随着t的变化逐次变化并不频繁地改变符号，而是几个正的e_t后面跟着几个负的，则表明随机误差项存u_t在正自相关，如图6.4所示。

二、DW检验法

*DW*检验是J.Durbin(杜宾)和G.S.Watson(沃特森)于1951年提出的一种适用于小样本的检验方法。*DW*检验只能用于检验随机误差项具有一阶自回归形式的自相关问题。这种检验方法是建立经济计量模型中最常用的方法，一般的计算机软件都可以计算出*DW*值。

随机误差项的一阶自回归形式为：

$$u_t = \rho u_{t-1} + v_t$$ (6.20)

为了检验序列的相关性，构造的原假设是：

$$H_0: \rho = 0$$ (6.21)

为了检验上述假设，构造**DW** 统计量首先要求出回归估计式的残差**et** 定义**DW**统计量为：

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$ (6.22)

DW值的取值范围

$$DW=\frac{\sum_{t=2}^{n}e_t^2+\sum_{t=2}^{n}e_{t-1}^2-2\sum_{t=2}^{n}e_te_{t-1}}{\sum_{t=1}^{n}e_t^2} \quad (6.23)$$

在认为：$\sum_{t=2}^{n}e_t^2\approx\sum_{t=2}^{n}e_{t-1}^2\approx\sum_{t=1}^{n}e_t^2$ 则：

$$DW\approx 2\left[1-\frac{\sum_{t=2}^{n}e_te_{t-1}}{\sum_{t=1}^{n}e_t^2}\right] \quad (6.24)$$

由于 $\hat{\rho}\approx\frac{\sum_{t=2}^{n}e_te_{t-1}}{\sum_{t=1}^{n}e_t^2}$

因此，$DW \approx 2(1-\hat{\rho})$ 。所以，***DW***值与 $\hat{\rho}$ 的对应关系如表**6.1**所示。

表 6.1

$\hat{\rho}$	***DW***
-1	**4**
(-1,0)	**(2,4)**
0	**2**
(0,1)	**(0,2)**
1	**0**

由上述讨论可知***DW***的取值范围为： $0 \leqslant DW \leqslant 4$

根据样本容量 ***n*** 和解释变量的数目 ***k'*** （不包括常数项)查**DW**分布表，得临界值 ***dL*** 和 ***dU*** ，然后依下列准则考察计算得到的**DW**值，以决定模型的自相关状态。

表6.2 DW 检验决策规则

$0 \leqslant DW \leqslant d_L$	误差项$u_1,u_2,\ldots,u_n$间存在正相关
$d_L < DW \leqslant d_U$	不能判定是否有自相关
$d_U < \mathrm{DW} < 4 - d_U$	误差项$u_1,u_2,\ldots,u_n$间无自相关
$4 - d_U \leqslant DW < 4 - d_L$	不能判定是否有自相关
$4 - d_L \leqslant DW \leqslant 4$	误差项$u_1,u_2,\ldots,u_n$间存在负相关

表6.2可以用坐标图更加直观地表示出来:

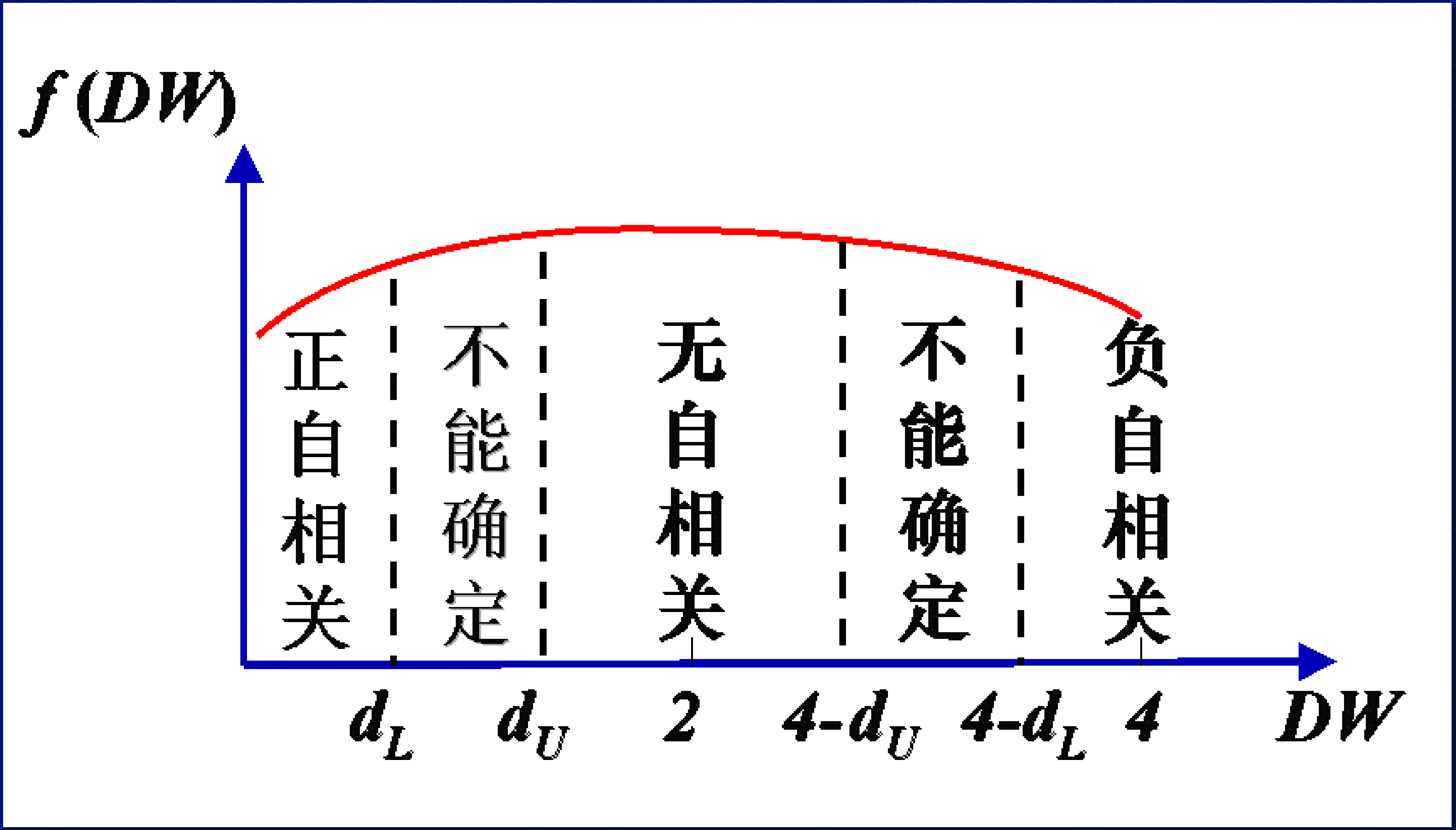


图6.5 *DW* 检验示意图

需要注意的是，*DW*检验尽管有着广泛的应用，但也有明显的缺点和局限性。

① DW检验有两个不能确定的区域，一旦DW值落在这两个区域，就无法判断。这时，只有增大样本容量或选取其他方法；

② DW统计量的上、下界表要求$n \geqslant 15$，这是因为样本如果再小，利用残差就很难对自相关的存在性做出比较正确的诊断；

③ DW检验不适应随机误差项具有高阶序列相关的检验；

④ 只适用于有常数项的回归模型并且解释变量中不能含滞后的被解释变量。

第四节　自相关的补救

对于虚假自相关，由于是模型设定偏误造成的，只能通过改变模型的设定去消除。对于设定正确的模型，如随机误差项有自相关，则为真实的自相关。对于真实的自相关可采用如下方法予以消除。

- 广义差分法
- Cochrane – Orcutt迭代法
- 其它方法：一阶差分法
  德宾两步法

一、广义差分法

对于自相关的结构已知的情形可采用广义差分法解决。

由于随机误差项u_t是不可观测的，通常我们假定u_t为一阶自回归形式，即 $u_t = \rho u_{t-1} + v_t$

（6.25） 其中，$|\rho| < 1$，v_t为经典误差项。

当自相关系数为已知时，使用广义差分法，自相关问题就可彻底解决。我们以一元线性回归模型为例说明广义差分法的应用。

对于一元线性回归模型

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (6.26)$$

将模型（6.26）滞后一期可得

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad (6.27)$$

用ρ乘式（6.27）两边，得

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_t + \rho u_{t-1} \qquad (6.28)$$

用式（6.26）减去式（6.28）可得

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + u_t - \rho u_{t-1} \quad (6.29)$$

模型，随机误差项无序列相关。式（6.29）中，$u_t - \rho u_{t-1} = v_t$是经典误差项。因此，模型（6.29）已经是经典线性回归

令 $$Y_t^* = Y_t - \rho Y_{t-1}, X_t^* = X_t - \rho X_{t-1}$$

$$\beta_1^* = \beta_1(1-\rho)$$

则式（6.29）可表示为：

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + v_t \qquad (6.30)$$

对模型（6.30）使用普通最小二乘估计就会得到参数估计的最佳线性无偏估计量。

式（6.29）称为广义差分方程，因为被解释变量与解释变量均为现期值减去前期值的一部分，由此而得名。

在进行广义差分时，解释变量X与被解释变量Y均以差分形式出现，因而样本容量由n减少为$n-1$，即丢失了第一个观测值。如果样本容量较大，减少一个观测值对估计结果影响不大。但是，如果样本容量较小，则对估计精度产生较大的影响。此时，可采用普莱斯－温斯滕（Prais-Winsten）变换，将第一个观测值变换为

$$Y_1\sqrt{1-\rho^2} \text{和} X_1\sqrt{1-\rho^2}$$

补充到差分序列 Y_t^*, X_t^* 中，再使用普通最小二乘法估计参数。

二、Cochrane － Orcutt 迭代法

在实际应用中，自相关系数 ρ 往往是未知的，ρ 必须通过一定的方法估计。最简单的方法是据 ***DW*** 统计量估计 ρ 。由 ***DW*** 与 ρ 的关系可知：

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

但是，式（6.31）得到的是一个粗略的结果，$\hat{\rho}$ 是对 ρ 精度不高的估计。其根本原因在于我们对有自相关的回归模型使用了普通最小二乘法。为了得到 ρ 的精确的估计值 $\hat{\rho}$ ，人们通常采用科克伦—奥克特（Cochrane—Orcutt）迭代法。

该方法利用残差e_t去估计未知的ρ。对于一元线性回归模型

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (6.32)$$

假定u_t为一阶自回归形式，即：

$$u_t = \rho u_{t-1} + v_t \quad (6.33)$$

Econometrics

科克伦－奥克特迭代法估计 ρ 的步骤如下：

1 使用普遍最小二乘法估计模型

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (6.32)$$

并获得残差： $e_t^{(1)}$

2 利用残差$e_t^{(1)}$做如下的回归

$$e_t^{(1)} = \hat{\rho}^{(1)} e_{t-1}^{(1)} + v_t \quad (6.34)$$

3 利用式（6.34）得到的 $\hat{\rho}^{(1)}$ ，对模型（6.32）进行广义差分，即

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

(6.35)

令

$$\begin{cases} Y_t^* = Y_t - \hat{\rho}^{(1)} Y_{t-1} \\ X_t^* = X_t - \hat{\rho}^{(1)} X_{t-1} \\ \beta_1^* = \beta_1 (1 - \hat{\rho}^{(1)}) \end{cases}$$

对式（6.35）使用普通最小二乘法，可得样本回归函数为:

$$\hat{Y}_t^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_t^* + e_t^{(2)} \quad (6.36)$$

4

因为据式（6.34）得到 $\hat{\rho}^{(1)}$ 并不是对 ρ 的最佳估计，必须进一步迭代，寻求最佳估计。由前一步估计的结果有：

$$\hat{\beta}_1 = \hat{\beta}_1^* / (1 - \hat{\rho}^{(1)}) \text{和} \hat{\beta}_2 = \hat{\beta}_2^*$$

将 $\hat{\beta}_1, \hat{\beta}_2$ 代入原回归方程（6.32）,求得新的残差如下：

$$e_t^{(3)} = Y_t - \beta_1 - \beta_2 X_t \qquad (6.37)$$

5 利用残差 $e_t^{(3)}$ 做如下的回归

$$e_t^{(3)} = \hat{\rho}^{(2)} e_{t-1}^{(3)} + v_t \quad (6.38)$$

这里得到的 $\hat{\rho}^{(2)}$ 就是 ρ 的第二轮估计值。

我们并不能确认 $\hat{\rho}^{(2)}$ 是否是ρ 的最佳估计值，我们还要继续估计ρ 的第三轮估计值 $\hat{\rho}^{(3)}$ 。这种方法是迭代法，当估计的 $\hat{\rho}^{(k)}$ 与 $\hat{\rho}^{(k+1)}$ 相差很小时，就找到了ρ 的最佳估计值。在模型实践中，通常采用科克伦—奥克特两步法。第一步，据式（6.34）估计ρ 。第二步，用估计值 $\hat{\rho}$ 做广义差分，估计广义差分方程的参数。这种两步法与上述迭代方法结果很相近。

三、其它方法简介

（一）一阶差分法

一阶差分法是消除序列相关的一种简单有效的方法。我们仍以一元线性回归模型来说明一阶差分法的应用。

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

式中，***ut***为一阶自回归**AR(1)**。将模型（**6.39**）变换为：

$$\Delta Y_t = \beta_2 \Delta X_t + u_t - u_{t-1}$$

如果原模型存在完全一阶正自相关，即 ρ **=1**则

$$u_t = u_{t-1} + v_t \qquad \textbf{(6.41)}$$

其中，***vt*** 为经典误差项。则式（**6.40**）的随机误差项为经典误差项，无自相关问题。对式（**6.40**）使用普通最小二乘法估计参数，可得到最佳线性无偏估计量。

（二）德宾两步法

当自相关系数未知时，也可采用德宾提出的两步法，消除自相关。将广义差分方程（6.29）表示为：

$$Y_t = \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + \rho Y_{t-1} + v_t$$

采用如下的两个步骤消除自相关。

第一步，把式（6.42）作为一个多元回归模型，使用普通最小二乘法估计参数。把Y_{t-1}的回归系数$\hat{\rho}$看作ρ的一个估计值。

第二步，求得$\hat{\rho}$后，使用$\hat{\rho}$进行广义差分，求得序列：

$$Y_t^* = Y_t - \hat{\rho} Y_{t-1} \quad 和 \quad X_t^* = X_t - \hat{\rho} X_{t-1}$$

然后使用普通最小二乘法对广义差分方程估计参数，求得最佳线性无偏估计量。

第五节　案例分析

研究范围：中国农村居民收入—消费模型 (1985~2003)

研究目的：消费模型是研究居民消费行为的工具和手段。通过消费模型的分析可判断居民消费边际消费倾向，而边际消费倾向是宏观经济系统中的重要参数。

建立模型

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

Y_t—居民消费，X_t—居民收入，u_t—随机误差项。

数据收集：1985~2003年农村居民人均收入和消费(见表6.3)

表6.3 1985-2003年农村居民人均收入和消费 单位：元

年份	全年人均纯收入（现价）	全年人均消费性支出（现价）	消费价格指数（1985=100）	人均实际纯收入（1985可比价）	人均实际消费性支出（1985可比价）
1985	397.60	317.42	100.0	397.60	317.40
1986	423.80	357.00	106.1	399.43	336.48
1987	462.60	398.30	112.7	410.47	353.42
1988	544.90	476.70	132.4	411.56	360.05
1989	601.50	535.40	157.9	380.94	339.08
1990	686.30	584.63	165.1	415.69	354.11
1991	708.60	619.80	168.9	419.54	366.96
1992	784.00	659.80	176.8	443.44	373.19
1993	921.60	769.70	201.0	458.51	382.94

续　表

年份	全年人均纯收入（现价）	全年人均消费性支出（现价）	消费价格指数（1985=100）	人均实际纯收入（1985可比价）	人均实际消费性支出（1985可比价）
1994	1221.00	1016.81	248.0	492.34	410.00
1995	1577.70	1310.36	291.4	541.42	449.69
1996	1923.10	1572.10	314.4	611.67	500.03
1997	2090.10	1617.15	322.3	648.50	501.77
1998	2162.00	1590.33	319.1	677.53	498.28
1999	2214.30	1577.42	314.3	704.52	501.75
2000	2253.40	1670.00	314.0	717.64	531.85
2001	2366.40	1741.00	316.5	747.68	550.08
2002	2475.60	1834.00	315.2	785.41	581.85
2003	2622.24	1943.30	320.2	818.86	606.81

据表6.3的数据使用普通最小二乘法估计消费模型得:

$$\hat{Y}_t = 106.7528 + 0.5998X_t \qquad (6.44)$$

R^2 = 0.9788, F = 786.0548, df=17, DW = 0.7706, 该回归方程可决系数较高, 回归系数均显著。对样本量为19、一个解释变量的模型、5%显著水平, 查DW统计表可知, dL=1.18, dU= 1.40, 模型中DW<dL, 显然消费模型中有自相关。这一点也可从残差图中看出, 点击EViews方程输出窗口的按钮Resids可得到残差图, 如图6.6所示。

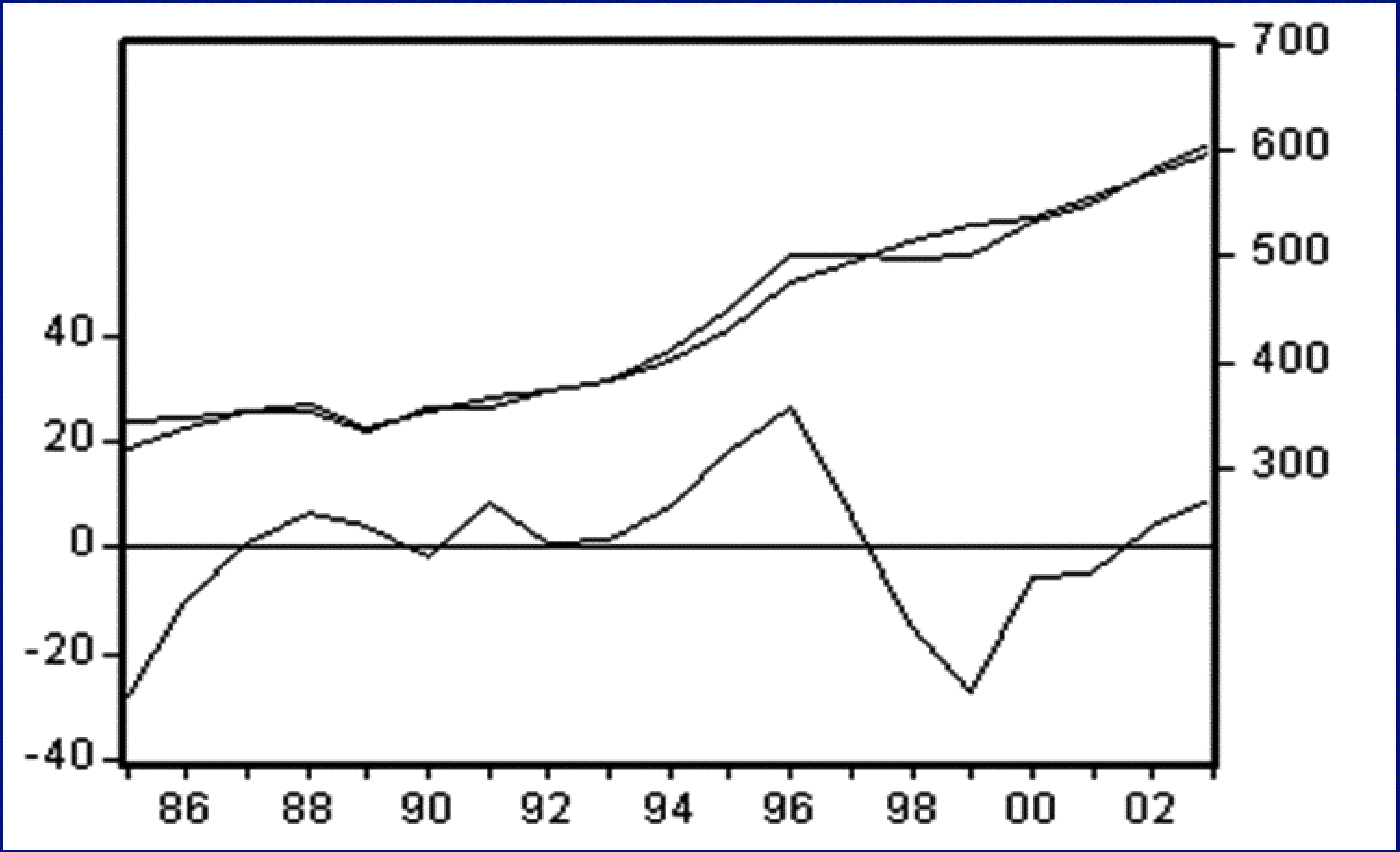


图6.6残差图

自相关问题的处理:

使用科克伦－奥克特的两步法解决自相关问题据模型**(6.44)**可得残差序列***et***，在**EViews**中，每次回归的残差存放在**resid**序列中，为了对残差进行回归分析，需生成命名为***e***的残差序列。在主菜单选择**Quick/Generate Series**或点击工作文件窗口工具栏中的**Procs/Generate Series**，在弹出的对话框中输入***e*** **= resid**，点击**OK**得到残差序列***et***。使用***et***进行滞后一期的自回归，在**EViews**命令栏中输入**ls e e (-1)**可得回归方程:

$$e_t = 0.4960 e_{t-1} \quad (6.45)$$

由式(6.45)可知 $\hat{\rho}$ =0.4960，对原模型进行广义差分，得到广义差分方程：

$$Y_t - 0.4960 Y_{t-1} = \beta_1 (1 - 0.4960) + \beta_2 (X_t - 0.4960 X_{t-1}) + u_t$$

(6.39)

对式（6.39）的广义差分方程进行回归，在EViews命令栏中输入ls *Y*—0.4960Y*(—1) c *X*—0.4960**X*(—1)回车后可得方程输出结果如表6.4。**

表6.4 广义差分方程输出结果

Dependent Variable: Y-0.496014*Y(-1)

Method: Least Squares

Date: 03/26/05 Time: 12:32

Sample(adjusted): 1986 2003

Included observations: 18 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	60.44431	8.964957	6.742287	0.0000
X-0.496014*X(-1)	0.583287	0.029410	19.83325	0.0000

R-squared	0.960914	Mean dependent var	231.9218
Adjusted R-squared	0.958472	S.D. dependent var	49.34525
S.E. of regression	10.05584	Akaike info criterion	7.558623
Sum squared resid	1617.919	Schwarz criterion	7.657554
Log likelihood	-66.02761	F-statistic	393.3577
Durbin-Watson stat	1.397928	Prob(F-statistic)	0.000000

由表6.4可得回归方程为:

$$\hat{Y}_t^* = 60.4443 + 0.5833X_t^* \quad (6.46)$$

$R^2 = 0.9609 \qquad F = 393.3577$

$df = 16 \qquad DW = 1.3979$

式中，$\hat{Y}_t^* = Y_t - 0.4960Y_{t-1}$　　$X_t^* = X_t - 0.4960X_{t-1}$

由于使用了广义差分数据，样本容量减少了1个，为18个。查5%显著水平的DW统计表可知 d_L= 1.16，d_U= 1.39，模型中DW=1.3979>d_U，说明广义差分模型中已无自相关。同时，可决系数R^2、t、F统计量均达到理想水平。

对比模型（**6.44**）和（**6.46**），很明显普通最小二乘法低估了回归系数的标准误。原模型中 $SE(\hat{\beta}_2) = 0.0214$，广义差分模型中为 $SE(\hat{\beta}_2) = 0.0294$。

如果我们使用普莱斯—温斯腾变换补充第一个观测值，即 $X_1\sqrt{1-0.4960^2}$　$Y_1\sqrt{1-0.4960^2}$

得到普莱斯—温斯腾变换的广义差分模型为:

$$Y_t^* = 60.4443 + 0.5833 X_t^*$$

对比模型（**6.47**）和（**6.46**）可发现两者的参数估计值和各检验统计量的差别很微小，说明在本例中使用普莱斯—温斯腾变换与直接使用科克伦—奥克特两步法的估计结果无显著差异，这是因为本例中的样本还不算太小。如果实际应用中样本较小，则两者的差异就会较大。通常对于小样本，应采用普莱斯—温斯腾变换补充第一个观测值。

由差分方程（6.46）可知:

$$\hat{\beta}_1 = \frac{60.4443}{1-0.4960} = 119.9292 \quad (6.48)$$

由此，我们得到最终的中国农村居民消费模型:

$$Y_t = 119.9292 + 0.5833 X_t$$

由模型(6.49)的中国农村居民消费模型可知，中国农村居民的边际消费倾向为0.5833，即中国农民每增加收入1元，将增加消费支出0.5833元。

本章小结

1．当总体回归模型的随机误差项在不同观测点上彼此相关时就产生了自相关问题。

2．自相关的出现有多种原因。时间序列的惯性、模型设定错误、数据的处理等等。

3．在出现自相关时，普通最小二乘估计量依然是无偏、一致的，但不再是有效的。通常的t检验和F检验都不能有效地使用。

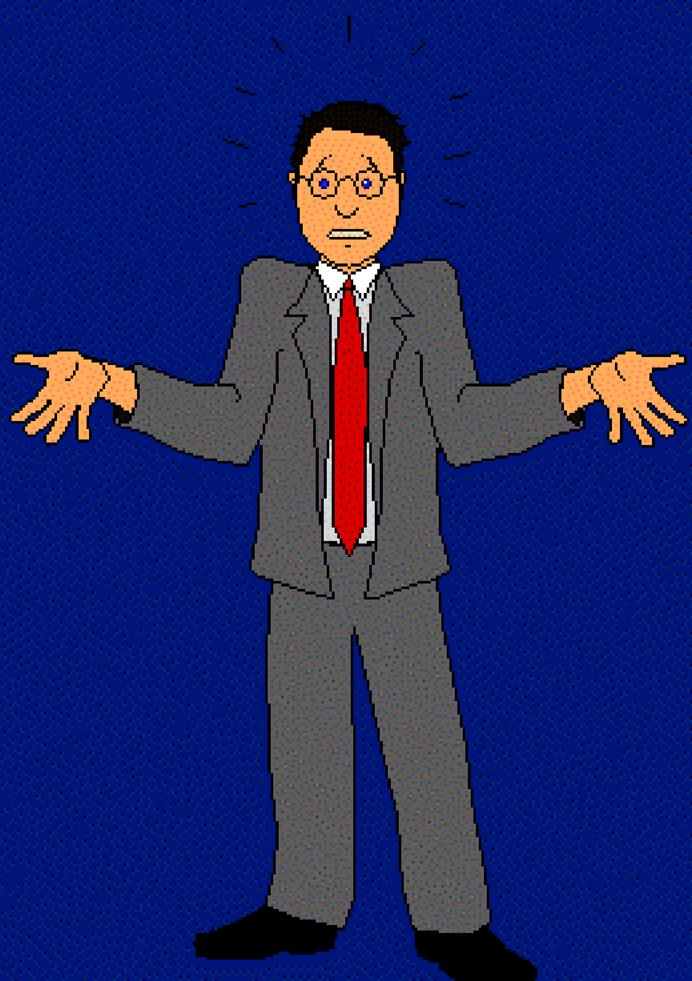
4．为了研究问题的方便和考虑实际问题的代表意义，我们通常将自相关设定为一阶自相关即AR(1)模式。用一阶自相关系数 ρ 表示自相关的程度与方向。当然，实际问题也存在AR(m)模式或其它模式。

5．由于u_t是不可观测的，通常我们使用u_t的估计量e_t判断u_t的特性。我们可通过e_t的图形判断自相关的存在，也可使用依据e_t计算的DW统计量判断自相关的存在。

6．如果自相关系数 ρ 是已知的，我们可以使用广义差分法消除序列相关。

7．如果自相关系数是 ρ 未知的，我们可采用科克伦—奥克特迭代法求得 ρ 的估计值，然后用广义差分法消除序列相关。

第六章 结束了！



THANKS