

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:

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

取我国1985年—2011年农村居民消费 Y 和居民收入 X 的数据为样本数据, 用普通最小二乘法估计其参数, 结果为

$$\hat{Y}_t = 44.1517 + 0.7207 X_t$$

$$(10.1079) \quad (0.0121)$$

$$t = (4.3680) \quad (59.6060)$$

$$R^2 = 0.9930 \quad F = 3552.876 \quad DW = 0.5300$$

检验结果表明：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 统计量为3552.876，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。为什么呢？

第六章 自相关

本章讨论四个问题：

- 什么是自相关
- 自相关的后果
- 自相关的检验
- 自相关性的补救

第一节 什么是自相关

本节基本内容:

- 自相关的概念
- 自相关产生的原因
- 自相关的表现形式

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。可以表示为：

$$Cov(u_i, u_j) = E(u_i, u_j) = 0 \quad (i \neq j)$$

一阶自相关系数

自相关系数 ρ 的定义与普通相关系的公式形式相同

$$\rho = \frac{\sum_{t=2}^n u_t u_{t-1}}{\sqrt{\sum_{t=2}^n u_t^2} \sqrt{\sum_{t=2}^n u_{t-1}^2}} \quad (6.1)$$

ρ 的取值范围为 $-1 \leq \rho \leq 1$

式 (6.1) 中 u_{t-1} 是 u_t 滞后期一期的随机误差项。
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

经济系统的惯性

经济活动的滞后效应

数据处理造成的相关

蛛网现象

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

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：

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。

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


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，即：

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

而建立模型时，模型设定为： $Y_t = \beta_1 + \beta_2 X_{2t} + u_t$

则 X_{3t} 对 Y_t 的影响便归入随机误差项 u_t 中，由于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的符号判断

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。

自相关是 u_1, u_2, \dots, u_n 序列自身的相关, 因随机误差项的关联形式不同而具有不同的自相关形式。

自相关多出现在时间序列数据中。

自相关的形式

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) 的随机项为 u_1, u_2, \dots, u_n ，如果自相关形式为

$$u_t = \rho u_{t-1} + v_t \quad -1 < \rho < 1$$

其中 ρ 为自相关系数， v_t 为经典误差项，即

$$E(v_t) = 0, \text{Var}(v_t) = \sigma^2, \text{Cov}(v_t, v_{t+s}) = 0, s \neq 0$$

则此式称为一阶自回归模式，记为 AR(1)。因为模型中 u_{t-1} 是 u_t 滞后一期的值，因此称为一阶。此式中的 ρ 也称为一阶自相关系数。

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

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v'_t$$

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。此式称为二阶自回归模式，记为 AR(2)。

一般地，如果 u_1, u_2, \dots, u_t 之间的关系为

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_m u_{t-m} + v_t$$

其中， v_t 为经典误差项。则称此式为 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

本节基本内容:

- 一阶自回归形式的性质
- 自相关对参数估计的影响
- 自相关对模型检验的影响
- 自相关对模型预测的影响

一、一阶自回归形式的性质

对于一元线性回归模型：

$$Y = \beta_1 + \beta_2 X + u$$

假定随机误差项 u 存在一阶自相关：

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

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 $\text{Var}(v_t) = \sigma_v^2$ ，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：

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

逐次代入可得：

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

这表明随机误差项 u_t 可表示为独立同分布的随机误差序列 $v_t, v_{t-1}, v_{t-2}, \dots$ 的加权和，权数分别为 $1, \rho, \rho^2, \dots$ 。当 $0 < \rho < 1$ 时，这些权数是随时间推移而呈几何衰减的；

而当 $-1 < \rho < 0$ 时，这些权数是随时间推移而交错振荡衰减的。

可以推得：

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

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

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，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：

$$\text{Cov}(u_t, u_{t-1}) = E(u_t u_{t-1}) = \frac{\rho \sigma_v^2}{1 - \rho^2}$$

$$\text{Cov}(u_t, u_{t-2}) = E(u_t u_{t-2}) = \frac{\rho^2 \sigma_v^2}{1 - \rho^2}$$

以此类推，可得：

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

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

二、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 $\text{Var}(\hat{\beta}_2)$

并且 $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k}$ 将低估真实的 σ^2

对于一元线性回归模型，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：

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

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 $\hat{\beta}_2$ 是否是有效呢？

例如，一元回归中

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

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

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

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 $\rho > 0$ ，同时 X 序列自身也呈正相关，因此前式中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 $\text{Var}(\hat{\beta}_2)$ 。

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

三、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

考虑自相关时的检验

由于 $\text{Var}(\hat{\beta}_2)$ 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 检验统计量为：

$$t = \frac{\text{估计值}}{\text{估计量的标准误}} = \frac{\hat{\beta}_2}{\text{se}(\hat{\beta}_2)}$$

由于 $\text{SE}(\hat{\beta}_2)$ 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。

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 $\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2}$ ，将会导致错误结果。

当 $\rho > 0$ ，即有正相关时，对所有的 j 有 $\rho^j > 0$ 。

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，普通最小二乘法的方差 $\text{Var}(\hat{\beta}_2) = \sigma^2 / \sum x_t^2$ 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，这会夸大估计量的估计精度，即得到较小的标准误差。

因此在有自相关时，普通最小二乘估计 $\hat{\beta}_2$ 的标准误差就不可靠了。

一个被低估了的标准误差意味着一个较大的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统计量不能用来判断回归系数的显著性。

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t检验不再可靠。

四、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 $\hat{\sigma}^2 = \sum e_i^2 / n - k$ 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

本节基本内容:

- 图示检验法
- DW检验法
- Breusch-Godfrey检验 (LM检验)

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。

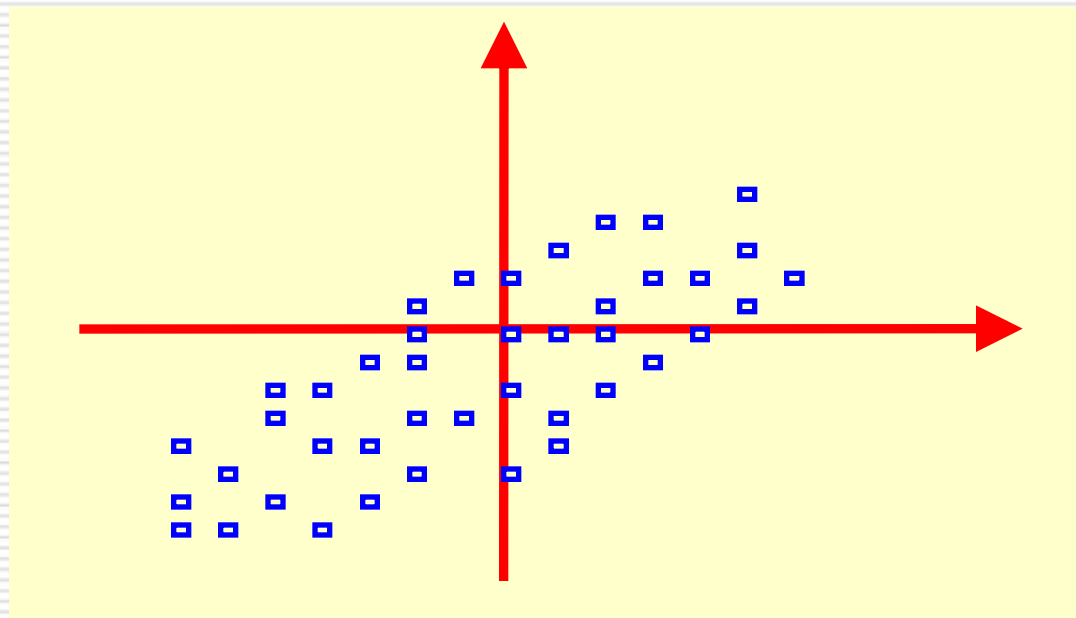


图 6.1 e_t 与 e_{t-1} 的关系

绘制 e_{t-1}, e_t 的散点图。用 (e_{t-1}, e_t) ($t=1, 2, \dots, n$) 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。

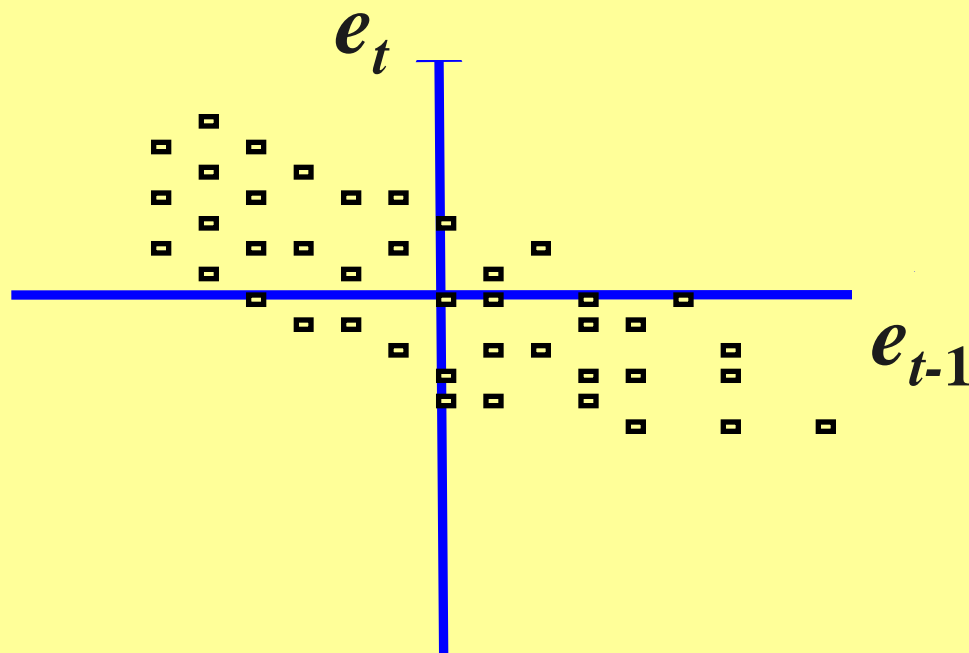
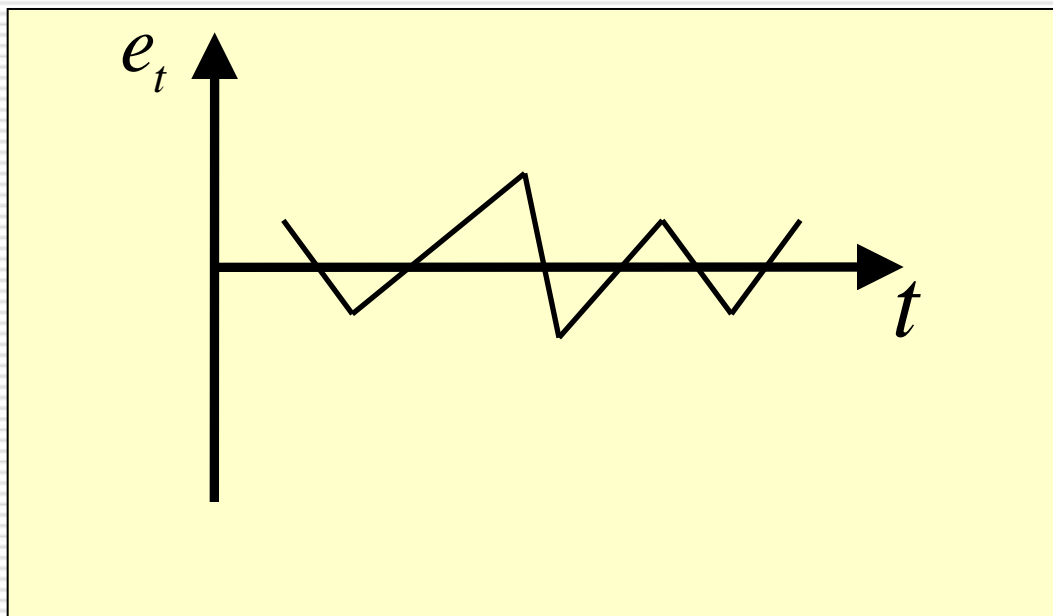


图 6.2 e_t 与 e_{t-1} 的关系

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。



按照时间顺序绘制回归残差项 e_t 的图形。如果 e_t ($t = 1, 2, \dots, n$) 随着 t 的变化逐次有规律地变化, e_t 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

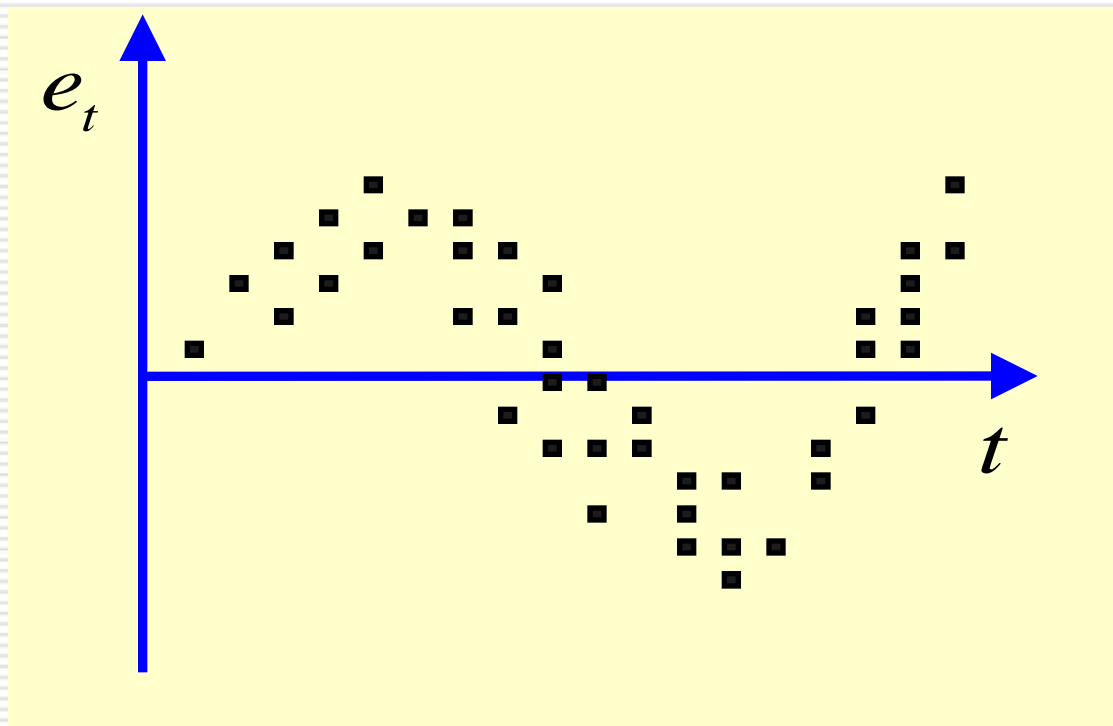


图: e_t 的分布

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 u_t 存在正自相关。

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。

- 1. 解释变量 X 是非随机的
- 2. 随机误差项为已结自回归形式
- 3. 线性模型的解释变量不包含滞后的被解释变量
- 4. 截距项不为零，即只适用于有常数项的回归模型
- 5. 数据序列无缺失项

随机误差项的一阶自回归形式为：

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

为了检验序列的相关性，构造的原假设是：

$$H_0 : \rho = 0$$

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 e_t 定义DW统计量为：

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

$$\begin{aligned}
 DW &= \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} && \left(\text{由 } \sum_{t=2}^n e_t^2 \approx \sum_{t=2}^n e_{t-1}^2 \approx \sum_{t=1}^n e_t^2 \right) \\
 &\approx 2 \left[1 - \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \right] && \left(\text{由 } \hat{\rho} \approx \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \right) \\
 &= 2(1 - \hat{\rho})
 \end{aligned}$$

由 $DW \approx 2(1 - \hat{\rho})$
可得DW 值与 $\hat{\rho}$ 的对应关系如表所示。

$\hat{\rho}$	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：

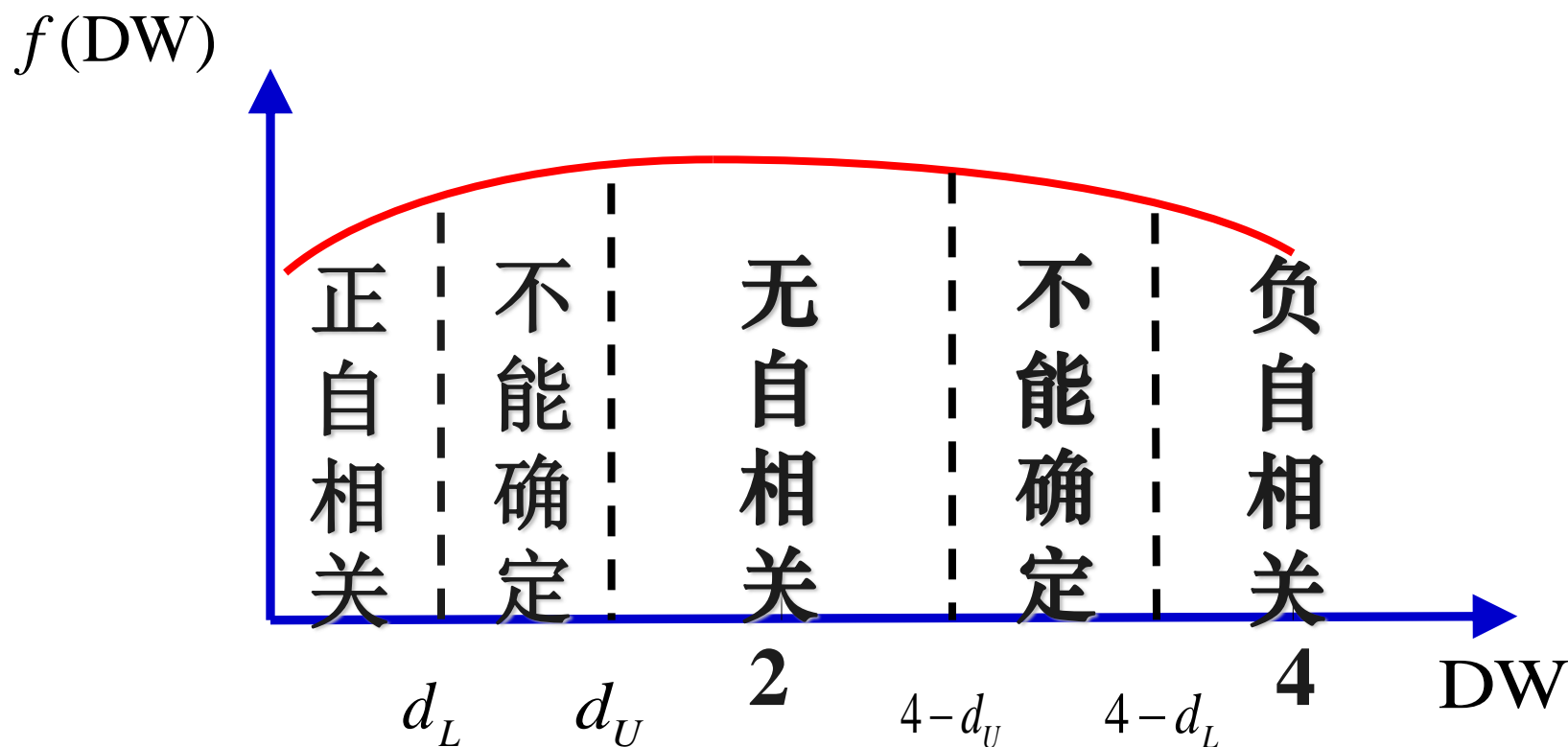
$$0 \leq DW \leq 4$$

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 d_L 和 d_U ，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。

DW检验决策规则

$0 \leq DW \leq d_L$	误差项 u_1, u_2, \dots, u_n 间存在 正相关
$d_L < DW \leq d_U$	不能判定是否有自相关
$d_U < DW < 4 - d_U$	误差项 u_1, u_2, \dots, u_n 间 无自相关
$4 - d_U \leq DW < 4 - d_L$	不能判定是否有自相关
$4 - d_L \leq DW \leq 4$	误差项 u_1, u_2, \dots, u_n 间存在 负相关

用坐标图更直观表示DW检验规则：



DW检验的缺点和局限性

- 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
- DW统计量的上、下界表要求 $n \geq 15$ ，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
- DW检验不适应随机误差项具有高阶序列相关的检验
- 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

三、自相关的BG检验(LM检验)

问题的提出: DW检验应用广泛且方便, 但有一定局限:

1. 只适用于一阶自相关的检验, 高阶自相关怎么检验?
2. 有一些限制条件, 如无滞后被解释变量等.

布劳殊和戈弗雷提出更一般的**BG**检验或称拉格朗日乘数检验
检验的思想:

对于模型:
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + L + \beta_k X_{kt} + u_t$$

若存在高阶自相关:
$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + L + \rho_p u_{t-p} + \varepsilon_t$$

其中 ε_t 满足基本假定 (白噪声)

即
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + L + \beta_k X_{kt} + \rho_1 u_{t-1} + \rho_2 u_{t-2} + L + \rho_p u_{t-p} + \varepsilon_t$$

可通过检验 $H_0: \rho_1 = \rho_2 = L = \rho_p = 0$ 检验是否存在高阶自相关性
实际是检验模型中是否遗漏了 $u_{t-1}, u_{t-2}, L, u_{t-p}$

检验思想(LM检验)

基于所分析模型普通最小二乘估计的残差对解释变量和一定数量滞后残差的辅助回归，如果滞后残差足以解释当前残差的变异，就拒绝误差项无自相关的原假设。

检验方法：拉格朗日乘数 (LM) 检验

1.用OLS估计模型： $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + L + \beta_k X_{ki} + u_i$

并得残差序列 e_t （用于替代 u_t ）

2.作辅助回归：将 e_t 对模型中所有解释变量和残差的滞后值 e_{t-1}, L, e_{t-p} 进行辅助回归，即

$$e_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + L + \beta_k X_{kt} + \rho_1 e_{t-1} + \rho_2 e_{t-2} + L + \rho_p e_{t-p} + v_t$$

3.计算辅助回归的可决系数 R^2 ， 构建统计量 TR^2

可证明，在大样本情况下，渐近地有 $TR^2 : \chi^2(p)$

（P为约束个数或自相关阶数； T为辅助回归的样本容量）

5.检验：给定显著性水平 α ， 查自由度为 p 的临界值 $\chi_\alpha^2(p)$
若 TR^2 大于 $\chi_\alpha^2(p)$ ， 则拒绝 $H_0 : \rho_1 = \rho_2 = L = \rho_p = 0$ 表明可能存在直到 p 阶的自相关

若 TR^2 小于 $\chi_\alpha^2(p)$ ， 则表明不存在自相关。

EViews操作举例:

View /Residual Test /Serial Correlation LM Test

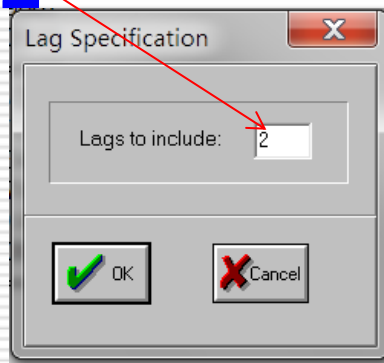
例如: 对于

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

Dependent Variable: Y
Method: Least Squares
Date: 09/09/13 Time: 16:56
Sample: 1985 2011
Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	44.15166	10.10791	4.368029	0.0002
X	0.720659	0.012090	59.60601	0.0000

R-squared	0.993013	Mean dependent var	586.8085
Adjusted R-squared	0.992733	S.D. dependent var	267.6878
S.E. of regression	22.81932	Akaike info criterion	9.164279
Sum squared resid	13018.03	Schwarz criterion	9.260267
Log likelihood	-121.7178	Hannan-Quinn criter.	9.192821
F-statistic	3552.876	Durbin-Watson stat	0.529988
Prob(F-statistic)	0.000000		

例如取阶数 $p=2$ BG检验结果: $TR^2 = 27 \times 0.53241 = 14.37507$

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	13.09420	Prob. F(2,23)	0.0002
Obs*R-squared	14.37507	Prob. Chi-Square(2)	0.0008

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 09/09/13 Time: 17:01

Sample: 1985 2011

Included observations: 27

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.177653	7.261598	0.024465	0.9807
X	-9.60E-05	0.008722	-0.011008	0.9913
RESID(-1)	0.778856	0.212339	3.667978	0.0013
RESID(-2)	-0.068992	0.215440	-0.320237	0.7517

R-squared	0.532410	Mean dependent var	1.04E-13
Adjusted R-squared	0.471420	S.D. dependent var	22.37618
S.E. of regression	16.26827	Akaike info criterion	8.552264
Sum squared resid	6087.101	Schwarz criterion	8.744240
Log likelihood	-111.4556	Hannan-Quinn criter.	8.609348
F-statistic	8.729466	Durbin-Watson stat	1.955487
Prob(F-statistic)	0.000476		

p值

BG检验（拉格朗日乘数检验）的适用范围

1. **DW**检验只能用于一阶自相关的检验，而**BG**检验（拉格朗日乘数检验）适合于高阶自相关的检验。
2. 适合检验模型的解释变量中有滞后被解释变量 Y_{t-1} 的情况。
3. **BG**检验的缺陷是滞后长度 p 不能先验确定。实际检验中可从**1**阶、**2**阶、**L** 逐次向更高阶检验，并结合用辅助回归中各 e_t 滞后项参数的显著性去帮助判断自相关的阶数。

利用**Evews**可以直接进行**BG**检验：方法是在回归方程窗口点 **View/Residual Test/Serial Correlation LM Test**，在对话框中选定滞后期数 p ，点“**ok**”得到计算的 TR^2 及对应的 $\chi^2(p)$ 检验的 p 值

若 $\alpha > p$ 则拒绝 $H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$ ，表明存在自相关

若 $\alpha < p$ 则不拒绝 H_0 ，表明不存在自相关

第四节 自相关的补救

本节基本内容:

- 广义差分法
- 科克伦—奥克特迭代法
- 其他方法简介

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

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

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

将模型滞后一期可得

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

用 ρ 乘式两边，得

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

两式相减, 可得

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

式中, $u_t - \rho u_{t-1} = v_t$ 是经典误差项。因此, 模型已经是经典线性回归。令:

$$Y_t^* = Y_t - \rho Y_{t-1}, \quad X_t^* = X_t - \rho X_{t-1}, \quad \beta_1^* = \beta_1(1 - \rho)$$

则上式可以表示为:

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + v_t$$

对模型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。

这称为广义差分方程，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。

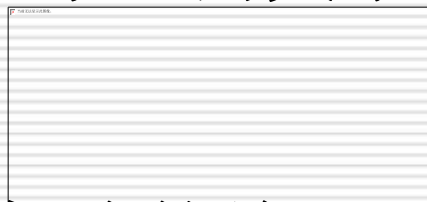
在进行广义差分时，解释变量 X 与被解释变量 Y 均以差分形式出现，因而样本容量由 n 减少为 $n-1$ ，即丢失了第一个观测值。如果样本容量较大，减少一个观测值对估计结果影响不大。但是，如果样本容量较小，则对估计精度产生较大的影响。此时，可采用普莱斯-温斯滕（Prais-Winsten）变换，将第一个观测值变换为：

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

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:



但是,这 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 $\hat{\rho}$, 通常采用科克伦—奥克特 (Cochrane – Orcutt) 迭代法。

该方法利用残差 去估计未知的 ρ 。对于一元线性回归模型

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

假定 u_t 为一阶自回归形式，即：

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

科克伦—奥克特迭代法估计 ρ 的步骤如下：

1. 使用普通最小二乘法估计模型

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

并获得残差： $e_t^{(1)}$

2. 利用残差 $e_t^{(1)}$ 做如下的回归

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

3. 利用 $\hat{\rho}^{(1)}$ ，对模型进行广义差分，即

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

$$\text{令 } Y_t^* = Y_t - \hat{\rho}^{(1)} Y_{t-1} \quad X_t^* = X_t - \hat{\rho}^{(1)} X_{t-1}$$

$$\beta_1^* = \beta_1 (1 - \hat{\rho}^{(1)})$$

使用普通最小二乘法，可得样本回归函数为：

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

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

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

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

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

5. 利用残差 $e_t^{(3)}$ 做如下的回归

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

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，还要继续估计 ρ 的第三轮估计值 $\hat{\rho}^{(3)}$ 。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。

三、其它方法简介

(一) 一阶差分法

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

式中， u_t 为一阶自回归AR(1)。将模型变换为：

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

如果原模型存在完全一阶正自相关，即 $\rho=1$

则 $u_t = u_{t-1} + v_t$

其中， v_t 为经典误差项。则随机误差项为经典误差项，无自相关问题。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：

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

第一步，把上式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： $Y_t^* = Y_t - \hat{\rho}Y_{t-1}$ 和 $X_t^* = X_t - \hat{\rho}X_{t-1}$

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

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—2011年农村居民人均收入和消费
(数据见教材)

模型的建立、估计与检验

使用普通最小二乘法估计消费模型得：

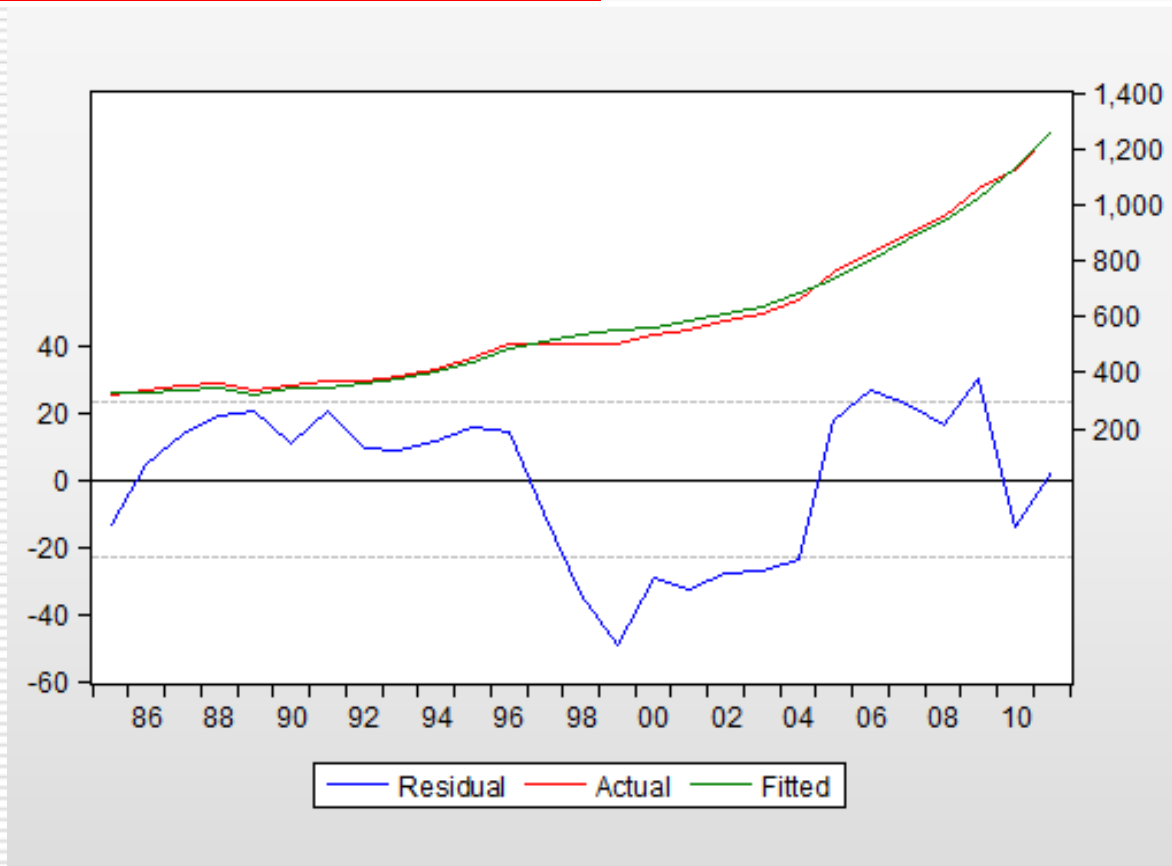
$$\hat{Y}_t = 44.1517 + 0.7207X_t$$

(10.1079) (0.0121)

$$t = (4.3680) \quad (59.6060)$$

$$R^2 = 0.9930 \quad F = 3552.876 \quad DW = 0.5300$$

该回归方程可决系数较高，回归系数均显著。对样本量为27、一个解释变量的模型、5%显著水平，查DW统计表可知， $dL=1.316$ ， $dU=1.469$ ，模型中 $DW < d_L$ ，显然消费模型中有正自相关。这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。



残差图

用BG检验作自相关检验

EViews中回归后点“View/Residual Diagnostics /Serial Correlation LM Test”，在“lags to include”中选取滞后阶数例如“2”，回车即的检验结果：

$$LM = TR^2 = 27 \times 0.53241 = 14.3751$$

其p值为0.000756，表明存在自相关。

结论：由于本案例存在明显的自相关，直接回归估计的结论是不可靠的。

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	13.09420	Prob. F(2,23)	0.0002
Obs*R-squared	14.37507	Prob. Chi-Square(2)	0.0008

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 09/08/13 Time: 20:17

Sample: 1985 2011

Included observations: 27

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.177653	7.261598	0.024465	0.9807
X	-9.60E-05	0.008722	-0.011008	0.9913
RESID(-1)	0.778856	0.212339	3.667978	0.0013
RESID(-2)	-0.068992	0.215440	-0.320237	0.7517

R-squared	0.532410	Mean dependent var	1.04E-13
Adjusted R-squared	0.471420	S.D. dependent var	22.37618
S.E. of regression	16.26827	Akaike info criterion	8.552264
Sum squared resid	6087.101	Schwarz criterion	8.744240
Log likelihood	-111.4556	Hannan-Quinn criter.	8.609348
F-statistic	8.729466	Durbin-Watson stat	1.955487
Prob(F-statistic)	0.000476		

自相关问题的处理

用回归残差序列 e_t 估计自相关系数

在EViews命令栏中输入“ls e e (-1)”可得回归方程

$$\hat{e}_t = 0.7283e_{t-1}$$

用 $\hat{\rho}=0.7283$ 对原模型进行广义差分回归，在EViews命令栏中直接输入“ls Y-0.7283*Y (-1) c X-0.7283*X (-1)” /回车，即得广义差分回归的输出结果

由广义差分方程有

$$\hat{\beta}_1 = \frac{13.6640}{1-0.7283} = 50.2908$$

得到中国农村居民消费模型为

$$\hat{Y}_t = 50.2908 + 0.7162X_t$$

Dependent Variable: Y-0.7283*Y(-1)
Method: Least Squares
Date: 09/08/13 Time: 20:21
Sample (adjusted): 1986 2011
Included observations: 26 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13.66397	6.492744	2.104498	0.0460
X-0.7283*X(-1)	0.716223	0.023371	30.64567	0.0000
R-squared	0.975082	Mean dependent var	188.8166	
Adjusted R-squared	0.974044	S.D. dependent var	97.49753	
S.E. of regression	15.70779	Akaike info criterion	8.419994	
Sum squared resid	5921.631	Schwarz criterion	8.516771	
Log likelihood	-107.4599	Hannan-Quinn criter.	8.447862	
F-statistic	939.1570	Durbin-Watson stat	1.779430	
Prob(F-statistic)	0.000000			

科克伦—奥克特迭代法

在Eviews的命令栏输入“LS Y C X AR (1)”/回车，即自动迭代得科克伦—奥克特迭代法估计结果：

由 $DW=1.7813$ 可以判断， $dU = 1.461$ ， $dU < DW < 4-dU$ ，说明在5%显著性水平下广义差分后模型中已无自相关。

结论：中国农村居民的边际消费倾向为0.7162，农民人均实际纯收入每增加1元，平均说来人均实际消费支出将增加0.7162元。

Dependent Variable: Y
Method: Least Squares
Date: 09/08/13 Time: 20:23
Sample (adjusted): 1986 2011
Included observations: 26 after adjustments
Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	50.31957	24.86137	2.024007	0.0547
X	0.716205	0.024113	29.70176	0.0000
AR(1)	0.729402	0.141543	5.153200	0.0000

R-squared	0.996688	Mean dependent var	597.1696
Adjusted R-squared	0.996400	S.D. dependent var	267.4109
S.E. of regression	16.04561	Akaike info criterion	8.496914
Sum squared resid	5921.616	Schwarz criterion	8.642079
Log likelihood	-107.4599	Hannan-Quinn criter.	8.538717
F-statistic	3460.310	Durbin-Watson stat	1.781333
Prob(F-statistic)	0.000000		

Inverted AR Roots	.73
-------------------	-----

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。可通过 e_t 的图形判断自相关的存在，也可使用DW 统计量或者用BG 检验判断自相关的存在。

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