

第 8 章 自 相 关

8.1 自相关的后果

除了异方差外，违反球形扰动项的另一情形是扰动项存在自相关。

对于扰动项 $\{\varepsilon_1, \dots, \varepsilon_n\}$ ，如果存在 $i \neq j$ ，使得 $E(\varepsilon_i \varepsilon_j | \mathbf{X}) \neq 0$ ，即协方差矩阵 $\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X})$ 的非主对角线元素不全为 0，则存在自相关(autocorrelation)或序列相关(serial correlation)。

在有自相关的情况下：

(1) OLS 估计量依然是无偏、一致且渐近正态的，因为在证明这些性质时，并未用到“无自相关”的假定。

(2) OLS 估计量方差 $\text{Var}(\hat{\boldsymbol{\beta}} | \mathbf{X})$ 的表达式不再是 $\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ ，因为 $\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X}) \neq \sigma^2 \mathbf{I}$ 。因此，使用普通标准误的 t 检验、 F 检验失效。

(3) 高斯-马尔可夫定理不再成立，即 OLS 不再是 BLUE。

为了直观地理解为何在自相关的情况下，OLS 不再是 BLUE，假设扰动项存在正自相关，即 $E(\varepsilon_i \varepsilon_j | \mathbf{X}) > 0$ ，参见图 8.1。

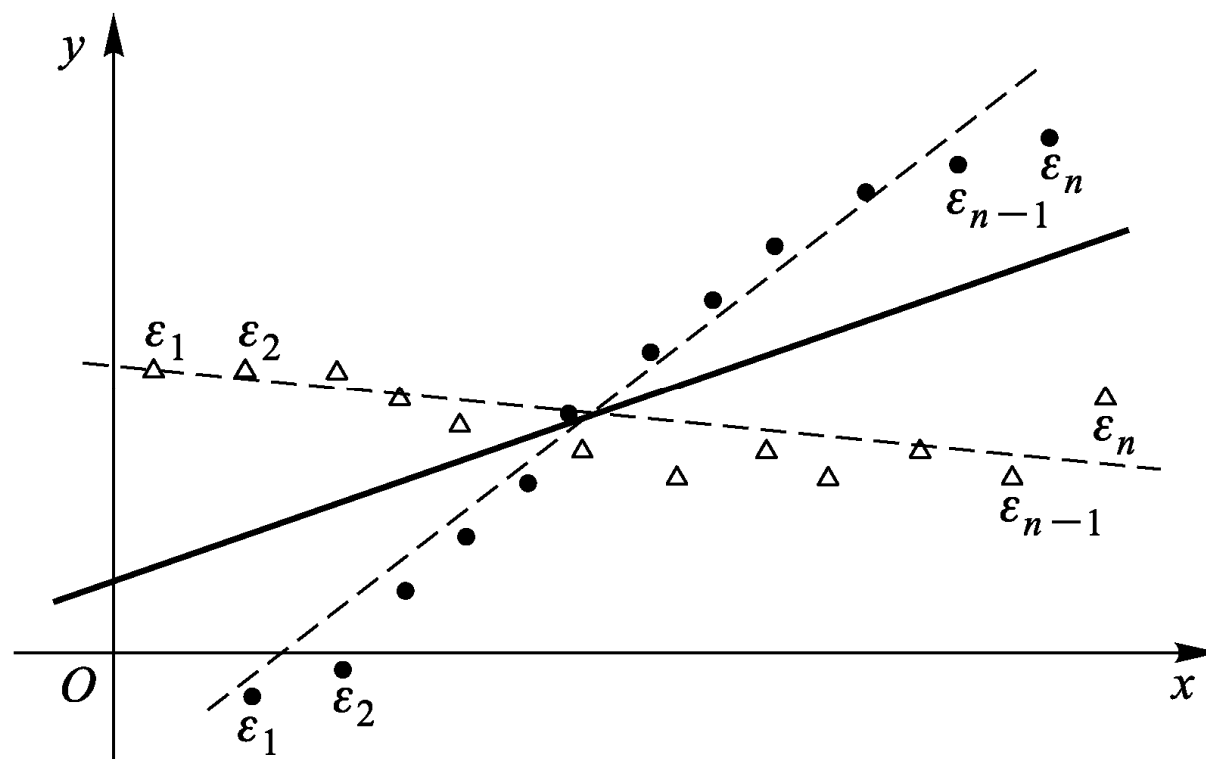


图 8.1 自相关的后果

从信息的角度来看，由于 OLS 估计忽略了扰动项自相关所包含的信息，故不是最有效率的估计方法。

8.2 自相关的例子

(1) 时间序列数据中的自相关:

由于经济活动通常具有某种连续性或持久性，自相关现象在时间序列中比较常见。

比如，相邻两年的 GDP 增长率、通货膨胀率。

又比如，某意外事件或新政策的效应需要随时间逐步释放出来。

再比如，最优资本存量需要通过若干年的投资才能逐渐达到(滞后的调整过程)。

(2) 横截面数据中的自相关:

一般来说，截面数据不容易出现自相关，但相邻的观测单位之间也可能存在**溢出效应**(spillover effect 或 neighborhood effect)，这种自相关也称为**空间自相关**(spatial autocorrelation)。

比如，相邻的省份、国家之间的经济活动相互影响(通过贸易、投资、劳动力流动等)；

相邻地区的农业产量受到类似天气变化的影响；

同一社区内的房屋价格存在相关性。

(3) 对数据的人为处理:

如果数据中包含移动平均数(moving average)、内插值(参见第 9 章)或季节调整(参见第 13 章)时, 则可从理论上判断存在自相关。

统计局提供的某些数据可能已经事先经过了这些人为处理。

(4) 设定误差(misspecification):

如果模型设定中遗漏了某个自相关的解释变量, 并被纳入到扰动项中, 则会引起扰动项的自相关。

这种由于设定误差而导致的自相关, 即便在横截面数据中也可能存在。

8.3 自相关的检验

1. 画图

由于残差 $\{e_t\}_{t=1}^n$ 可大致视为扰动项的实现值 $\{\varepsilon_t\}_{t=1}^n$ ，故可通过残差来考察扰动项的自相关。

一个直观的方法是将残差 e_t 与残差滞后 e_{t-1} 画成散点图。

进一步，可以计算残差的各阶样本相关系数，比如残差的一阶相关系数 $\hat{\rho}_1$ ，二阶相关系数 $\hat{\rho}_2$ ，……， k 阶相关系数 $\hat{\rho}_k$ ，等等。

由于这些相关系数 $\hat{\rho}_k$ 是滞后阶数 k 的函数，将 $(k, \hat{\rho}_k)$ 画图，即可得到残差的自相关图(correlogram)，参见图 8.6。

2. BG 检验(Breusch, 1978; Godfrey, 1978)

考虑以下多元线性模型：

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t \quad (8.1)$$

假设扰动项 ε_t 存在一阶自相关，即

$$\varepsilon_t = \gamma \varepsilon_{t-1} + u_t \quad (8.2)$$

其中， u_t 为白噪声。

方程(8.2)没有常数项，因为 $E(\varepsilon_t) = 0$ 。

为了检验是否存在一阶自相关，只要在方程(8.2)中检验 $H_0: \gamma = 0$ 即可。

由于可能存在高阶自相关，可考虑扰动项的 p 阶自回归：

$$\varepsilon_t = \gamma_1 \varepsilon_{t-1} + \cdots + \gamma_p \varepsilon_{t-p} + u_t \quad (8.3)$$

并检验原假设 “ $H_0: \gamma_1 = \cdots = \gamma_p = 0$ ”。

由于 ε_t 不可观测，故用 e_t 替代，并引入解释变量 (x_{t2}, \cdots, x_{tK}) ，进行如下辅助回归：

$$e_t = \gamma_1 e_{t-1} + \cdots + \gamma_p e_{t-p} + \delta_2 x_{t2} + \cdots + \delta_K x_{tK} + v_t \quad (t = p+1, \cdots, n)$$

(8.4)

由于残差 e_t 是解释变量 (x_{t2}, \cdots, x_{tK}) 的函数，故如果遗漏 (x_{t2}, \cdots, x_{tK}) ，可能导致扰动项 v_t 与 $(e_{t-1}, \cdots, e_{t-p})$ 相关，使得估计不一致。

在辅助回归(8.4)中，“无自相关”的原假设相当于检验 $H_0: \gamma_1 = \cdots = \gamma_p = 0$ ，通常使用 nR^2 形式的 LM 统计量进行检验：

$$LM = (n-p)R^2 \xrightarrow{d} \chi^2(p) \quad (8.5)$$

由于辅助回归(8.4)使用了 e_{t-p} ，损失 p 个样本观测值，故样本容量仅为 $(n-p)$ 。

如果 LM 统计量超过了 $\chi^2(p)$ 的临界值，则拒绝无自相关的原假设。此检验称为 Breusch-Godfrey 检验(Breusch, 1978; Godfrey, 1978, 简记 BG)。

Davidson and MacKinnon(1993)建议，把残差中因滞后而缺失的项用其期望值 0 来代替，以保持样本容量仍为 n ，然后使用 LM 统计量 $nR^2 \xrightarrow{d} \chi^2(p)$ 。

Davidson-MacKinnon 方法为 Stata 的默认设置。

3. Q 检验

记 ρ_1, \dots, ρ_p 分别为扰动项的 1 至 p 阶自相关系数。

检验自相关的另一思路是，检验各阶自相关系数均为 0，即 $H_0 : \rho_1 = \dots = \rho_p = 0$ 。

定义残差的各阶样本自相关系数为

$$\hat{\rho}_j \equiv \frac{\sum_{t=j+1}^n e_t e_{t-j}}{\sum_{t=1}^n e_t^2} \quad (j=1, \dots, p) \quad (8.6)$$

如果 $H_0 : \rho_1 = \dots = \rho_p = 0$ 成立，则 $\hat{\rho}_j$ 应离 0 不远。

根据大数定律, $\hat{\rho}_j$ 依概率收敛至 0。

根据中心极限定理, $\sqrt{n}\hat{\rho}_j$ 服从渐近正态分布。

因此, $\sqrt{n}\hat{\rho}_j$ 的平方和(对 j 求和)为渐近卡方分布, 这就是 Box-Pierce Q 统计量(Box and Pierce, 1970):

$$Q_{\text{BP}} \equiv n \sum_{j=1}^p \hat{\rho}_j^2 \xrightarrow{d} \chi^2(p) \quad (8.7)$$

经过改进的 Ljung-Box Q 统计量(Ljung and Box, 1979)为

$$Q_{\text{LB}} \equiv n(n+2) \sum_{j=1}^p \frac{\hat{\rho}_j^2}{n-j} \xrightarrow{d} \chi^2(p) \quad (8.8)$$

这两种 Q 统计量在大样本下是等价的，但 Ljung-Box Q 统计量的小样本性质更好，故为 Stata 所采用。

如何确定自相关阶数 p 呢？

没有确定的规则。

如果 p 太小，则可能忽略了高阶自相关的存在。

如果 p 较大(与样本容量 n 相比)，则 Q 统计量的小样本分布可能与 $\chi^2(p)$ 相差较远。

Stata 默认的 p 值为 $\min\{\text{floor}(n/2) - 2, 40\}$ ，其中 $\text{floor}(n/2)$ 为不超过 $n/2$ 的最大整数，并在 $[\text{floor}(n/2) - 2]$ 与 40 之间取其小者。

4. DW 检验

DW 检验(Durbin and Watson, 1950)是较早出现的自相关检验, 现已不常用。

它的主要缺点是只能检验一阶自相关, 且必须在解释变量满足严格外生性的情况下才成立(BG 检验无此限制)。

DW 检验的统计量为

$$\begin{aligned} DW \equiv d &\equiv \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{\sum_{t=2}^n e_t^2 - 2\sum_{t=2}^n e_t e_{t-1} + \sum_{t=2}^n e_{t-1}^2}{\sum_{t=1}^n e_t^2} \\ &\approx 2 - 2\frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \equiv 2(1 - \hat{\rho}_1) \end{aligned}$$

(8.9)

其中, $\hat{\rho}_1$ 为残差的一阶自相关系数。

当 $d = 2$ 时, $\hat{\rho}_1 \approx 0$, 无一阶自相关。

当 $d = 0$ 时, $\hat{\rho}_1 \approx 1$, 存在一阶正自相关。

当 $d = 4$ 时, $\hat{\rho}_1 \approx -1$, 存在一阶负自相关。

DW 检验的另一缺点是, 其 d 统计量还依赖于数据矩阵 \mathbf{X} , 无法制成统计表, 须使用其上限分布 d_U 与下限分布 d_L ($d_L < d < d_U$) 来间接地检验。

即便如此, 仍存在“无结论区域”。

DW 检验的具体检验方法，根据 d_U 与 d_L 的临界值，可做如下判断(参见图 8.2):

- (1) 如果 $0 < d \leq d_L$ ，则存在正自相关；
- (2) 如果 $d_L < d < d_U$ ，则无法确定；
- (3) 如果 $d_U \leq d \leq 4 - d_U$ ，则无自相关；
- (4) 如果 $4 - d_U < d < 4 - d_L$ ，则无法确定；
- (5) 如果 $4 - d_L \leq d$ ，则存在负自相关。

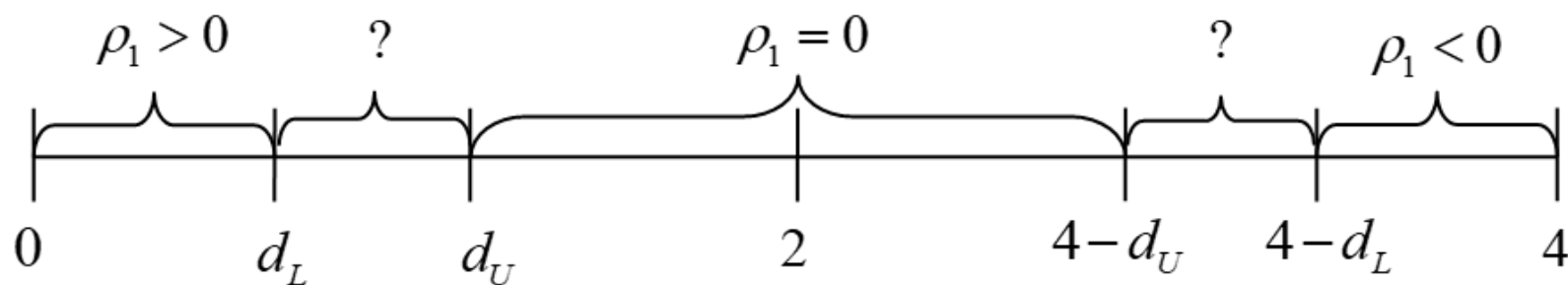


图 8.2 DW 检验的无结论区域

8.4 自相关的处理

如果经过检验，发现存在自相关，则大致有以下四种处理方法。

1. 使用“OLS + 异方差自相关稳健的标准误”

在自相关的情况下，OLS 估计量依然无偏且一致，故仍可使用 OLS 来估计回归系数。

为了正确地进行统计推断，须使用异方差自相关稳健的标准误 (Heteroskedasticity and Autocorrelation Consistent Standard Error, 简记 HAC)，即在存在异方差与自相关的情况下也成立的稳健标准误。

这种方法称为 **Newey-West 估计法**(Newey and West, 1987)。

它只改变标准误的估计值，并不改变回归系数的估计值。

根据第 6 章，异方差稳健的协方差矩阵为夹心估计量：

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}} | \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \widehat{\text{Var}}(\boldsymbol{\varepsilon} | \mathbf{X}) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \quad (8.10)$$

异方差自相关稳健的协方差矩阵也是夹心估计量，但考虑到自相关的存在，“三明治”中间的“菜” $\widehat{\text{Var}}(\boldsymbol{\varepsilon} | \mathbf{X})$ 更为复杂。

在计算 HAC 标准误时，如果仅考虑前几阶自相关系数(比如只考虑一阶自相关系数 ρ_1)将导致此标准误不一致，因为忽略了高阶自相关。

如果同时考虑所有各阶相关系数，即 $(\rho_1, \dots, \rho_{n-1})$ ，则待估参数多达 $(n-1)$ ，且随样本容量 n 同步增长，也将导致估计量不一致。

另外，对 ρ_{n-1} 的估计将很不准确，因为只有一对数据 (e_1, e_n) 可用于此估计；类似地，对 ρ_{n-2} 的估计也不准确，以此类推。

正确的做法是, 包括足够多阶数的自相关系数, 并让此阶数 p 随着样本容量 n 而增长。

一般建议取 $p = n^{1/4}$ 或 $p = 0.75 n^{1/3}$, 称为**截断参数**(truncation parameter), 即比 p 更高阶的自相关系数将被截断而不考虑。

由于 HAC 标准误取决于截断参数 p , 故实践中, 建议使用不同的截断参数, 以考察 HAC 标准误是否对于截断参数敏感。

2. 准差分法

在自相关的情况下, 由于 OLS 未充分利用此信息, 故不是最有效率的 BLUE。

根据加权最小二乘法的思路, 如果能够变换原模型, 使得转换后的扰动项变为球形扰动项(不再有自相关), 则可得到更有效率的估计。

假设原模型为

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t \quad (t = 1, \cdots, n) \quad (8.11)$$

其中, 扰动项 ε_t 存在自相关, 且为一阶自回归形式:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad (8.12)$$

其中, 自回归系数 $|\rho| < 1$, 且 u_t 为白噪声。

将原模型(8.11)滞后一期，然后在方程两边同时乘以 ρ 可得

$$\rho y_{t-1} = \rho\beta_1 + \rho\beta_2 x_{t-1,2} + \cdots + \rho\beta_K x_{t-1,K} + \rho\varepsilon_{t-1} \quad (8.13)$$

将原方程(8.11)减去方程(8.13)可得

$$y_t - \rho y_{t-1} = (1-\rho)\beta_1 + \beta_2(x_{t2} - \rho x_{t-1,2}) + \cdots + \beta_K(x_{tK} - \rho x_{t-1,K}) + \underbrace{(\varepsilon_t - \rho\varepsilon_{t-1})}_{u_t} \quad (8.14)$$

其中， $t = 2, \cdots, n$ ，故损失一个样本观测值。

在方程(8.14)中，新扰动项 $(\varepsilon_t - \rho\varepsilon_{t-1}) = u_t$ (白噪声)，故满足球型扰动项的假定。

对方程(8.14)进行 OLS 估计，可提高估计效率。

这种方法称为 Cochrane-Orcutt 估计法(Cochrane and Orcutt, 1949, 简记 CO)。

此法也称为准差分法(quasi differences)，因为在做变换时，只是减去滞后值的一部分(比如 $y_t - \rho y_{t-1}$)，而非全部(比如 $y_t - y_{t-1}$)。

由于使用准差分法将损失一个样本容量，故仍不是最有效率的 BLUE。

为了得到 BLUE 估计量，考虑补上损失的第一个方程：

$$y_1 = \beta_1 + \beta_2 x_{12} + \cdots + \beta_K x_{1K} + \varepsilon_1 \quad (8.15)$$

由于 $\{u_t\}_{t=1}^n$ 为白噪声，故 ε_1 与准差分后的新扰动项 $u_t = (\varepsilon_t - \rho\varepsilon_{t-1})$ 均不相关。

加入第一个方程(8.15)不会导致自相关，但会导致异方差。

第一个方程(8.15)的扰动项方差为 $\sigma_\varepsilon^2 \equiv \text{Var}(\varepsilon_t)$ ，而准差分方程(8.14)的扰动项方差为 $\sigma_u^2 \equiv \text{Var}(u_t)$ ，二者并不相等。

对方程(8.12)两边求方差，可得 σ_ε^2 与 σ_u^2 之间的关系：

$$\text{Var}(\varepsilon_t) = \rho^2 \text{Var}(\varepsilon_{t-1}) + \text{Var}(u_t) \quad (8.16)$$

将上式移项整理可得

$$\sigma_u^2 = (1 - \rho^2) \sigma_\varepsilon^2 \quad (8.17)$$

由此可知， σ_u^2 是 σ_ε^2 的 $(1-\rho^2)$ 倍。

除非 $\rho = 0$ (无自相关)，否则二者不会相等。

只要将第一个方程(8.15)两边同乘 $\sqrt{1-\rho^2}$ ，即可保证同方差：

$$\sqrt{1-\rho^2} y_1 = \sqrt{1-\rho^2} \beta_1 + \beta_2 \sqrt{1-\rho^2} x_{12} + \cdots + \beta_K \sqrt{1-\rho^2} x_{1K} + \sqrt{1-\rho^2} \varepsilon_1 \quad (8.18)$$

方程(8.18)的扰动项方差为

$$\text{Var}\left(\sqrt{1-\rho^2} \varepsilon_1\right) = (1-\rho^2) \sigma_\varepsilon^2 = \sigma_u^2 \quad (8.19)$$

故这 n 个方程满足同方差与无自相关的假定，为球形扰动项。

使用 OLS 估计这些方程，即可得到 BLUE。

这种方法称为 Prais-Winsten 估计法(Prais and Winsten, 1954, 简记 PW)。

无论 CO 法还是 PW 法均不可行(infeasible)，因为它们都假设知道一阶自回归系数 ρ 。

实践中，必须用数据估计一阶自回归系数 $\hat{\rho}$ 。

Stata 所默认的估计方法为使用 OLS 残差进行辅助回归：

$$e_t = \hat{\rho}e_{t-1} + error_t \quad (8.20)$$

另外，也可使用残差的一阶自相关系数来估计 $\hat{\rho}$ ：

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \quad (8.21)$$

或通过 DW 统计量来估计 $\hat{\rho}$ ：

$$\hat{\rho} = 1 - \frac{DW}{2} \quad (8.22)$$

实践中，常使用迭代法，即首先用 OLS 估计原模型，使用 OLS 残差作辅助回归(8.20)，得到 $\hat{\rho}^{(1)}$ (对 ρ 的第一轮估计)，再用 $\hat{\rho}^{(1)}$ 进行 CO 或 PW 估计；然后，使用 CO 或 PW 法的新残差估计 $\hat{\rho}^{(2)}$ (对 ρ 的第二轮估计)，再用 $\hat{\rho}^{(2)}$ 进行 CO 或 PW 估计，以此类推，直至收敛(即相邻两轮的 ρ 与系数估计值之差足够小)。

3. 广义最小二乘法(GLS)

更一般地，可能同时存在异方差与自相关。

可使用广义最小二乘法(Generalized Least Squares, 简记 GLS), 同时处理异方差与自相关。

假设扰动项的协方差矩阵 $\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X}) = \sigma^2 \mathbf{V}(\mathbf{X}) \neq \sigma^2 \mathbf{I}_n$, 其中 $\mathbf{V}(\mathbf{X})$ 为对称正定矩阵且已知, 但可能依赖于 \mathbf{X} 。

GLS 的基本思想是, 通过变量转换, 使得转换后的模型满足球形扰动项的假定。

首先介绍一个命题。

命题 对于对称正定矩阵 $\mathbf{V}_{n \times n}$ ，存在非退化矩阵 $\mathbf{C}_{n \times n}$ ，使得 $\mathbf{V}^{-1} = \mathbf{C}'\mathbf{C}$ 。

在一维情况下，“ \mathbf{V} 正定”即要求 \mathbf{V} 为正数，故 $\frac{1}{\mathbf{V}}$ 也是正数，可分解为 $\frac{1}{\sqrt{\mathbf{V}}} \cdot \frac{1}{\sqrt{\mathbf{V}}}$ ；反之，如果 \mathbf{V} 为 0 或负数，则无法进行此分解。

推广到多维情形，就是此命题。

此命题中的矩阵 \mathbf{C} 并不唯一，但不影响 GLS 的最终结果。

根据此命题，对于协方差矩阵 $\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X}) = \sigma^2 \mathbf{V}(\mathbf{X})$ ，首先找到非退化矩阵 \mathbf{C} ，使得 $\mathbf{V}^{-1} = \mathbf{C}'\mathbf{C}$ 。

其次，将原回归模型 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 两边同时左乘矩阵 \mathbf{C} ：

$$\mathbf{C}\mathbf{y} = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\varepsilon} \quad (8.23)$$

定义以下变量转换：

$$\tilde{\mathbf{y}} \equiv \mathbf{C}\mathbf{y}, \quad \tilde{\mathbf{X}} \equiv \mathbf{C}\mathbf{X}, \quad \tilde{\boldsymbol{\varepsilon}} \equiv \mathbf{C}\boldsymbol{\varepsilon} \quad (8.24)$$

则可将模型写为

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}} \quad (8.25)$$

变换后的回归模型仍然满足严格外生性，因为

$$E(\tilde{\boldsymbol{\varepsilon}} | \tilde{\mathbf{X}}) = E(\mathbf{C}\boldsymbol{\varepsilon} | \mathbf{CX}) = E(\mathbf{C}\boldsymbol{\varepsilon} | \mathbf{X}) = \mathbf{C} E(\boldsymbol{\varepsilon} | \mathbf{X}) = \mathbf{0} \quad (8.26)$$

其中，由于 \mathbf{C} 非退化，故 $E(\mathbf{C}\boldsymbol{\varepsilon} | \mathbf{CX}) = E(\mathbf{C}\boldsymbol{\varepsilon} | \mathbf{X})$ 。

球型扰动项的假定也得到满足，因为

$$\begin{aligned} \text{Var}(\tilde{\boldsymbol{\varepsilon}} | \tilde{\mathbf{X}}) &= E(\tilde{\boldsymbol{\varepsilon}} \tilde{\boldsymbol{\varepsilon}}' | \mathbf{X}) = E(\mathbf{C}\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \mathbf{C}' | \mathbf{X}) = \mathbf{C} E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' | \mathbf{X}) \mathbf{C}' = \sigma^2 \mathbf{C} \mathbf{V} \mathbf{C}' \\ &= \sigma^2 \mathbf{C} (\mathbf{V}^{-1})^{-1} \mathbf{C}' = \sigma^2 \mathbf{C} (\mathbf{C}' \mathbf{C})^{-1} \mathbf{C}' = \sigma^2 \mathbf{C} \mathbf{C}^{-1} (\mathbf{C}')^{-1} \mathbf{C}' = \sigma^2 \mathbf{I}_n \end{aligned} \quad (8.27)$$

因此，高斯-马尔可夫定理成立。

对变换后的方程(8.25)使用 OLS 即得到 GLS 估计量：

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{GLS}} &= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}} = [(\mathbf{CX})'(\mathbf{CX})]^{-1} (\mathbf{CX})' \mathbf{C}\mathbf{y} \\ &= (\mathbf{X}' \underbrace{\mathbf{C}'\mathbf{C}}_{\mathbf{V}} \mathbf{X})^{-1} \mathbf{X}' \underbrace{\mathbf{C}'\mathbf{C}}_{\mathbf{V}} \mathbf{y} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\end{aligned}\quad (8.28)$$

虽然 \mathbf{C} 不唯一，但 $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ 唯一，因为 $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ 不依赖于 \mathbf{C} 。

由于高斯-马尔可夫定理成立，故 $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ 是 BLUE，比 OLS 更有效率。

使用 GLS 的前提是，必须知道协方差矩阵 \mathbf{V} 。

由于 \mathbf{V} 通常未知，故 GLS 是不可行的。

实践中，须通过数据估计 $\hat{\mathbf{V}}$ ，再进行 GLS 估计，称为“可行广义最小二乘法” (Feasible GLS, 简记 FGLS)。

WLS 与 PW 都是 GLS 的特例。

FWLS 与可行的 PW 法都是 FGLS 的特例。

何时使用 FGLS 处理自相关？

在使用 FGLS 处理自相关时，如果对自相关系数 ρ 的估计比较准确，而且满足严格外生性，则 FGLS 比 OLS 更有效率。

如果不满足严格外生性，而仅满足前定解释变量(同期外生)的假定，则 FGLS 可能不一致，尽管 OLS 依然一致。

在使用准差分法时，变换后的新扰动项为 $(\varepsilon_t - \rho\varepsilon_{t-1})$ ，而新解释变量为 $(x_t - \rho x_{t-1})$ ；故在同期外生的假定下，二者仍可能存在相关性，比如 $\text{Cov}(\varepsilon_t, x_{t-1}) \neq 0$ ，导致不一致的估计。

FGLS 的适用条件比 OLS 更苛刻，不如 OLS 稳健。

4. 修改模型设定

在有些情况下，自相关的深层原因可能是模型设定有误。

比如，遗漏了自相关的解释变量；或将动态模型(解释变量中包含被解释变量的滞后值)误设为静态模型，而后者也可视为遗漏了解释变量。

假设真实模型为

$$y_t = \alpha + \beta x_t + \rho y_{t-1} + \varepsilon_t \quad (8.29)$$

由于 y_t 是 y_{t-1} 的函数，故 $\{y_t\}$ 存在自相关。

假设此模型被错误地设定为

$$y_t = \alpha + \beta x_t + \underbrace{(\rho y_{t-1} + \varepsilon_t)}_{v_t} \quad (8.30)$$

其中， ρy_{t-1} 被纳入到扰动项 v_t 中，导致扰动项 $\{v_t\}$ 出现自相关，因为 $\{y_{t-1}\}$ 存在自相关。

对于时间序列存在的自相关，有时可通过引入被解释变量的滞后来消除。

对于模型设定误差所导致的自相关，最好从改进模型设定着手解决，而不是机械地使用 FGLS。

8.5 处理自相关的 Stata 命令及实例

1. 时间序列算子

为了在 Stata 中使用时间序列算子(time-series operator), 首先要定义时间变量。

假设时间变量为 *year*, 可使用如下命令:

```
. tsset year
```

其中, “tsset” 表示 time series set, 它告诉 Stata, 该数据集为时间序列, 且时间变量为 *year*。

常用的时间序列算子包括滞后(lag)与差分(difference), 分别以“L.”与“D.”来表示(可以小写)。

一阶滞后算子为“L.”, 即 $L.x_t = x_{t-1}$ 。

二阶滞后算子为“L2.”, 即 $L2.x_t = x_{t-2}$, 以此类推。

如果要同时表示一阶至四阶滞后, 可简写为“L(1/4).”, 即 $L(1/4).x_t = (x_{t-1} \ x_{t-2} \ x_{t-3} \ x_{t-4})$ 。

比如, 命令 `. reg y L.x L2.x L3.x L4.x`

可简写为

`. reg y L(1/4).x`

类似地，“ $L(0/1).(x \ y)$ ”表示
 $L(0/1).(x_t \ y_t) = (x_t \ x_{t-1} \ y_t \ y_{t-1})$ ，其中“0”表示零阶滞后，即当前值。

一阶差分算子为“ $D.$ ”，即 $D.x_t = \Delta x_t = x_t - x_{t-1}$ 。

二阶差分算子为“ $D2.$ ”，即
 $D2.x_t = \Delta(\Delta x_t) = \Delta(x_t - x_{t-1}) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}$
 (二阶差分为一阶差分的差分)。

时间序列算子可以混合使用。“ $LD.$ ”表示一阶差分的滞后值，
 “ $DL.$ ”表示滞后值的一阶差分，二者实际上是等价的，因为
 $LD.x_t = L.(x_t - x_{t-1}) = x_{t-1} - x_{t-2} = D.x_{t-1} = DL.x_t$ 。

有关时间序列算子的更多说明，参见“`help tsvarlist`”。

2. 画残差图

假设在作完回归后，将残差记为 $e1$ ，可输入如下命令画残差与其滞后的散点图：

```
. scatter e1 L.e1
```

如果想看残差自相关图(即各阶自相关系数)，可使用命令

```
. ac e1
```

其中，“ac”表示 autocorrelation(自相关)。

3. BG 检验

作完 OLS 回归后，可使用如下命令进行 BG 检验：

```
. estat bgodfrey, lags(p) nomiss0
```

选择项 “lags(p)” 用来指定 BG 检验的滞后阶数 p ，默认 “lags(1)”，即 $p=1$ 。

选择项 “nomiss0” 表示进行不添加 0 的 BG 检验，默认以 0 代替缺失值，即 Davidson-MacKinnon 的方法。

如何确定滞后阶数 p ？一个简单方法是，看自相关图。

在使用 Stata 命令 `ac` 画自相关图时，所有落在 95% 的置信区域 (以阴影表示) 以外的自相关系数均显著地不等于 0。

确定滞后阶数 p 的另一方法是，设定一个较大的 p 值，作回归

$$e_t = \gamma_1 e_{t-1} + \cdots + \gamma_p e_{t-p} + \delta_2 x_{t2} + \cdots + \delta_K x_{tK} + v_t \quad (t = p+1, \cdots, n)$$

(8.31)

然后看最后一个系数 γ_p 的显著性。

如果 γ_p 不显著，则考虑滞后 $(p-1)$ 期，以此类推，直至显著为止。

4. Q 检验

假设将 OLS 残差记为 $e1$ ，则可使用如下命令进行 Q 检验：

```
. wntestq e1,lags(p)
```

“wntestq”指 white noise test Q ，因为白噪声没有自相关。

选择项 “lags(p)” 用来指定滞后阶数，默认滞后阶数为 $\min\{\text{floor}(n/2)-2, 40\}$ 。

进行 Q 检验的另一命令是

```
. corrgram e1,lags(p)
```

“corrgram”表示 correlogram，即画自相关图。

选择项 “lags(p)” 用来指定滞后阶数，而默认滞后阶数也是 $\min\{\text{floor}(n/2) - 2, 40\}$ 。

5. DW 检验

作完 OLS 回归后可使用命令 “estat dwatson” 显示 DW 统计量。

由于 DW 检验的局限性，Stata 并不提供其临界值。

6. HAC 稳健标准误

在 Stata 中进行 OLS 估计，但提供 Newey-West 标准误，可输入命令

```
. newey y x1 x2 x3, lag(p)
```

必选项 “lag(p)” 用来指定截断参数 p ，即用于计算 HAC 标准误的最高滞后阶数。

7. 处理一阶自相关的 FGLS

在 Stata 中使用准差分法处理自相关，可使用命令

```
. prais y x1 x2 x3, corc
```

选择项 “corc ” 表示使用 CO 估计法，默认使用 PW 估计法。

以 Hildreth and Lu(1960)对冰淇淋需求函数的研究作为例。

数据集 `icecream.dta` 包含了下列变量的 30 个月度时间序列数据：
consumption(人均冰淇淋消费量)，*income*(平均家庭收入)，*price*(冰淇淋价格)，*temp*(平均华氏气温)，*time*(时间)。

首先，打开数据集，并将其设为时间序列数据。

```
. use icecream.dta,clear  
. tsset time
```

其次，为了看冰淇淋的消费量与气温的时间趋势图，输入命令：

```
. twoway connect consumption  
time,msymbol(circle) yaxis(1) || connect temp  
time, msymbol(triangle) yaxis(2)
```

“connect”表示将观测点用线连接起来，选择项“msymbol(circle)”与“msymbol(triangle)”分别表示点的“图标”(marker symbol)分别为圆圈与三角形。

选择项“yaxis(1)”与“yaxis(2)”指定使用不同的纵坐标，因为冰淇淋消费量与气温的取值范围很不相同；结果参见图8.3。

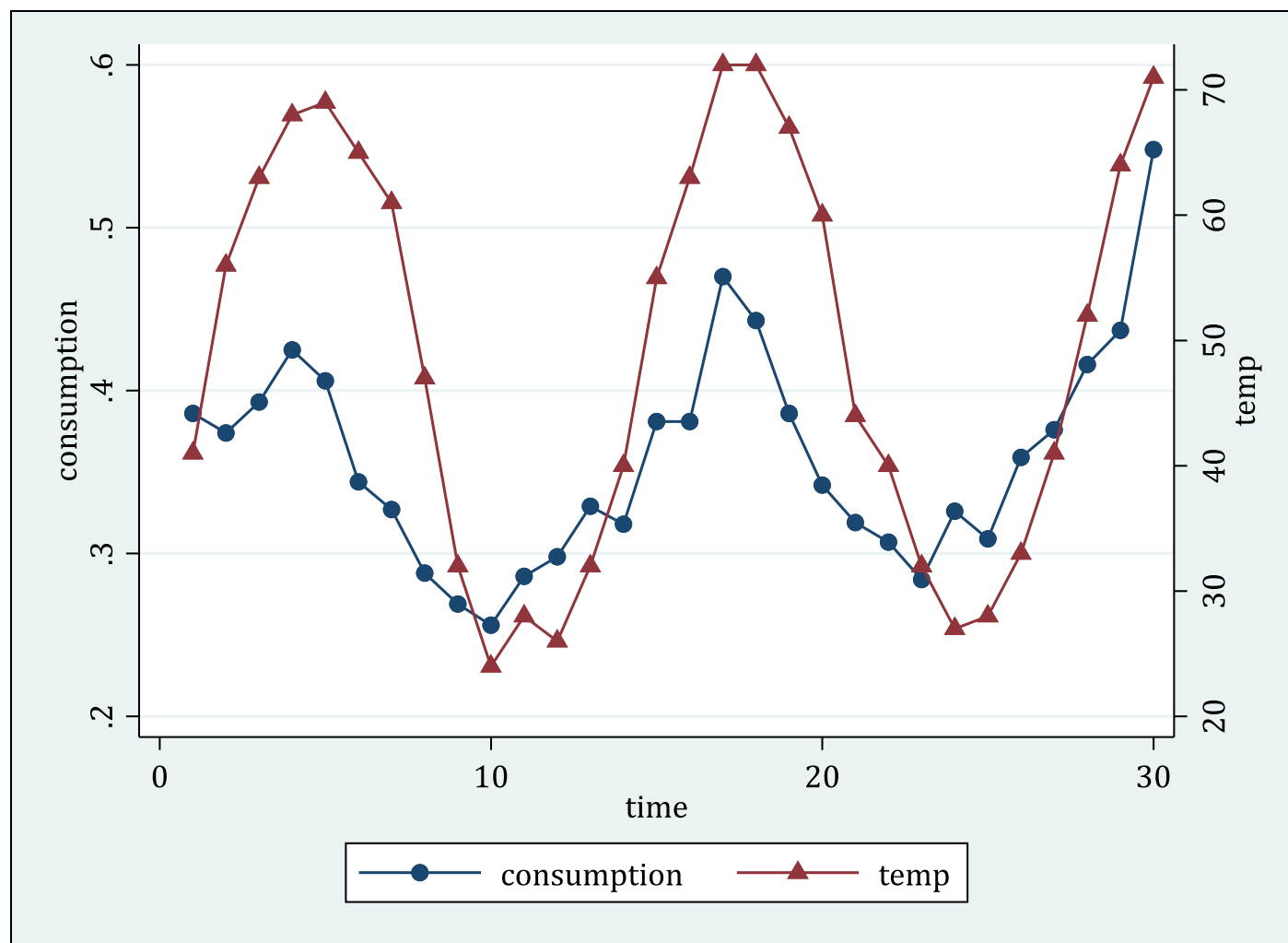


图 8.3 冰淇淋消费与气温的时间趋势

冰淇淋消费量与温度明显地正相关。

考虑以下线性回归模型：

$$\text{consumption}_t = \beta_0 + \beta_1 \text{temp}_t + \beta_2 \text{price}_t + \beta_3 \text{income}_t + \varepsilon_t \quad (8.32)$$

首先进行 OLS 回归：

```
. reg consumption temp price income
```

Source	SS	df	MS	Number of obs	=	30
				F(3, 26)	=	22.17
Model	.090250523	3	.030083508	Prob > F	=	0.0000
Residual	.035272835	26	.001356647	R-squared	=	0.7190
				Adj R-squared	=	0.6866
Total	.125523358	29	.004328392	Root MSE	=	.03683
consumption	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
temp	.0034584	.0004455	7.76	0.000	.0025426	.0043743
price	-1.044413	.834357	-1.25	0.222	-2.759458	.6706322
income	.0033078	.0011714	2.82	0.009	.0008999	.0057156
_cons	.1973149	.2702161	0.73	0.472	-.3581223	.752752

气温(*temp*)与收入(*income*)均在 1%的水平上显著为正,表示气温越高、收入越高,则冰淇淋的消费量越大。

价格(*price*)的系数为负,表明价格越高,则消费量越低,但并不显著(*p*值为 0.222)。

由于这是时间序列，我们怀疑其扰动项存在自相关。

首先计算残差(记为 $e1$)，及其滞后值($L.e1$)，然后画残差与残差滞后的散点图：

```
. predict e1,r
```

```
. twoway scatter e1 L.e1 || lfit e1 L.e1
```

其中，“`lfit`”表示 linear fit(线性拟合)，即画出 $e1$ 与 $L.e1$ 的拟合回归线；结果参见图 8.4。

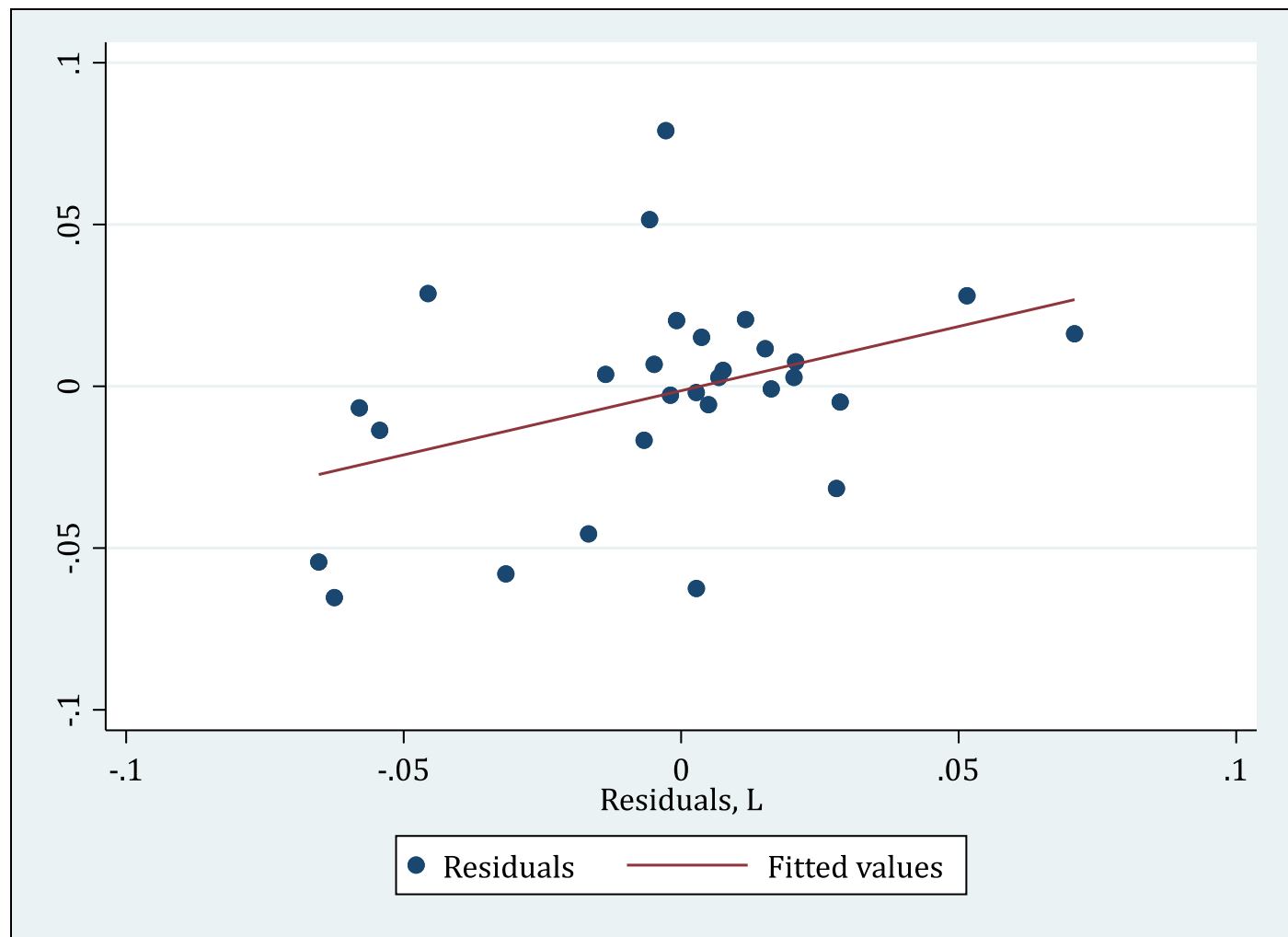


图 8.4 残差与残差滞后的散点图

扰动项很可能存在一阶正自相关。

作为对比,下面画残差与其二阶滞后的散点图,结果参见图 8.5:

```
. twoway scatter e1 l2.e1 || lfit e1 l2.e1
```

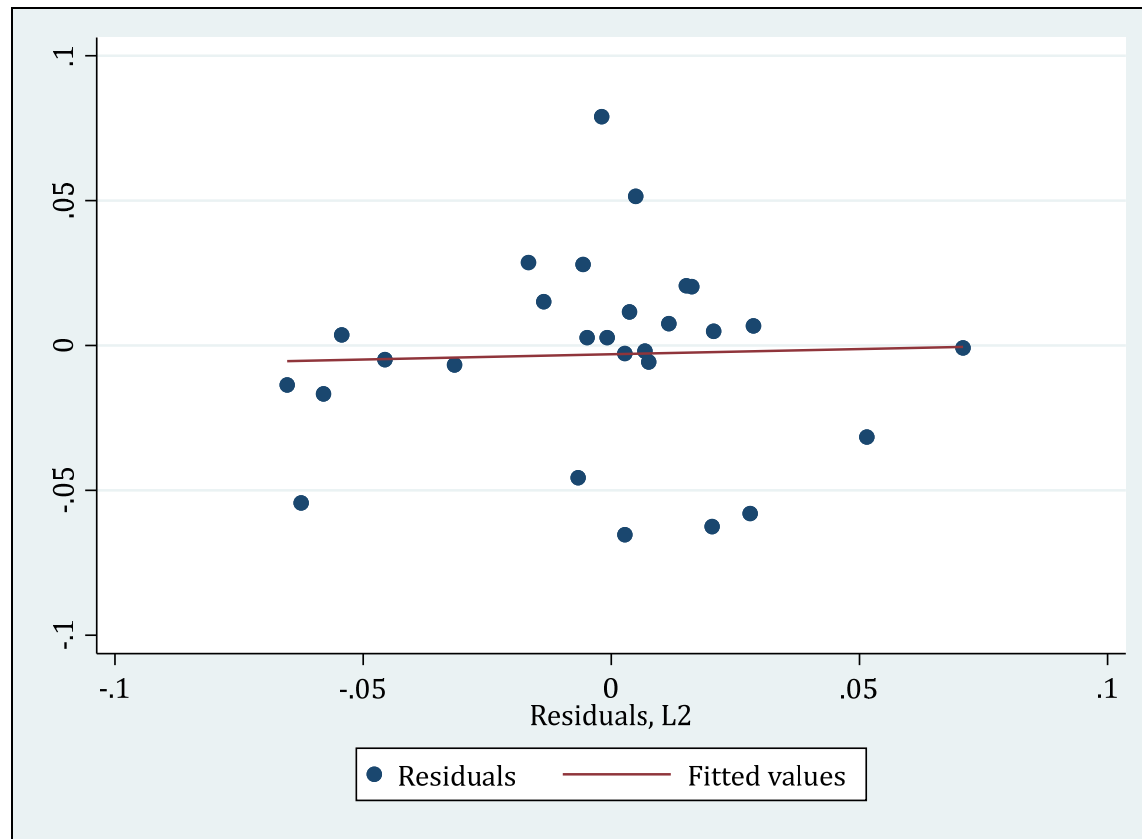


图 8.5 残差与二阶残差滞后的散点图

残差似乎不存在二阶自相关。

为看各阶自相关系数及显著性,画残差的自相关图,参见图 8.6。

. ac e1

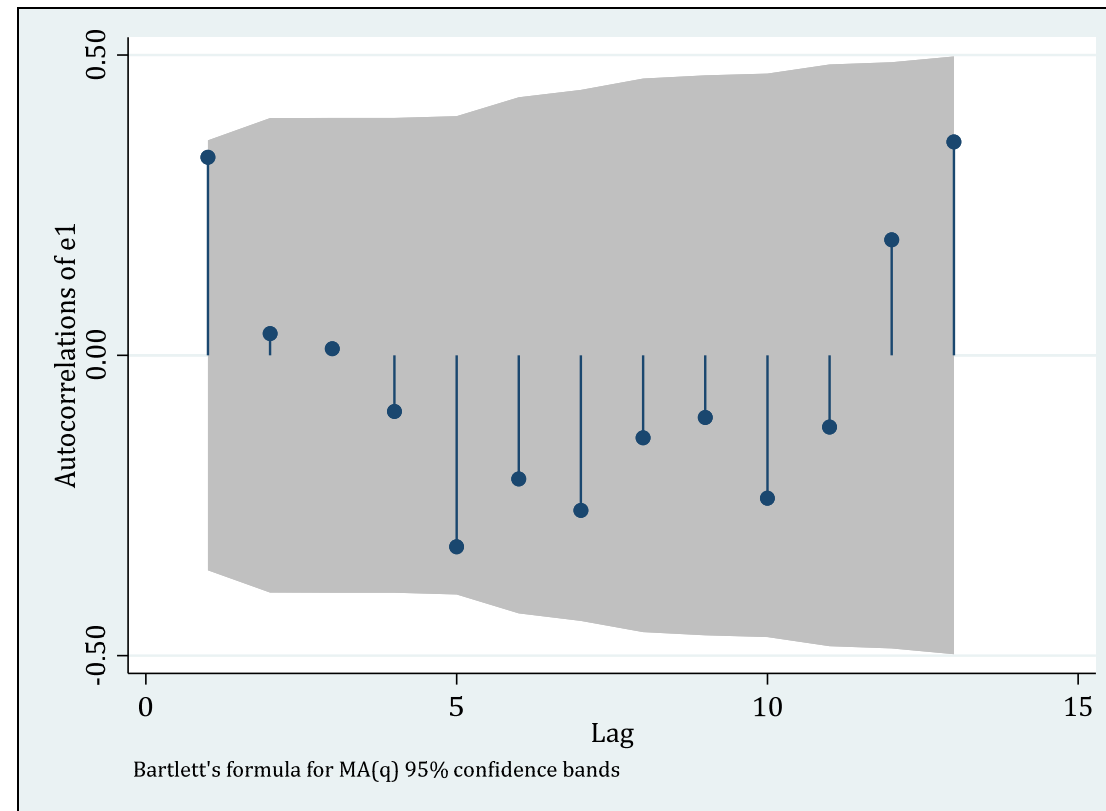


图 8.6 自相关图

图 8.6 的横轴为滞后阶数，纵轴为残差的自相关系数，而阴影部分为置信度为 95% 的置信区间(区域)。

图 8.6 显示，各阶自相关系数的取值均在 95% 的置信区间之内，故可接受各阶自相关系数为 0 的原假设。

但一阶自相关系数已很接近置信区间的边界，故仍怀疑存在一阶自相关，而更高阶自相关则可大致忽略。

下面进行正式的 BG 检验，考察是否存在一阶自相关：

```
. estat bgodfrey
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	4.237	1	0.0396

H0: no serial correlation

BG 检验的 p 值为 0.039 6，故可在 5% 的显著性水平上拒绝“无自相关”的原假设，而认为存在自相关。

如果不以 0 取代缺失值，可输入命令

```
. estat bgodfrey,nomiss0
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	4.704	1	0.0301

H0: no serial correlation

结果依然可在 5%的水平上拒绝“无自相关”的原假设。

下面进行 Q 检验。

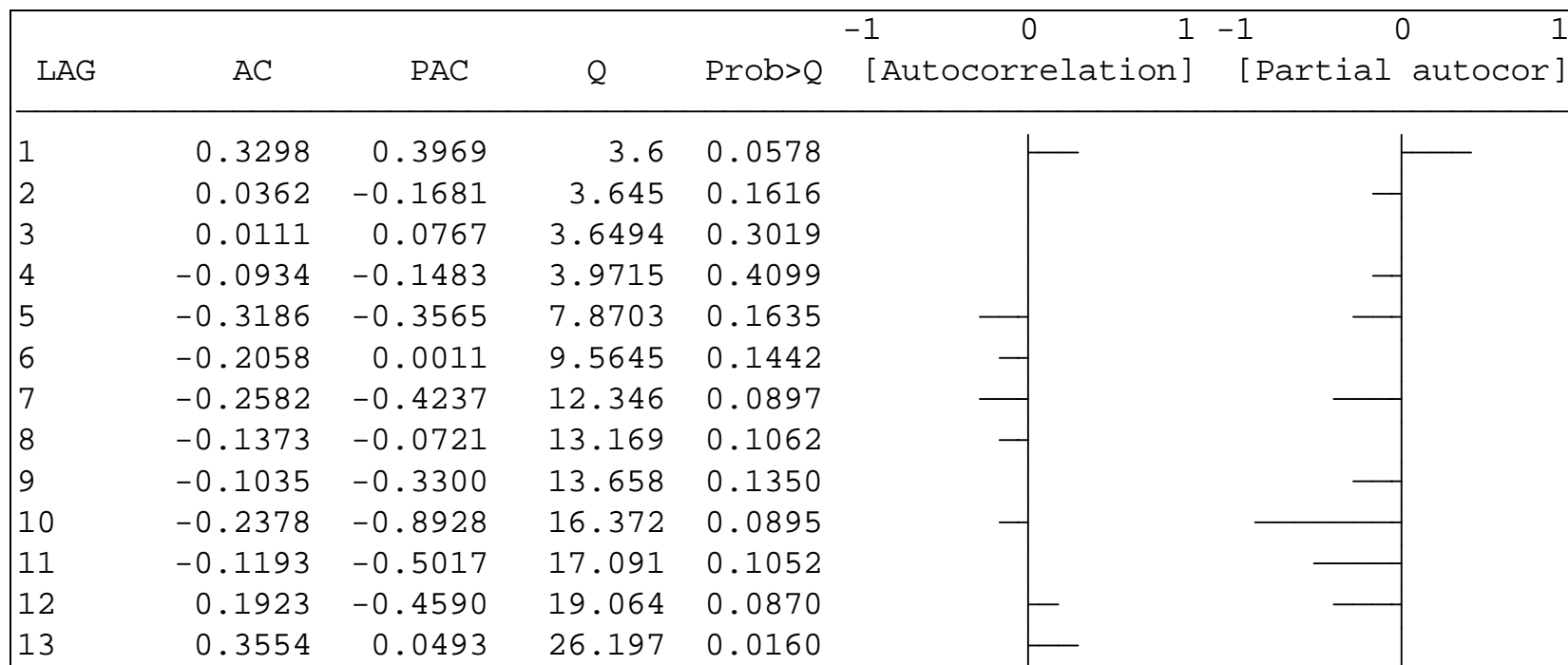
```
. wntestq e1
```

Portmanteau test for white noise	
Portmanteau (Q) statistic =	26.1974
Prob > chi2(13) =	0.0160

“Prob > chi2(13)= 0.016”表明默认的滞后阶数为 13 阶，且可在 5%的水平上拒绝“无自相关”的原假设。

下面使用命令 corrgram 进行 Q 检验。

```
. corrgram e1
```



上表汇报了从 1-13 阶的自相关系数(AC), Q 统计量(Q)及其相应 p 值(Prob> Q)。

第 13 阶 Q 统计量及其 p 值与命令 `wntestq` 的结果完全相同。

使用命令 `corrgram` 的好处在于,它同时计算了各阶 Q 统计量。

作为最后一个自相关检验,下面计算 DW 统计量:

```
. estat dwatson
```

Durbin-Watson d-statistic(4, 30) = 1.021169
--

由于 DW 统计量的局限性, Stata 并未提供其临界值。

由于 DW=1.02, 离 2 较远而靠近 0, 可大致判断存在正自相关。

由于扰动项存在自相关，故 OLS 估计所提供的标准误不准确，应使用异方差自相关稳健的 HAC 标准误。

由于 $n^{1/4} = 30^{1/4} \approx 2.34$ ，故取 Newey-West 估计量的截断参数为 $p=3$ ：

```
. newey consumption temp price income, lag(3)
```

Regression with Newey-West standard errors				Number of obs	=	30
Maximum lag = 3				F(3,	26)	= 27.63
				Prob > F	=	0.0000
consumption	Newey-West					
	Coefficient	std. err.	t	P> t	[95% conf. interval]	
temp	.0034584	.0004002	8.64	0.000	.0026357	.0042811
price	-1.044413	.9772494	-1.07	0.295	-3.053178	.9643518
income	.0033078	.0013278	2.49	0.019	.0005783	.0060372
_cons	.1973149	.3378109	0.58	0.564	-.4970655	.8916952

Newey-West 标准误与 OLS 标准误相差无几(但略大)。

考察 Newey-West 标准误是否对于截断参数敏感，将滞后阶数增大一倍，变为 6，再重新估计。

```
. newey consumption temp price income, lag(6)
```

Regression with Newey-West standard errors			Number of obs		=	30
Maximum lag = 6			F(3, 26)		=	52.97
			Prob > F		=	0.0000
consumption	Newey-West		t	P> t	[95% conf. interval]	
	Coefficient	std. err.				
temp	.0034584	.0003504	9.87	0.000	.0027382	.0041787
price	-1.044413	.9821798	-1.06	0.297	-3.063313	.9744864
income	.0033078	.00132	2.51	0.019	.0005945	.006021
_cons	.1973149	.3299533	0.60	0.555	-.4809139	.8755437

无论截断参数为 3 还是 6，Newey-West 标准误变化不大。

由于存在自相关，OLS 不再是 BLUE，故可考虑使用 FGLS，对模型进行更有效率的估计。

首先使用 CO 估计法：

```
. prais consumption temp price income, corc
```


Iteration 0: rho = 0.0000						
Iteration 1: rho = 0.4006						
Iteration 2: rho = 0.4008						
Iteration 3: rho = 0.4009						
Iteration 4: rho = 0.4009						
Iteration 5: rho = 0.4009						
Iteration 6: rho = 0.4009						
Iteration 7: rho = 0.4009						
Cochrane-Orcutt AR(1) regression with iterated estimates						
Source	SS	df	MS	Number of obs	=	29
				F(3, 25)	=	15.40
Model	.047040596	3	.015680199	Prob > F	=	0.0000
Residual	.025451894	25	.001018076	R-squared	=	0.6489
				Adj R-squared	=	0.6068
Total	.072492491	28	.002589018	Root MSE	=	.03191
consumption	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
temp	.0035584	.0005547	6.42	0.000	.002416	.0047008
price	-.8923963	.8108501	-1.10	0.282	-2.562373	.7775807
income	.0032027	.0015461	2.07	0.049	.0000186	.0063869
_cons	.1571479	.2896292	0.54	0.592	-.4393546	.7536504
rho	.4009256					
Durbin-Watson statistic (original) = 1.021169						
Durbin-Watson statistic (transformed) = 1.548837						

使用 CO 估计法得到的系数估计值与 OLS 比较接近，但样本容量降为 29(损失一个样本观测值)。

经过模型转换后 DW 值改进为 1.55。

然后使用 PW 估计法：

```
. prais consumption temp price income,nolog
```

其中，选择项 “nolog” 表示不显示迭代过程。

Prais-Winsten AR(1) regression with iterated estimates						
Source	SS	df	MS	Number of obs	=	30
				F(3, 26)	=	14.35
Model	.04494596	3	.014981987	Prob > F	=	0.0000
Residual	.027154354	26	.001044398	R-squared	=	0.6234
				Adj R-squared	=	0.5799
Total	.072100315	29	.002486218	Root MSE	=	.03232
consumption	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
temp	.0029541	.0007109	4.16	0.000	.0014929	.0044152
price	-1.048854	.759751	-1.38	0.179	-2.610545	.5128361
income	-.0008022	.0020458	-0.39	0.698	-.0050074	.0034029
_cons	.5870049	.2952699	1.99	0.057	-.0199311	1.193941
rho	.8002264					
Durbin-Watson statistic (original) = 1.021169						
Durbin-Watson statistic (transformed) = 1.846795						

虽然使用 PW 估计法使 DW 统计量进一步改进为 1.85，但收入 (*income*) 的系数估计值却变为负数(-0.0008)。

尽管它只是绝对值很小的负数，且在统计上不显著，但由于 PW 估计法使得收入的系数估计值与理论预期相反，似乎 PW 估计法反而不如 OLS 稳健。

自相关的存在可能是由于模型设定不正确。

考虑在解释变量中加入气温(*temp*)的滞后值，然后进行 OLS 回归：

```
. reg consumption temp L.temp price income
```

Source	SS	df	MS	Number of obs	=	29
				F(4, 24)	=	28.98
Model	.103387183	4	.025846796	Prob > F	=	0.0000
Residual	.021406049	24	.000891919	R-squared	=	0.8285
				Adj R-squared	=	0.7999
Total	.124793232	28	.004456901	Root MSE	=	.02987

consumption	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
temp						
--.	.0053321	.0006704	7.95	0.000	.0039484	.0067158
L1.	-.0022039	.0007307	-3.02	0.006	-.0037119	-.0006959
price	-.8383021	.6880205	-1.22	0.235	-2.258307	.5817025
income	.0028673	.0010533	2.72	0.012	.0006934	.0050413
_cons	.1894822	.2323169	0.82	0.423	-.2899963	.6689607

气温的滞后项(L.temp)在 1%的水平上显著地不等于 0，但符号为负(系数为-0.0022)。

当期气温仍然显著地为正(系数为 0.0053)。

这可能意味着，当气温上升时，对冰淇淋的需求上升，但不会在当月全部消费完，而增加冰箱中的冰淇淋库存，导致下期对冰淇淋的开支下降。

使用 BG 检验判断重新设定的模型是否存在自相关：

```
. estat bgodfrey
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.120	1	0.7292

H0: no serial correlation

由于 p 值为 0.73，故可放心接受“无自相关”的原假设。

下面计算 DW 统计量。

```
. estat dwatson
```

Durbin-Watson d-statistic(5, 29) = 1.582166
--

DW 值也改进为 1.58。

通过修改模型设定，加入气温的滞后项后，扰动项基本上不再存在自相关。

究竟应该使用以上哪种模型，一定程度上取决于研究者的判断。

可在研究报告中同时列出各种模型的结果，以此说明系数估计值与标准误的稳健性(不依估计方法的改变而剧烈变化)。