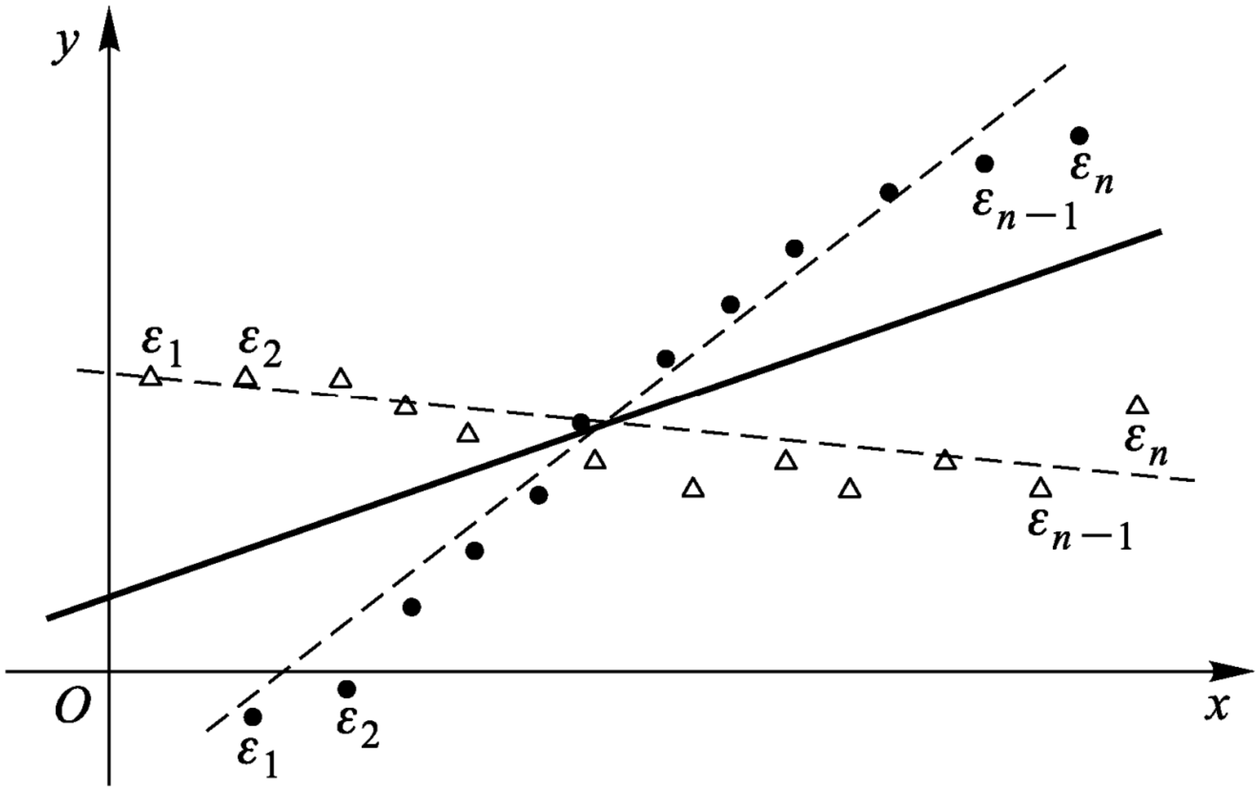
第6章　自相关

自相关的后果

- 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$\{\varepsilon_1, \cdots, \varepsilon_n\}$，如果存在$i \neq j$，使得$E(\varepsilon_i \varepsilon_j | \boldsymbol{X}) \neq 0$，即协方差矩阵$Var(\boldsymbol{\varepsilon} | \boldsymbol{X})$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)。

- 在自相关的情况下：
- (1) OLS 估计量依然无偏、一致且渐近正态，因为证明这些性质时，并未用到“无自相关”的假定。
- (2) OLS 估计量方差$Var(\hat{\boldsymbol{\beta}}|\boldsymbol{X})$的表达式不再是$\sigma^2(X'X)^{-1}$。因为$Var(\boldsymbol{\varepsilon}|\boldsymbol{X}) \neq \sigma^2\boldsymbol{I}$。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，OLS 不再是 BLUE。

- 为直观理解在自相关的情况下，OLS 不再是 BLUE，假设扰动项存在正自相关，即$E(\varepsilon_i \varepsilon_j|\boldsymbol{X}) > 0$。



- 实线表示真实的总体回归线。
- 如果$\varepsilon_1 > 0$，由于扰动项存在自相关，则$\varepsilon_2 > 0$的可能性也很大。

  如果$\varepsilon_{n-1} < 0$，则$\varepsilon_n < 0$的可能性也就很大。

  样本回归线(虚线)很可能左侧翘起、右侧下垂，使得对回归线斜率的估计过小。
- 反之，如果$\varepsilon_1 < 0$，由于扰动项存在正自相关，故$\varepsilon_2 < 0$的可能性也很大。

  如果$\varepsilon_{n-1} > 0$（图中右边小圆点），则$\varepsilon_n > 0$的可能性也就很大。

  样本回归线(虚线)很可能左侧下垂、右侧翘起，使得对回归线斜率的估计过大。

以上只是**可能**的情形。

- 由于自相关的存在，使得样本回归线上下摆动幅度增大，导致参数估计变得不准确。
- 从信息角度，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。

自相关的例子

- **(1) 时间序列**：由于经济活动通常具有连续性或持久性，自相关在时间序列中较常见。
- 例：相邻两年的 GDP 增长率、通货膨胀率。
- 例：某意外事件或新政策的效应需要随时间逐步释放出来。
- 例：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**：截面数据不易出现自相关，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)。
- 例：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)
- 例：相邻地区的农业产量受到类似天气变化的影响
- 例：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)、内插值或季节调整时，可从理论上判断存在自相关。
- 统计局提供的某些数据可能事先经过了人为处理。

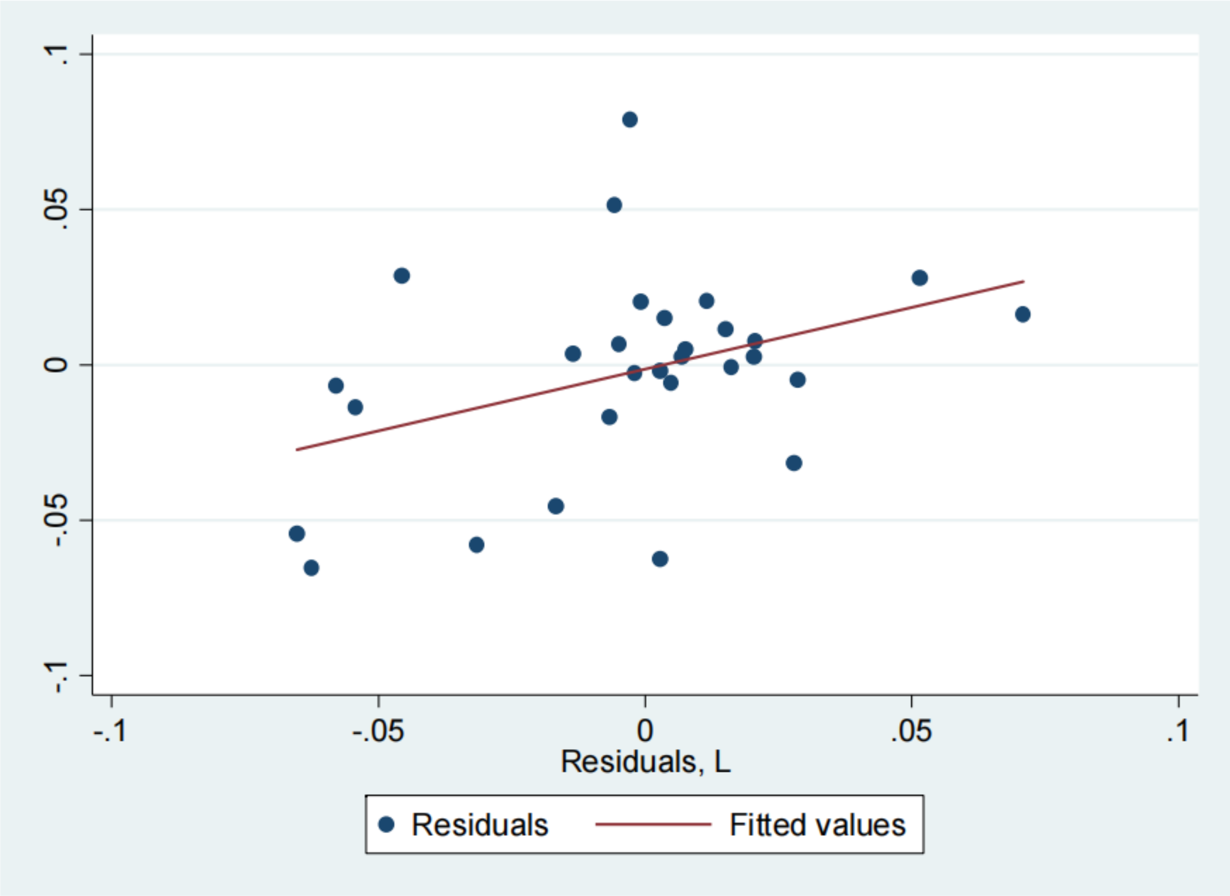
- **(4) 设定误差(misspecification)**：如果模型设定中遗漏了某个自相关的解释变量，并被纳入到扰动项中，会引起扰动项的自相关。

自相关的检验

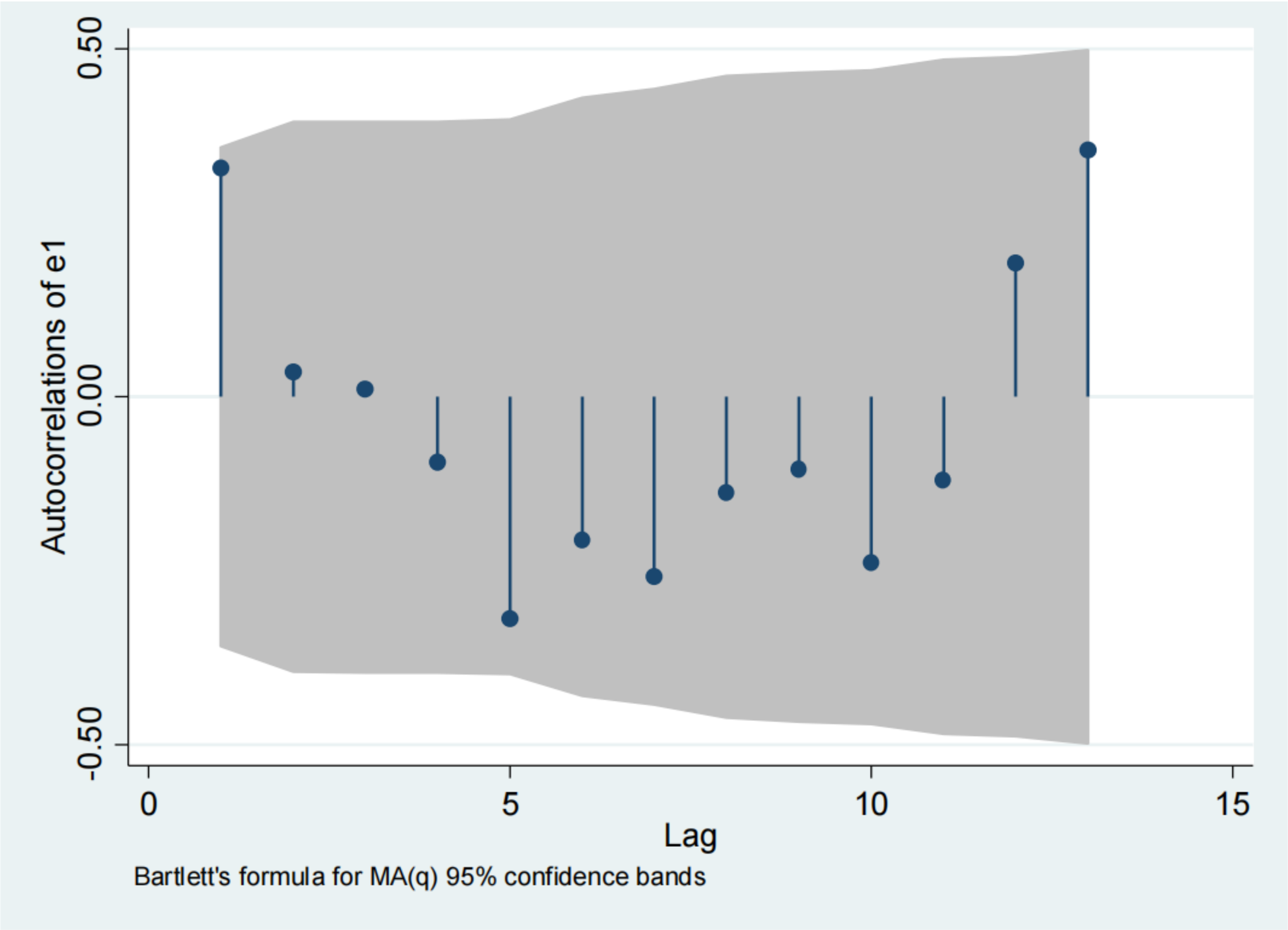
- 1．画图
- 2．BG 检验(Breusch, 1978；Godfrey, 1978)
- 3．Q 检验
- 4．DW 检验

- 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。
- 也可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$\hat{\rho}_2$，乃至k阶相关系数$\hat{\rho}_k$。
- 相关系数$\hat{\rho}_k$是之后阶数k的函数，将$(k, \hat{\rho}_k)$画图，可得残差的“自相关图”(correlogram)。

残差与残差滞后的散点图



• 自相关图



0.50
0.00
-0.50
Autocorrelations of e1
0
5
10
15
Lag
Bartlett's formula for MA(q) 95% confidence bands

- 2．BG 检验
- 考虑多元线性模型：

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t$$

- 假设扰动项ε_t存在一阶自相关，即

$$\varepsilon_t = \gamma \varepsilon_{t-1} + \mu_t$$

其中，μ_t为白噪声。$\varepsilon_t = \gamma \varepsilon_{t-1} + \mu_t$没有常数项，因为$E(\varepsilon_t) = 0$。

- 为检验是否存在一阶自相关，只要检验$H_0: \gamma = 0$即可。
- 由于可能存在高阶自相关，考虑扰动项的p阶自回归：

$$\varepsilon_t = \gamma_1 \varepsilon_{t-1} + \cdots + \gamma_p \varepsilon_{t-p} + \mu_t$$

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

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：

$$e_t = \gamma_1 e_{t-1} + \cdots + \gamma_p e_{t-p} + \delta_2 x_{t2} + \cdots + \delta_k x_{tk} + \nu_t \quad (t = p+1, \cdots, n)$$

- 由于残差e_t是解释变量$(x_{t2},\cdots,x_{tk})$的函数，如果遗漏$(x_{t2},\cdots,x_{tk})$，可能导致扰动项ν_t与$(e_{t-1},\cdots,e_{t-\mathrm{p}})$相关，使得估计不一致。
- 在辅助回归中，“无自相关”的原假设相当于检验H_0: $\gamma_1=\cdots=\gamma_p=0$，通常用$nR^2$形式的LM统计量进行检验：

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

- 由于辅助回归使用了$e_{t-\mathrm{p}}$，损失p个样本观测值，故样本容量仅为$(n-p)$。
- 如果LM 超过$\chi^2(p)$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 $nR^2\xrightarrow{d}\chi^2(p)$。 (Stata默认方法)

- 3．Q 检验
- 记$\rho_1, \cdots, \rho_p$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

  $H_0: \rho_1 = \cdots = \rho_p = 0$。
- 定义残差的各阶样本自相关系数为

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

- 如果$H_0: \rho_1 = \cdots = \rho_p = 0$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。
- 故$\sqrt{n}\hat{\rho}_j$的平方和（对j求和）为渐近卡方分布，即“Box-Pierce Q统计量”(Box and Pierce,1970):

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

- 经改进的“Ljun-Box Q”（Ljung and Box,1979）为

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

- 这两种 Q 统计量在大样本下等价，但 Ljung-Box Q 统计量的小样本性质更好（Stata所采用）。

- 如何确定自相关阶数 p呢?
- 如果p太小，可能忽略高阶自相关的存在。
- 如果p较大(与样本容量n相比)，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{floor(n/ 2) -2, 40}，其中floor(n/ 2) 为不超过n / 2的最大整数，并在[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，已不常用。
- 主要缺点是只能检验一阶自相关，且须在解释变量满足严格外生性的情况下才成立(BG 检验、Q检验无此限制)。
- DW 检验的统计量为

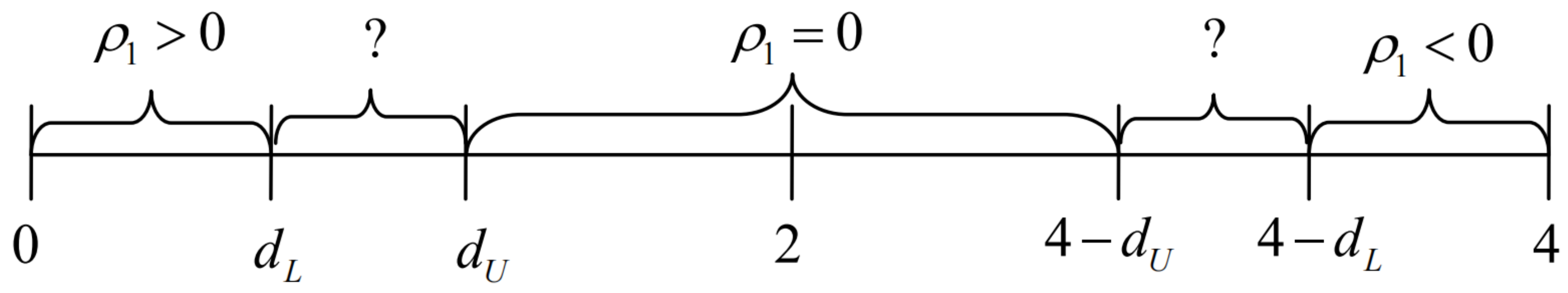
$$\mathrm{DW} \equiv d \equiv \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{\sum_{t=2}^{n} e_t^2 - 2\sum_{t=2}^{n} e_t e_{t-1} + \sum_{t=2}^{n} e_{t-1}^2}{\sum_{t=1}^{n} e_t^2}$$

$$\approx 2 - 2\frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \equiv 2(1 - \hat{\rho}_1)$$

- 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统计量的分布还依赖于数据矩阵$\boldsymbol{X}$，无法制表，须使用其上限分布d_U与下限分布$d_L(d_L < d < d_U)$来间接检验。
- 即便如此，仍存在“无结论区域”
- DW 统计量的本质就是残差的一阶自相关系数，没有太多信息。

- 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：
- （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$4 - d_L \leq d < 4$，则存在负自相关。



DW 检验的无结论区域

自相关的处理

- 1．使用“OLS + 异方差自相关稳健的标准误”
- 2．准差分法
- 3．广义最小二乘法(GLS)
- 4．修改模型设定

- 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，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)，只改变标准误的估计值，不改变回归系数的估计值。

- 异方差稳健的协方差矩阵为夹心估计量：

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 类似地，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”$\widehat{Var(\boldsymbol{\varepsilon}|\boldsymbol{X})}$更复杂。
- 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。
- 另一方面，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$(n-1)$随样本容量$n$同步增长，也将导致估计量不一致。

- 而且，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，因为只有两对数据(e_1, e_{n-1})、(e_2, e_n)可用于估计；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$p = 0.75n^{1/3}$称为“截断参数”(truncation parameter)，即比 p更高阶的自相关系数被截断而不考虑。
- 实践中，建议使用不同的截断参数p，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 WLS 的思路，如能够变换原模型，使得转换后的扰动项变为球形扰动项，可得到更有效率的估计。
- 假设原模型为

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t \ \ (t = 1, \cdots, n)$$

其中，扰动项ε_t存在自相关，且为一阶自回归形式:

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

其中，自回归系数$|\rho| < 1$，且u_t为白噪声。

- 将原模型滞后一期，然后在方程两边同乘ρ

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

- 将原方程减去上述方程可得

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

- 其中，$t = 2, \cdots, n$，故损失一个样本观测值。
- 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。
- 对方程进行OLS 估计，可提高估计效率，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$y_t - y_{t-1}$）。

- 准差分法将损失一个样本容量，仍不是BLUE。
- 为得到 BLUE 估计量，补上损失的第一个方程：

$$y_1 = \beta_1 + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1$$

- 由于$\{\mu_t\}_{t=1}^{n}$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。
- 加入第一个方程不会导致自相关；但会导致异方差。
- 第一个方程的扰动项方差为$\sigma_\varepsilon^2 \equiv Var(\varepsilon_t)$
- 准差分方程的扰动项方差为$\sigma_u^2 \equiv Var(u_t)$

- 对扰动项ε_t的一阶自回归方程$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$两边求方差，可得$\sigma_\varepsilon^2$与$\sigma_u^2$之间的关系：

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

- 整理可得$\sigma_u^2 = (1-\rho^2)\sigma_\varepsilon^2$
- 故σ_u^2是σ_ε^2的$(1-\rho^2)$倍，除非$\rho = 0$(无自相关)，否则二者不相等。
- 将第一个方程两边同乘$\sqrt{1-\rho^2}$：

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

- 该方程的扰动项方差为

$$Var(\sqrt{1-\rho^2}\varepsilon_1) = (1-\rho^2)\sigma_\varepsilon^2 = \sigma_u^2$$

- 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。
- 进行 OLS 估计，可得 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)，因为都假设知道一阶自回归系数ρ。
- 实践中，须用数据估计一阶自回归系数$\hat{\rho}$。
- （1）Stata 默认的方法使用 OLS 残差进行辅助回归：

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

- （2）可用残差的一阶自相关系数来估计$\hat{\rho}$

$$\hat{\rho} \equiv \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

- （3）通过DW统计量来估计$\hat{\rho}$

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

- 实践中，常使用迭代法。
- 首先，用 OLS 估计原模型，使用 OLS 残差作辅助回归得到$\hat{\rho}^{(1)}$（对ρ的第一轮估计），再用$\hat{\rho}$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)
- 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 Square，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$Var(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2\boldsymbol{V}(\boldsymbol{X}) \neq \sigma^2\boldsymbol{I}_n$，其中$\boldsymbol{V}(\boldsymbol{X})$为对称正定矩阵且已知，但可能依赖于$\boldsymbol{X}$。
- GLS 的基本思想是，通过变量转换，使得转换后的模型满足球型扰动项的假定。

- **命题** 对于对称正定矩阵$\boldsymbol{V}_{n\times n}$，存在非退化矩阵$\boldsymbol{C}_{n\times n}$，使得$\boldsymbol{V}^{-1}=\boldsymbol{C}'\boldsymbol{C}$。
- 在一维情况下，“$\boldsymbol{V}$正定”即要求$\boldsymbol{V}$为正数，故$\frac{1}{\boldsymbol{V}}$也是正数，可分解为$\frac{1}{\sqrt{\boldsymbol{V}}}\cdot\frac{1}{\sqrt{\boldsymbol{V}}}$；反之，如果$\boldsymbol{V}$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$\boldsymbol{C}$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$Var(\boldsymbol{\varepsilon}|\boldsymbol{X})=\sigma^2\boldsymbol{V}(\boldsymbol{X})$，首先找到非退化矩阵$\boldsymbol{C}$，使得$\boldsymbol{V}^{-1}=\boldsymbol{C}'\boldsymbol{C}$

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

$$\boldsymbol{Cy} = \boldsymbol{CX\beta} + \boldsymbol{C\varepsilon}$$

定义变量转换

$$\widetilde{\boldsymbol{y}} \equiv \boldsymbol{Cy}, \widetilde{\boldsymbol{X}} \equiv \boldsymbol{CX}, \tilde{\boldsymbol{\varepsilon}} \equiv \boldsymbol{C\varepsilon}$$

- 将模型写为

$$\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$$

- 变换后的回归模型仍满足严格外生性，因为

$$E\left(\tilde{\boldsymbol{\varepsilon}}\middle|\widetilde{\boldsymbol{X}}\right) = E(\boldsymbol{C\varepsilon}|\boldsymbol{CX}) = E(\boldsymbol{C\varepsilon}|\boldsymbol{X}) = \boldsymbol{C}E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{0}$$

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

- 球型扰动项的假定也得到满足，因为
- $Var\left(\tilde{\boldsymbol{\varepsilon}}\middle|\widetilde{\boldsymbol{X}}\right) = E(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}'|\boldsymbol{X}) = E(\boldsymbol{C}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{C}'|\boldsymbol{X}) = \boldsymbol{C}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X})\boldsymbol{C}' = \sigma^2\boldsymbol{C}\boldsymbol{V}\boldsymbol{C}' = \sigma^2\boldsymbol{C}(\boldsymbol{V}^{-1})^{-1}\boldsymbol{C}' = \sigma^2\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}' = \sigma^2\boldsymbol{C}\boldsymbol{C}^{-1}(\boldsymbol{C}')^{-1}\boldsymbol{C}' = \sigma^2\boldsymbol{I}_n$
- 因此，高斯-马尔可夫定理成立。
- 对变换后的方程使用 OLS 即得到 GLS 估计量：

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{y}} = [(\boldsymbol{C}\boldsymbol{X})'(\boldsymbol{C}\boldsymbol{X})]^{-1}(\boldsymbol{C}\boldsymbol{X})'\boldsymbol{C}\boldsymbol{y}$$

$$= (\boldsymbol{X}'\underbrace{\boldsymbol{C}'\boldsymbol{C}}_{\boldsymbol{V}^{-1}}\boldsymbol{X})^{-1}\boldsymbol{X}'\underbrace{\boldsymbol{C}'\boldsymbol{C}}_{\boldsymbol{V}^{-1}}\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$

- 虽然$\boldsymbol{C}$不唯一，但$\widehat{\boldsymbol{\beta}}_{GLS}$唯一，因为$\widehat{\boldsymbol{\beta}}_{GLS}$不依赖于$\boldsymbol{C}$。
- 由于高斯-马尔可夫定理成立，故$\widehat{\boldsymbol{\beta}}_{GLS}$是BLUE，比OLS更有效率。
- 使用GLS的前提是，须知道协方差矩阵$\boldsymbol{V}$。
- 由于$\boldsymbol{V}$通常未知，故 GLS 不可行。

- 实践中，须通过数据估计$\hat{\boldsymbol{V}}$，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，而 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，且满足严格外生性，则 FGLS 比 OLS 更有效率。
- 但如不满足严格外生性，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$Cov(\varepsilon_t, x_{t-1}) \neq 0$，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

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

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

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

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