

第 8 章 自 相 关

8.1 自相关的后果

如果存在 $i \neq j$, 使得 $E(\varepsilon_i \varepsilon_j | \mathbf{X}) \neq 0$, 即 $\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X})$ 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 估计量依然无偏且一致, 因为在证明这些性质时, 并未用到“无自相关”的假定;

(2) OLS 估计量依然服从渐近正态分布;

(3) OLS 估计量方差 $\text{Var}(\mathbf{b} | \mathbf{X})$ 的表达式不再是 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, 因为 $\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X}) \neq \sigma^2 \mathbf{I}$, 通常的 t 检验、 F 检验也失效了;

(4) 高斯-马尔可夫定理不再成立, 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。

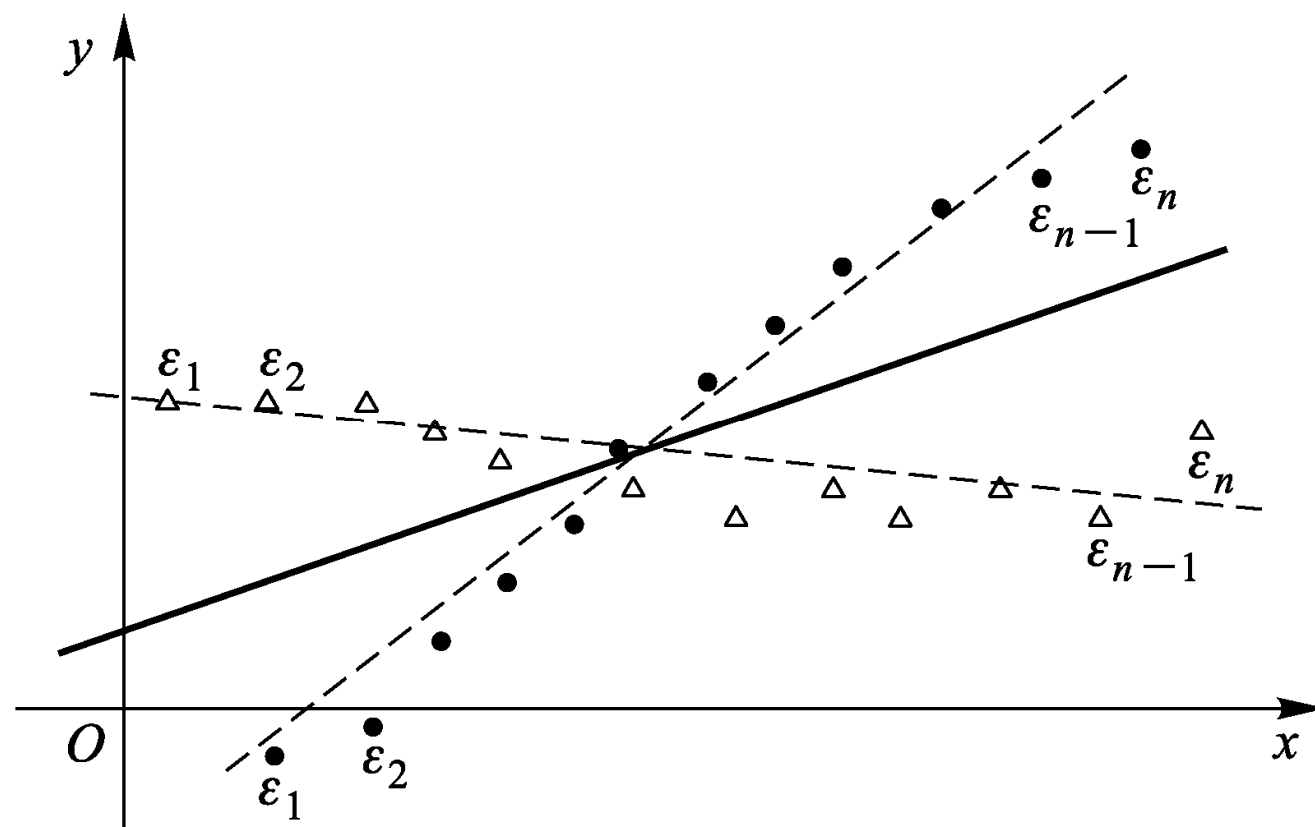


图 8.1 自相关的后果

8.2 自相关的例子

(1) 时间序列的自相关：经济活动具有持久性，比如，相邻两年的 **GDP** 增长率、通货膨胀率；意外事件或新政策的效应需逐步释放；最优资本存量需若干年投资才能达到(滞后的调整过程)。

(2) 截面数据的自相关：相邻单位间可能存在“溢出效应”(spillover effect or neighborhood effect)，称为“空间自相关”(spatial autocorrelation)。比如，相邻省份、国家间的经济活动相互影响；相邻地区的农产量受类似天气影响；同一社区内房屋价格相关。

(3) 对数据的人为处理：数据中包含移动平均数(moving average)、内插值或季节调整时。

(4) 设定误差(misspecification): 模型设定中遗漏了某个自相关的解释变量, 被纳入到扰动项中。

8.3 自相关的检验

1. 画图

可将 e_t 与 e_{t-1} 画成散点图。

也可画残差的“自相关图”(correlogram), 显示各阶样本自相关系数(命令 ac)或偏自相关系数(命令 pac)。此法虽直观, 不严格。

2. BG 检验

对于 $y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_K x_{tK} + \varepsilon_t$ ，假设存在一阶自相关，即 $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$ ，其中 u_t 为白噪声，并检验 $H_0: \rho = 0$ 。

由于可能存在高阶自相关，考虑 p 阶自回归：

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \cdots + \rho_p \varepsilon_{t-p} + u_t$$

检验 $H_0: \rho_1 = \cdots = \rho_p = 0$ 。由于 $\{\varepsilon_t\}$ 不可观测，故用 $\{e_t\}$ 替代，并引入所有解释变量，考虑辅助回归：

$$e_t \xrightarrow{\text{OLS}} x_{t1}, \dots, x_{tK}, e_{t-1}, \dots, e_{t-p} \quad (t = p+1, \dots, n)$$

由于使用 e_{t-p} ，损失 p 个样本值，故样本容量仅为 $(n-p)$ 。

使用 nR^2 形式的 LM 统计量：

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

如果 $(n-p)R^2$ 超过 $\chi^2(p)$ 的临界值，拒绝“无自相关”的原假设。
此检验被称为“Breusch-Godfrey 检验”，简称 BG 检验。

Davidson and MacKinnon(1993)建议:

把残差向量 \mathbf{e} 中因滞后而缺失的项, 用其期望值 $E(\mathbf{e}) = \mathbf{0}$ 来代替;

保持样本容量仍为 n , 使用统计量:

$$nR^2 \xrightarrow{d} \chi^2(p)$$

Davidson-MacKinnon 方法为 Stata 的默认设置。

3. Box-Pierce Q 检验

残差的各阶样本自相关系数：

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

如果 $H_0 : \rho_1 = \dots = \rho_p = 0$ 成立, 则 $\hat{\rho}_j \xrightarrow{p} 0$, $\sqrt{n}\hat{\rho}_j \xrightarrow{d}$ 正态分布, $j = 1, 2, \dots, p$ 。

残差的各阶样本自相关系数平方和的 n 倍, 就是 “Box-Pierce Q 统计量”：

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

经改进的“Ljung-Box Q 统计量”:

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

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, 为 Stata 所采用。

如何确定自相关阶数 p ? 如果 p 太小, 可能忽略高阶自相关的存在; 如果 p 较大, 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 $p = \min\{\text{floor}(n/2) - 2, 40\}$, 其中 $\text{floor}(n/2)$ 为不超过 $n/2$ 的最大整数。

4. DW 检验

“DW 检验” (Durbin and Watson, 1950)较早出现，已不常用。只能检验一阶自相关，且要求解释变量满足严格外生性。

DW 检验的统计量为

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

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, $\hat{\rho}_1 \approx 1$, 一阶正自相关;

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

DW 统计量依赖于数据矩阵 \mathbf{X} , 无法制表, 须使用上限分布 d_U 与下限分布 d_L ($d_L < d < d_U$) 来判断。得到 d_U 与 d_L 的临界值后, 仍存在无结论区域。

DW 统计量本质就是残差的一阶自相关系数, 不能指望它提供太多的信息。

8.4 自相关的处理

1. 使用“OLS + 异方差自相关稳健的标准误”

仍用 OLS 来估计回归系数，但使用“异方差自相关稳健的标准误”(Heteroskedasticity and Autocorrelation Consistent Standard Error, 简记 HAC)。

此法称为“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。

为什么第 5 章的“异方差稳健标准误”不适用于自相关的情形？问题出在假定 5.5，即 $\mathbf{g}_i \equiv \mathbf{x}_i \varepsilon_i$ 为鞅差分序列的假定。

命题 如果回归模型含有截距项, 则假定 5.5 意味着扰动项 ε_i 无自相关。

证明: 根据假定 5.5, \mathbf{g}_i 为鞅差分序列, 故

$$E(\mathbf{g}_i | \mathbf{g}_{i-1}, \dots, \mathbf{g}_1) = E(\mathbf{x}_i \varepsilon_i | \mathbf{g}_{i-1}, \dots, \mathbf{g}_1) = 0$$

因为模型含有截距项, 故向量 $\mathbf{g}_i \equiv \mathbf{x}_i \varepsilon_i$ 的第一个元素为 ε_i 。因此, $E(\varepsilon_i | \mathbf{g}_{i-1}, \dots, \mathbf{g}_1) = 0$ 。由于 $\{\varepsilon_{i-1}, \dots, \varepsilon_1\} \subset \{\mathbf{g}_{i-1}, \dots, \mathbf{g}_1\}$ (前者是后者的子集, 故前者的信息完全包含于后者之中), 根据迭代期望定律可得

$$\begin{aligned} E(\varepsilon_i | \varepsilon_{i-1}, \dots, \varepsilon_1) &= E_{\mathbf{g}_{i-1}, \dots, \mathbf{g}_1} \left[E(\varepsilon_i | \varepsilon_{i-1}, \dots, \varepsilon_1) | \mathbf{g}_{i-1}, \dots, \mathbf{g}_1 \right] \\ &= E_{\mathbf{g}_{i-1}, \dots, \mathbf{g}_1} \underbrace{\left[E(\varepsilon_i | \mathbf{g}_{i-1}, \dots, \mathbf{g}_1) \right]}_{=0} = 0 \end{aligned}$$

因此, ε_i 均值独立于 $(\varepsilon_{i-1}, \dots, \varepsilon_1)$, 故扰动项 ε_i 无自相关。

根据第 5 章, 异方差稳健的协方差矩阵 $\mathbf{S}_{XX}^{-1} \hat{\mathbf{S}} \mathbf{S}_{XX}^{-1}$ 为夹心估计量, 其中 $\mathbf{S}_{XX} \equiv \frac{1}{n} \mathbf{X}'\mathbf{X}$, $\hat{\mathbf{S}} \equiv \frac{1}{n} \sum_{i=1}^n e_i^2 \mathbf{x}_i \mathbf{x}_i'$ 。

异方差自相关稳健的协方差矩阵也是夹心估计量, 其形式为 $\mathbf{S}_{XX}^{-1} \hat{\mathbf{Q}} \mathbf{S}_{XX}^{-1}$, 其中

$$\hat{\mathbf{Q}} = \hat{\mathbf{S}} + \frac{1}{n} \sum_{j=1}^p \sum_{t=j+1}^n \left(1 - \frac{j}{p+1} \right) e_t e_{t-j} (\mathbf{x}_t \mathbf{x}_{t-j}' + \mathbf{x}_{t-j} \mathbf{x}_t')$$

p 为自相关的阶数，也称“截断参数” (truncation parameter)。
建议令 $p = n^{1/4}$ 或 $p = 0.75n^{1/3}$ ，再取整数。

考虑一元回归情形，

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$$

OLS 估计量为，

$$\hat{\beta}_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) [\beta_1 (x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})]}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

其中，由于 $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{\varepsilon}$ ，故 $y_i - \bar{y} = \beta_1(x_i - \bar{x}) + \varepsilon_i - \bar{\varepsilon}$ 。因此，

$$\hat{\beta}_1 - \beta_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})\varepsilon_i}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

其中， $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})\varepsilon_i - \frac{1}{n} \bar{\varepsilon} \underbrace{\sum_{i=1}^n (x_i - \bar{x})}_{=0}$ 。记

$v_i \equiv (x_i - \bar{x})\varepsilon_i$ ，在大样本中， $\hat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{n} \sum_{i=1}^n v_i}{\sigma_x^2}$ ，其中 σ_x^2 为 x_i 的方差。

故在大样本中，

$$\text{Var}(\hat{\beta}_1) = \frac{\text{Var}\left(\frac{1}{n} \sum_{i=1}^n v_i\right)}{(\sigma_x^2)^2}$$

考虑 $n=2$ 的最简单情形，则上式分子为，

$$\begin{aligned}\text{Var}\left(\frac{1}{n} \sum_{i=1}^n v_i\right) &= \text{Var}\left[\frac{1}{2}(v_1 + v_2)\right] = \frac{1}{4}[\text{Var}(v_1) + \text{Var}(v_2) + 2\text{Cov}(v_1, v_2)] \\ &= \frac{1}{2}\sigma_v^2 + \frac{1}{2}\rho_1\sigma_v^2 = \frac{1}{2}\sigma_v^2(1 + \rho_1) \equiv \frac{1}{2}\sigma_v^2 f_2\end{aligned}$$

其中， $\sigma_v^2 \equiv \text{Var}(v_i)$ ， $\rho_1 \equiv \text{corr}(v_1, v_2)$ 为一阶自相关系数，而

$f_2 \equiv (1 + \rho_1)$ 是修正系数。

如不存在自相关, $\rho_1=0$, $f_2=1$, 则 $\text{Var}\left(\frac{1}{n}\sum_{i=1}^n v_i\right) = \frac{1}{2}\sigma_v^2$, 得到通常的方差公式。

如存在自相关, $\rho_1 \neq 0$, 方差公式有所不同。

考虑样本容量为 n 的一般情况, 则 $\text{Var}\left(\frac{1}{n}\sum_{i=1}^n v_i\right) = \frac{1}{n}\sigma_v^2 f_n$, 其中 $f_n \equiv 1 + 2\sum_{j=1}^{n-1}\left(\frac{n-j}{n}\right)\rho_j$ 为对应于样本容量为 n 的修正系数, 而 ρ_j 为 j 阶自相关系数; 因此,

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma_v^2}{n(\sigma_x^2)^2} \cdot f_n$$

上式是普通方差公式的 f_n 倍。 f_n 包含未知的自相关系数 ρ_j ，需对其进行估计，比如

$$\hat{f}_n \equiv 1 + 2 \sum_{j=1}^{n-1} \left(\frac{n-j}{n} \right) \hat{\rho}_j$$

其中， $\hat{\rho}_j$ 为 j 阶样本自相关系数。但待估计参数 $(\rho_1, \dots, \rho_{n-1})$ 太多，且随样本容量 n 增长，导致此估计量不一致。

反之，仅考虑前几阶自相关系数(比如，只考虑 ρ_1)的估计量也不一致，因为忽略了高阶自相关。

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 随着样本容量的增长而增长。

一般建议取 $p = n^{1/4}$ 或 $p = 0.75n^{1/3}$ ，作为截断参数。

实践中，建议使用不同的截断参数，考察 HAC 标准误是否对于截断参数的取值敏感。

2. 使用“OLS + 聚类稳健的标准误”

如果样本观测值可以分为不同的“聚类”(clusters)，在同一聚类里的观测值互相相关，而不同聚类之间的观测值不相关，这种样本称为“聚类样本”(cluster sample)。

【例】在 Nerlove(1963)对美国电力企业的研究中，同一个州的电力企业可能受到相同州政策的影响而自相关，但不同州之间的电力企业可能不相关。此时，“州”(state)被称为“聚类变量”(cluster variable)。

【例】如果以全班同学为样本，则聚类变量可能是宿舍或专业。

如果将观测值按聚类的归属顺序排列，则扰动项的协方差矩阵为“块对角”(block diagonal)。

仍可用 OLS 来估计系数，但需使用“聚类稳健的标准误”(cluster robust standard errors)。

假设样本容量为 N ，包括 M 个聚类，其中第 j 个聚类包含 M_j 个个体。记第 j 个聚类个体 i 的解释变量为 \mathbf{x}_{ij} ，残差为 e_{ij} ，然后定义 $\mathbf{u}_j \equiv \sum_{i=1}^{M_j} e_{ij} \mathbf{x}_{ij}$ ，则聚类稳健的协方差矩阵可以写为

$$\frac{N-1}{N-K} \frac{M}{M-1} (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{j=1}^M \mathbf{u}_j' \mathbf{u}_j \right) (\mathbf{X}'\mathbf{X})^{-1}$$

其中, $\frac{N-1}{N-K} \frac{M}{M-1}$ 为对自由度的调整。

聚类稳健的标准误也是夹心估计量。在推导过程中并未假定同方差, 故也是异方差稳健的。

使用聚类稳健标准误的前提是, 聚类中的个体数 M_j 较少, 而聚类数很多 ($M \rightarrow \infty$); 则聚类稳健标准误是真实标准误的一致估计。

处理面板数据时, 常使用聚类稳健的标准误。

3. 使用可行广义最小二乘法(FGLS)

首先估计 $\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X})$ 。为减少待估参数，假设扰动项为 AR(1):

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t, \quad |\rho| < 1, \quad u_t \text{ 为白噪声}$$

记扰动项的 j 阶协方差 $\rho_j \equiv \text{Cov}(\varepsilon_t, \varepsilon_{t-j} | \mathbf{X})$ ，则

$$\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X}) = \begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{n-2} \\ \vdots & \vdots & & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & \rho_0 \end{pmatrix}$$

容易证明, $\rho_0 = \sigma^2 = \text{Var}(\varepsilon_t) = \frac{\sigma_u^2}{1-\rho^2}$, 其中 $\sigma_u^2 \equiv \text{Var}(u_t)$ 。

$\rho_1 = \rho\sigma^2$, 故 $\frac{\rho_1}{\rho_0} = \frac{\rho\sigma^2}{\sigma^2} = \rho$ 为一阶自相关系数; $\rho_2 = \rho^2\sigma^2, \dots,$
 $\rho_{n-1} = \rho^{n-1}\sigma^2$, 故

$$\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X}) = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix} \equiv \sigma^2 \mathbf{V}$$

只要估计唯一的参数 ρ , 就可使用 FGLS。Stata 默认的估计方

法为使用 OLS 对残差进行辅助回归, $e_t = \hat{\rho}e_{t-1} + error_t$ 。也可通过

残差一阶自相关系数 $\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}$, 或 $\hat{\rho} = 1 - \frac{DW}{2}$ 来估计 ρ 。

并将 V 的逆矩阵分解为 $V^{-1} = C'C$ 。可以证明

$$C = \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \dots & 0 & 0 \\ -\rho & 1 & \dots & 0 & 0 \\ 0 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

以 $\sqrt{1-\rho^2}\mathbf{C}$ 左乘原模型，并定义 $\tilde{\mathbf{y}} \equiv \sqrt{1-\rho^2}\mathbf{C}\mathbf{y}$, $\tilde{\mathbf{X}} \equiv \sqrt{1-\rho^2}\mathbf{C}\mathbf{X}$, $\tilde{\boldsymbol{\varepsilon}} \equiv \sqrt{1-\rho^2}\mathbf{C}\boldsymbol{\varepsilon}$, 则变换后的扰动项 $\tilde{\boldsymbol{\varepsilon}}$ 满足球型扰动项的假设，故高斯-马尔可夫定理成立(此变换是 GLS 的特例):

$$\tilde{\mathbf{y}} = \sqrt{1-\rho^2}\mathbf{C}\mathbf{y} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \dots & 0 & 0 \\ -\rho & 1 & \dots & 0 & 0 \\ 0 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_n - \rho y_{n-1} \end{pmatrix}$$

$$\tilde{\mathbf{X}} = \sqrt{1-\rho^2} \mathbf{C} \mathbf{X} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \dots & 0 & 0 \\ -\rho & 1 & \dots & 0 & 0 \\ 0 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} x_{11} & \dots & x_{1K} \\ x_{21} & \dots & x_{2K} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nK} \end{pmatrix}$$

$$= \begin{pmatrix} \sqrt{1-\rho^2} x_{11} & \dots & \sqrt{1-\rho^2} x_{1K} \\ x_{21} - \rho x_{11} & \dots & x_{2K} - \rho x_{1K} \\ \vdots & & \vdots \\ x_{n1} - \rho x_{n-1,1} & \dots & x_{nK} - \rho x_{n-1,K} \end{pmatrix}$$

$$\tilde{\boldsymbol{\varepsilon}} = \sqrt{1-\rho^2} \mathbf{C} \boldsymbol{\varepsilon} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \dots & 0 & 0 \\ -\rho & 1 & \dots & 0 & 0 \\ 0 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2} \varepsilon_1 \\ \varepsilon_2 - \rho \varepsilon_1 \\ \vdots \\ \varepsilon_n - \rho \varepsilon_{n-1} \end{pmatrix}$$

写出每个观测值(个体)的回归方程:

$$\sqrt{1-\rho^2} y_1 = \sqrt{1-\rho^2} \beta_1 + \sqrt{1-\rho^2} \beta_2 x_{12} + \dots + \sqrt{1-\rho^2} \beta_K x_{1K} + \tilde{\varepsilon}_1$$

$$y_2 - \rho y_1 = (1-\rho) \beta_1 + \beta_2 (x_{22} - \rho x_{12}) + \dots + \beta_K (x_{2K} - \rho x_{1K}) + \tilde{\varepsilon}_2$$

.....

$$y_n - \rho y_{n-1} = (1-\rho) \beta_1 + \beta_2 (x_{n2} - \rho x_{n-1,2}) + \dots + \beta_K (x_{nK} - \rho x_{n-1,K}) + \tilde{\varepsilon}_n$$

第一个方程的形式与其他方程不同。用 OLS 估计变换后的模型，即为“Prais-Winsten 估计法”（简记 PW）。

为计算方便，将第一个方程删去，称为“Cochrane-Orcutt 估计法”（简记 CO）。该法有更简洁的推导过程。原模型为，

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t$$

将上式滞后一期，然后方程两边同时乘以 ρ 得

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

将两方程相减可得：

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

新扰动项 $\varepsilon_t - \rho\varepsilon_{t-1} = u_t$ 满足球型扰动项的古典假定。

此法也称“准差分法”(quasi differences)。

在操作中，常使用迭代法，首先用 OLS 估计原模型，作辅助回归得到 $\hat{\rho}^{(1)}$ (对 ρ 的第一轮估计)，再用 $\hat{\rho}^{(1)}$ 进行 FGLS 估计，使用新的残差估计 $\hat{\rho}^{(2)}$ (对 ρ 的第二轮估计)，再用 $\hat{\rho}^{(2)}$ 进行 FGLS 估计，……，直至收敛。

使用 FGLS 处理自相关，如果对自相关系数的估计较准确，且满足严格外生性的假定，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的假定，则

FGLS 可能不一致，尽管 OLS 依然一致。

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 $(x_{tk} - \rho x_{t-1,k})$ ，二者可能存在相关性，导致不一致估计。

总之，FGLS 不如 OLS 稳健。

4. 修改模型设定

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

$$y_t = \rho y_{t-1} + \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t$$

由于 y_t 是 y_{t-1} 的函数，故 $\{y_t\}$ 存在自相关。假设这个模型被错误地估计成

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \underbrace{[\rho y_{t-1} + \varepsilon_t]}_{=v_t}$$

ρy_{t-1} 被纳入到扰动项 v_t 中，导致扰动项 $\{v_t\}$ 自相关。

时间序列的自相关，有时可通过引入被解释变量的滞后值来消除。由于模型设定误差而导致的自相关，最好改进模型。