stationary: strict: $\forall \{t_1, \cdots, t_n\}$, $\{x_{t_1}, \cdots, x_{t_n}\}$ 的分布和 $\{x_{t_1+k}, \cdots, x_{t_n+k}\}$ 分布一样

weak: $E[x_t]$ 与 t 无关, $cov(x_t, x_{t+k})$ 与 t 无关.

white noise: 1. weak stationary 2. $E[x_t]=0$ 3. $cov(x_t, x_{t+k})=0$

第 20 章　平稳时间序列 四 描述性

ergodic: $E[fg] = E[f]\,E[g]$, 独立性 (足够远)

根据时间序列的随机过程特性，可分为“平稳序列”(stationary)与“非平稳序列”(non-stationary)两大类，需使用不同的计量方法。

20.1　时间序列的数字特征

记随机变量 y 的观测值为 $\{y_1, y_2, \cdots, y_T\}$，并假设为(严格)平稳过程。

故该序列的期望、方差等数字特征不随时间而变。

期望$\mu \equiv \mathrm{E}(y)$反映序列的平均水平，用$\bar{y} \equiv \frac{1}{T}\sum_{t=1}^{T} y_t$来估计。

方差$\mathrm{Var}(y)$反映序列的波动幅度，用$\frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2$来估计。

定义 时间序列$\{y_{\mathrm{t}}\}$的 k **阶自协方差**(autocovariance of order k)

$$\gamma_k \equiv \mathrm{Cov}(y_t, y_{t+k}) = \mathrm{E}\left[(y_t - \mu)(y_{t+k} - \mu)\right]$$

它反映同一变量(y)相隔 k 期之间的自相关程度。对γ_k的估计值为“样本自协方差”:

$$\hat{\gamma}_k \equiv \frac{1}{T-k}\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})$$

定义 时间序列$\{y_t\}$的 k **阶自相关系数**(autocorrelation of order k)

$$\rho_k \equiv \mathrm{Corr}(y_t, y_{t+k}) \equiv \frac{\mathrm{Cov}(y_t, y_{t+k})}{\mathrm{Var}(y_t)}$$

对于平稳过程，ρ_k不依赖于时间，仅是滞后阶数 k 的函数，故称为“自相关函数”(Autocorrelation Function，简记 ACF)。

将(k, ρ_k)画成图，即为“自相关图”(correlogram)。由于$\rho_k = \rho_{-k}$，故一般只画自相关图的正半边。

对ρ_k的估计值为“样本自相关系数”:

$$\hat{\rho}_k \equiv \hat{\gamma}_k / \hat{\gamma}_0$$

y_t与y_{t+k}之间的相关性可能由二者之间的变量$\{y_{t+1},\cdots,y_{t+k-1}\}$引起。

定义 时间序列$\{y_t\}$的 k **阶偏自相关系数**(partial autocorrelation of order k)为

$$\rho_k^* \equiv \mathrm{Corr}(y_t, y_{t+k} \mid y_{t+1},\cdots,y_{t+k-1})$$

即给定$\{y_{t+1},\cdots,y_{t+k-1}\}$条件下，$y_t$与$y_{t+k}$的条件相关系数。

ρ_k^*只是 k 的函数，称为“偏自相关函数”(Partial Autocorrelation Function，简记 PACF)。

对以下 k 阶自回归方程进行 OLS 估计：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_k y_{t-k} + \varepsilon_t$$ PACF

则 $\hat{\beta}_k$ 就是“k 阶样本偏自相关系数” $\hat{\rho}_k^*$。

20.2 自回归模型

对于样本数据 $\{y_1, y_2, \cdots, y_T\}$，最简单的预测方法为 AR(1)：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \quad (t = 2, \cdots, T)$$

ε_t 为白噪声， $\mathrm{E}(\varepsilon_t) = 0$， $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$， 且无自相关。

假设 $|\beta_1| < 1$， 则 $\{y_t\}$ 为渐近独立的平稳过程。 ergodic stationary.

由于 y_{t-1} 依赖于 $\{\varepsilon_{t-1}, \cdots, \varepsilon_1\}$，而 ε_t 与 $\{\varepsilon_{t-1}, \cdots, \varepsilon_1\}$ 不相关，故 y_{t-1} 与 ε_t 不相关，OLS 一致。

但使用 OLS 将损失一个样本容量。为提高估计效率，考虑 MLE。

假设$\{\varepsilon_t\}$为 iid，服从$N(0, \sigma_\varepsilon^2)$。对方程两边取期望

$$\mathrm{E}(y) = \beta_0 + \beta_1 \mathrm{E}(y)$$

故$\{y_t\}$的无条件期望为$\dfrac{\beta_0}{1-\beta_1}$。对方程两边取方差可得

$$\mathrm{Var}(y) = \beta_1^2 \mathrm{Var}(y) + \sigma_\varepsilon^2$$

故$\{y_t\}$的无条件方差为$\dfrac{\sigma_\varepsilon^2}{1-\beta_1^2}$。故

$$y_1 \sim N\left(\frac{\beta_0}{1-\beta_1}, \frac{\sigma_\varepsilon^2}{1-\beta_1^2}\right)$$

y_1的其(无条件)密度函数为

$$f_{y_1}(y_1) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2/(1-\beta_1^2)}} \exp\left\{\frac{-\left[y_1 - \left(\beta_0/(1-\beta_1)\right)\right]^2}{2\sigma_\varepsilon^2/(1-\beta_1^2)}\right\}$$

在给定y_1的条件下，y_2的条件分布为

$$y_2 | y_1 \sim N(\beta_0 + \beta_1 y_1, \sigma_\varepsilon^2)$$

其条件密度为

$$f_{y_2|y_1}(y_2|y_1) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{ \frac{-(y_2 - \beta_0 - \beta_1 y_1)^2}{2\sigma_\varepsilon^2} \right\}$$

y_1与y_2的联合分布密度为

$$f_{y_1, y_2}(y_1, y_2) = f_{y_1}(y_1) f_{y_2|y_1}(y_2 | y_1)$$

$f_{y_3|y_2} \cdot f_{y_4|y_3}$

依次类推， $y_t|y_{t-1} \sim N(\beta_0 + \beta_1 y_{t-1}, \sigma_\varepsilon^2)$, $t = 2, \cdots, T$ 。

整个样本数据$\{y_1, y_2, \cdots, y_T\}$的联合概率密度为

$$f_{y_1, \cdots, y_T}(y_1, \cdots, y_T) = f_{y_1}(y_1) \prod_{t=2}^{T} f_{y_t|y_{t-1}}(y_t | y_{t-1})$$

对数似然函数为

$$\ln L(\beta_0, \beta_1, \sigma_\varepsilon^2; y_1, \cdots, y_T) = \ln f_{y_1}(y_1) + \sum_{t=2}^{T} \ln f_{y_t|y_{t-1}}(y_t \mid y_{t-1})$$

代入 $f_{y_1}(y_1)$ 与 $f_{y_t|y_{t-1}}(y_t \mid y_{t-1})$ 的表达式可得

$$\ln L = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\left[\sigma_\varepsilon^2 / (1-\beta_1^2)\right] - \frac{\left[y_1 - \left(\beta_0/(1-\beta_1)\right)\right]^2}{2\sigma_\varepsilon^2 / (1-\beta_1^2)}$$

$$-\frac{T-1}{2}\ln 2\pi - \frac{T-1}{2}\ln \sigma_\varepsilon^2 - \sum_{t=2}^{T} \frac{(y_t - \beta_0 - \beta_1 y_{t-1})^2}{2\sigma_\varepsilon^2}$$

此估计量称为“精确最大似然估计量”(Exact MLE)。

没有解析解.

如果 T 较大，第一个观测值对似然函数的贡献较小，可忽略。考虑在y_1给定的情况下，$\{y_2, \cdots, y_T\}$的条件分布，则对数似然函数简化为

$$\ln L = -\frac{T-1}{2}\ln 2\pi - \frac{T-1}{2}\ln \sigma_\varepsilon^2 - \sum_{t=2}^{T}\frac{(y_t - \beta_0 - \beta_1 y_{t-1})^2}{2\sigma_\varepsilon^2}$$

此估计量称为“条件最大似然估计量”(conditional MLE)。

为使上式最大化，须让右边第三项的分子最小化，

$$\min_{\{\beta_0,\,\beta_1\}} \sum_{t=2}^{T}(y_t - \beta_0 - \beta_1 y_{t-1})^2$$

其结果与 OLS 一样。

更一般地，考虑 AR(p)：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t$$

对 AR(p)的估计与 AR(1)类似。

在使用“条件 MLE”时，考虑在给定$\{y_1, \cdots, y_p\}$的情况下，$\{y_{p+1}, \cdots, y_T\}$的条件分布。

由于条件 MLE 等价于 OLS，而后者并不依赖于正态性假定，故条件 MLE 的一致性也不依赖于正态性假定。

如何估计$\hat{p}$？

方法一：由大到小的序贯 t 规则(general-to-specific sequential t rule)。设最大滞后期$p_{\max}$，令$\hat{p} = p_{\max}$进行估计，并对最后一个滞后期系数的显著性进行 t 检验。如果接受该系数为 0，则令$\hat{p} = p_{\max} - 1$，重新进行估计，再对最后一个滞后期的系数进行 t 检验，如果显著，则停止；否则，令$\hat{p} = p_{\max} - 2$；以此类推。

p不要选太大. GARCH(1,1).

方法二：使用信息准则，选择$\hat{p}$使得 AIC，BIC 或 HQIC 最小化，分别记为$\hat{p}_{\text{AIC}}$，$\hat{p}_{\text{BIC}}$与$\hat{p}_{\text{HQIC}}$。比如，

$$\min_{p} \text{AIC} \equiv \ln\left(\boldsymbol{e}'\boldsymbol{e}/T\right) + \frac{2}{T}(p+1)$$

其中，$\boldsymbol{e}'\boldsymbol{e}$为残差平方和。

20.3　移动平均模型

另一类时间序列模型为“移动平均过程”(Moving Average Process，简记 MA)。一阶移动平均过程为 MA(1)：

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

$\{\varepsilon_t\}$为白噪声，ε_t的系数被标准化为 1。

使用条件 MLE 估计。

假设$\{\varepsilon_t\}$为 iid，且服从$N(0, \sigma_\varepsilon^2)$。如果已知$\varepsilon_{t-1}$，则

$$y_t \big| \varepsilon_{t-1} \sim N(\mu + \theta\varepsilon_{t-1}, \sigma_\varepsilon^2)$$

假设$\varepsilon_0 = 0$，则可知$\varepsilon_1 = y_1 - \mu$。

给定ε_1，则可知$\varepsilon_2 = y_2 - \mu - \theta\varepsilon_1$。

以此类推，使用公式"$\varepsilon_t = y_t - \mu - \theta\varepsilon_{t-1}$"计算出全部$\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T\}$。

给定$\varepsilon_0 = 0$的条件，可写下样本数据$\{y_1, y_2, \cdots, y_T\}$的似然函数。

q 阶移动平均过程，记为 MA(q)：

$$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

也可以进行条件 MLE 估计，即在给定"$\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{-q+1} = 0$"的

条件下，最大化样本数据的似然函数。

20.4 ARMA

将 AR(p)与 MA(q)结合起来，得到 ARMA(p, q)：

$$\underbrace{y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p}} + \underbrace{\varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}}$$

$\{\varepsilon_t\}$为白噪声。

给定$\{y_1, \cdots, y_p\}$与“$\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{-q+1} = 0$”, 可进行条件 MLE 估计。

如何用数据来估计$(\hat{p}, \hat{q})$？

MA(q) ACF

PACF $y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$

L：滞后乘子. $y_t = \mu + (1+\theta L)\varepsilon_t$

$(1+\theta L)^{-1}(y_t - \mu) = \varepsilon_t$

y_t y_{t-1} …… y_{t-j}

ε_t

ε_{t-1} ε_{t-1}

ε_{t-q} ε_{t-q}

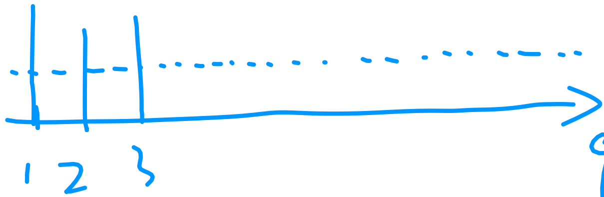
ε_{t-q-1}

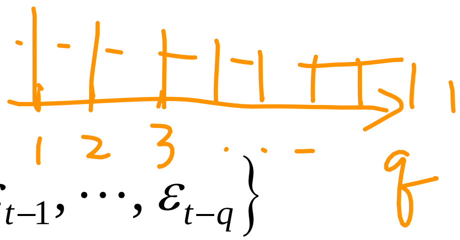
ε_{t-j}

ε_{t-j-1}

先考察数据的自相关函数(ACF)与偏自相关函数(PACF)，以判断是否存在 $p=0$ 或 $q=0$ 的情形。

如果 $p=0$，则为 MA(q)：





$$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

如果 $j>q$，则 $\mathrm{Cov}(y_t, y_{t-j}) = 0$，因为产生 y_t 的扰动项 $\{\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q}\}$ 与产生 y_{t-j} 的扰动项 $\{\varepsilon_{t-j}, \varepsilon_{t-j-1}, \cdots, \varepsilon_{t-j-q}\}$ 无交集。

对于 MA(q)模型，ACF 函数在 $j>q$ 时都等于零，出现"截尾"。

另一方面，AR(1)的 ACF 函数呈指数衰减，称为"拖尾" (tails off to zero)，不存在截尾。

AR(1)： $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$.

ACF

$= \beta_0 + \beta_1(\beta_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$

$\frac{\gamma(1)}{\gamma(0)} = \frac{\mathrm{cov}(y_t, y_{t-1})}{\mathrm{Var}(y_t)} = \beta_1$

$\frac{\gamma(2)}{\gamma(0)} = \frac{\mathrm{cov}(y_t, y_{t-2})}{\mathrm{Var}(y_t)} = \beta_1^2$

PACF $y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t$

如果$q=0$，则 ARMA(p, q)简化为 AR(p)模型：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t$$

假设真实模型为 AR(p)，却用 OLS 来估计 AR(p+1)，即 $y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \beta_{p+1} y_{t-p-1} + \varepsilon_t$，则 $\operatorname*{plim}_{T\to\infty} \hat{\beta}_{p+1} = 0$，因为 $\beta_{p+1} = 0$。而 $\hat{\beta}_{p+1}$ 正是对 $(p+1)$ 阶偏自相关函数的估计。

对于 AR(p)模型，PACF 函数在 $j > p$ 时都等于零，即出现截尾。

另一方面，MA(q)模型的 PACF 函数逐渐衰减，拖尾，不存在截尾。

总之，对于 AR(p)模型，其 ACF 函数拖尾，而 PACF 函数截尾。如出现这种情形，可判断为 AR(p)。

对于 MA(q)模型，其 ACF 函数截尾，而 PACF 函数拖尾。如出现这种情形，可判断为 MA(q)。

如果以上两种情形均不符合，则考虑一般的 ARMA(p, q)模型，其中 p, q 均不为零。

Box, Jenkins and Reinsel (1994)认为，对大多数情况，$p \le 2$ 与 $q \le 2$ 就足够了。可让 $p_{\max}$ 与 $q_{\max}$ 更大些，使用信息准则或序贯 t 规则。

在估计模型之后，需进行诊断性分析，以确定 ARMA(p, q)模型的假定是否成立。

最重要的假定是，扰动项$\{\varepsilon_t\}$为白噪声。

如果模型过小，即 $\hat{p} < p$ 或 $\hat{q} < q$，则相当于遗漏解释变量，导致扰动项出现自相关，不再是白噪声。

可使用 Q 检验来检验模型的残差是否存在自相关。

如果残差存在自相关，应考虑使模型更大些，重新对模型进行估计，再检验新模型的残差是否为白噪声，如此反复。

20.5 自回归分布滞后模型

在自回归模型中，可引入其他变量构成“自回归分布滞后模型”(Autoregressive Distributed Lag Model，简记 ADL(p, q))：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \gamma_1 x_{t-1} + \cdots + \gamma_q x_{t-q} + \varepsilon_t$$

例 记通货膨胀率为π_t，失业率为un_t，预测模型为

$$\Delta\pi_t = \beta_0 + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_3\Delta\pi_{t-3} + \beta_4\Delta\pi_{t-4} + \gamma_1 un_{t-1} + \varepsilon_t$$

此 ADL(4,1)为经验菲利普斯曲线(empirical Phillip Curve)。

如果$\gamma_1 < 0$，则失业率越低，物价越有上涨的压力；而通胀的调整也受到过去通胀变化的滞后作用。

可引入更多解释变量。比如，共有 K 个解释变量$\{x_{1t}, \cdots, x_{Kt}\}$，其中第 j 个解释变量x_{jt}共有q_j个滞后值被包括在模型中，$j = 1, \cdots, K$：

$$
\begin{aligned}
y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \gamma_{11} x_{1,\,t-1} + \cdots + \gamma_{1q_1} x_{1,\,t-q_1} + \cdots \\
+ \gamma_{K1} x_{K,\,t-1} + \cdots + \gamma_{Kq_K} x_{K,\,t-q_K} + \varepsilon_t
\end{aligned}
$$

如果自回归分布滞后模型满足以下假定，则可用 OLS 来估计。

(i) $\mathrm{E}(\varepsilon_t \mid y_{t-1}, y_{t-2}, \cdots, x_{1,\,t-1}, x_{1,\,t-2}, \cdots, x_{K,\,t-1}, x_{K,\,t-2}, \cdots) = 0$。此假定类似于严格外生性假设，意味着扰动项$\varepsilon_t$与所有解释变量的整个历史全部无关。

这保证了对滞后期$(p, q_1, \cdots, q_K)$的设定是正确的。如果滞后期设定不正确，比如，真实模型还应该包括$y_{t-(p+1)}$，但$\beta_{p+1} y_{t-(p+1)}$却被纳入ε_t中，则ε_t便与解释变量相关，导致 OLS 不一致。

(ii) $\{y_t, x_{1t}, \cdots, x_{Kt}\}$为渐近独立的平稳序列。

(iii) $\{y_t, x_{1t}, \cdots, x_{Kt}\}$有非零的有限四阶矩。

(iv) 解释变量无完全多重共线性。

对滞后期数的选择可使用信息准则，或使用 t, F 检验来检验最后一期系数的显著性。

更一般地，可以在 ARMA 模型中引入其他变量，称为“ARMAX”模型。

20.6 ARMA 模型的 Stata 命令及实例

20.7 误差修正模型

从经济理论而言，变量之间可能存在长期均衡关系，而变量的短期变动为向这个长期均衡关系的部分调整。

“误差修正模型” (Error Correction Model，简记 ECM)正是这一思想在计量的体现。

均衡

考虑 AR(1)：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$$

$|\beta_1| < 1$，故$\{y_t\}$为平稳过程。

对方程两边求期望，令长期均衡值$y^* = \mathrm{E}(y_t) = \mathrm{E}(y_{t-1})$，可得

$$y^* = \frac{\beta_0}{1-\beta_1}$$

将$\beta_0 = (1-\beta_1)y^*$代入原方程，并在方程两边减去y_{t-1}可得

$$\Delta y_t = (1-\beta_1)y^* - (1-\beta_1)y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \underbrace{(1-\beta_1)(y^* - y_{t-1})}_{\text{error correction}} + \varepsilon_t$$

AR(1)的误差修正模型，将Δy_t表达为对长期均衡的偏离$(y^* - y_{t-1})$的部分调整(即误差修正)，加上扰动项。

考虑 ADL 模型：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + \varepsilon_t$$

$|\beta_1| < 1$。假设经济理论认为(y, x)之间存在长期均衡关系$y = \phi + \theta x$。

方程两边求期望，令$y^* = \mathrm{E}(y_t) = \mathrm{E}(y_{t-1})$，$x^* = \mathrm{E}(x_t) = \mathrm{E}(x_{t-1})$，可得

$$y^* = \beta_0 + \beta_1 y^* + \gamma_0 x^* + \gamma_1 x^*$$

$$(1 - \beta_1) y^* = \beta_0 + (\gamma_0 + \gamma_1) x^*$$

$$y^* = \frac{\beta_0}{(1-\beta_1)} + \frac{(\gamma_0+\gamma_1)}{(1-\beta_1)} x^*$$

故$\phi = \dfrac{\beta_0}{1-\beta_1}$，$\theta = \dfrac{\gamma_0+\gamma_1}{1-\beta_1}$。

θ称为“长期乘数”(long-run multiplier)，衡量当 x 永久性地变化一单位时，导致 y 的永久性变化幅度。

$\beta_0 = (1-\beta_1)\phi$，$\gamma_0+\gamma_1 = (1-\beta_1)\theta$。

在方程两边减去y_{t-1}，并在方程右边加上、减去$\gamma_0 x_{t-1}$：

$$\Delta y_t = \beta_0 - (1-\beta_1) y_{t-1} + \gamma_0 (x_t - x_{t-1}) + (\gamma_0+\gamma_1) x_{t-1} + \varepsilon_t$$

代入$\beta_0=(1-\beta_1)\phi$，可得

$$\Delta y_t=(1-\beta_1)\phi-(1-\beta_1)y_{t-1}+\gamma_0\Delta x_t+(\gamma_0+\gamma_1)x_{t-1}+\varepsilon_t$$

代入$\gamma_0+\gamma_1=(1-\beta_1)\theta$，则有

$$\Delta y_t=\gamma_0\Delta x_t+\underbrace{(\beta_1-1)(y_{t-1}-\phi-\theta x_{t-1})}_{\text{error correction}}+\varepsilon_t$$

$(\beta_1-1)(y_{t-1}-\phi-\theta x_{t-1})$称为“误差修正项”(error correction term)。$\{\phi,\theta\}$为长期参数，$\{\gamma_0,\beta_1-1\}$为短期参数。

ECM 的形式看似非线性，但仍是线性回归，可把它还原为 ADL 模型，用 OLS 来估计。

20.8　MA(∞)与滞后算子

MA(q)的自相关函数(ACF)存在截尾。如果希望自相关系数永远不为 0，可考虑MA(∞)：

$$y_t = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$$

其中，$\theta_0 = 1$。

无穷多个随机变量之和，能收敛到某随机变量吗？

一个充分条件是，序列$\{\theta_j\}_{j=0}^{\infty}$为“绝对值可加总”(Absolutely Summable，简记 AS)，即$\sum_{j=0}^{\infty} \left|\theta_j\right| < \infty$(有限)。

在 AS 的条件下，不仅 MA(∞)有定义，而且是弱平稳过程；如果$\{\varepsilon_t\}$为 iid，则 MA(∞)严格平稳。

时间序列分析常引入“滞后算子”(lag operator)。

定义 $Ly_t = y_{t-1}$， $L^2 y_t = L(Ly_t) = y_{t-2}$， …， $L^p y_t = y_{t-p}$。特别地，$L^0 y_t = 1\cdot y_t = y_t$。

滞后算子的运算相当于幂函数。比如，$L^p \cdot L^q = L^{p+q}$。

由于$\Delta y_t = y_t - y_{t-1} = (1-L)y_t$，故差分算子$\Delta \equiv 1-L$。

对于 AR(*p*)，$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t$，可用滞后算子表示：

$$y_t = \beta_0 + \beta_1 L y_t + \cdots + \beta_p L^p y_t + \varepsilon_t$$

移项可得

$$y_t - \beta_1 L y_t - \cdots - \beta_p L^p y_t = \beta_0 + \varepsilon_t$$

向右提取公因子y_t

$$(1 - \beta_1 L - \cdots - \beta_p L^p) y_t = \beta_0 + \varepsilon_t$$

定义“滞后多项式”(lag polynomial) $\beta(L) \equiv 1 - \beta_1 L - \cdots - \beta_p L^p$，则

$$\beta(L) y_t = \beta_0 + \varepsilon_t$$

如果存在$\beta(L)^{-1}$，则可以在两边左乘$\beta(L)^{-1}$，得到 AR(p)的“解”。

定义 对于任意一个实数序列$(\alpha_0, \alpha_1, \alpha_2, \cdots)$，定义其对应的**滤波**(filter)为$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$。

命题 假设$\{x_t\}$为弱平稳过程，序列$\{\theta_j\}_{j=0}^{\infty}$为绝对值可加总(AS)，则$y_t = \alpha + \sum_{j=0}^{\infty} \theta_j x_{t-j} = \alpha + \sum_{j=0}^{\infty} \theta_j L^j x_t$有定义，且为弱平稳。进一步，如果$\{x_t\}$为 iid，则$\{y_t\}$严格平稳。

定义 对于两个滤波“$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$”与“$\beta(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots$”，定义其乘积为

已知

$$\delta(L) \equiv \alpha(L)\beta(L) \equiv (\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots)(\beta_0 + \beta_1 L + \beta_2 L^2 + \cdots)$$

$$= \alpha_0\beta_0 + \underbrace{(\alpha_0\beta_1 + \alpha_1\underbrace{\beta_0}_{\frac{1}{\alpha_0}})}_{=0} L + \underbrace{(\alpha_2\beta_0 + \alpha_1\beta_1 + \alpha_0\beta_2)}_{=0} L^2 + \cdots = 1$$

$\beta_0 = \frac{1}{\alpha_0}$

$\alpha_0\beta_1 + \alpha_1 \cdot \frac{1}{\alpha_0} = 0$

$\Rightarrow \beta_1 = -\frac{\alpha_1}{\alpha_0^2}$

滤波的乘积满足交换律。

最感兴趣的情形是$\delta(L)=1$，即令上式的常数项$\alpha_0\beta_0$为 1，而其余各项的系数均为 0。

称$\beta(L)$为$\alpha(L)$的“逆”(inverse)，并记为$\alpha(L)^{-1}$。

只要$\alpha_0 \neq 0$，则$\alpha(L)^{-1}$都有定义且唯一，因为可以得到满足$\alpha(L)\beta(L)=1$的唯一解$(\beta_0, \beta_1, \beta_2, \cdots)$，即$\beta_0 = 1/\alpha_0$，$\beta_1 = -\alpha_1\beta_0/\alpha_0$，等等。

$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$.

$(1-\alpha_1 L) y_t = \alpha_0 + \varepsilon_t$

例 $(1-L)^{-1} = 1 + L + L^2 + L^3 + \cdots$

证明： $(1-L)(1+L+L^2+L^3+\cdots)=1$。但$(1-L)^{-1}$不再是 AS。

$\beta_1=\beta$, $\beta_2=\beta^2$

例 $(1-\beta L)^{-1}=1+\beta L+\beta^2L^2+\beta^3L^3+\cdots$

证明： 将上例中的"L"换成"βL"即可。如果$|\beta|<1$，则$(1-\beta L)^{-1}$为 AS。

命题 对于$y_t=\beta_0+\beta_1 y_{t-1}+\varepsilon_t$，假设$|\beta_1|<1$，则此 AR(1)是 MA($\infty$)。

证明：

方法一(迭代法)：

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \\
&= \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= (\beta_0 + \beta_0\beta_1) + \beta_1^2 y_{t-2} + \beta_1\varepsilon_{t-1} + \varepsilon_t \\
&= (\beta_0 + \beta_0\beta_1) + \beta_1^2(\beta_0 + \beta_1 y_{t-3} + \varepsilon_{t-2}) + \beta_1\varepsilon_{t-1} + \varepsilon_t \\
&= \beta_0(1 + \beta_1 + \beta_1^2) + \beta_1^3 y_{t-3} + \beta_1^2\varepsilon_{t-2} + \beta_1\varepsilon_{t-1} + \varepsilon_t \\
&= \cdots \\
&= \beta_0(1 + \beta_1 + \beta_1^2 + \cdots) + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_1^2\varepsilon_{t-2} + \beta_1^3\varepsilon_{t-3} + \cdots \\
&= \frac{\beta_0}{1-\beta_1} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_1^2\varepsilon_{t-2} + \beta_1^3\varepsilon_{t-3} + \cdots
\end{aligned}
$$

方法二(滤波求逆法)：

由于$(1-\beta_1 L)y_t = \beta_0 + \varepsilon_t$，故

$$
\begin{aligned}
y_t &= (1-\beta_1 L)^{-1}(\beta_0 + \varepsilon_t) \\
&= (1+\beta L + \beta^2 L^2 + \cdots)\beta_0 + (1+\beta L + \beta^2 L^2 + \cdots)\varepsilon_t \\
&= \beta_0(1+\beta_1 + \beta_1^2 + \cdots) + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_1^2\varepsilon_{t-2} + \beta_1^3\varepsilon_{t-3} + \cdots \\
&= \frac{\beta_0}{1-\beta_1} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_1^2\varepsilon_{t-2} + \beta_1^3\varepsilon_{t-3} + \cdots
\end{aligned}
$$

可将平稳 AR(1)看成过去所有扰动项的总效应之和，离现在越远的扰动项其影响力呈几何级数递减。

从 AR(1)的MA(∞)表达式可以看出：

$$IRF(j) \equiv \frac{\partial y_{t+j}}{\partial \varepsilon_t} = \beta_1^j$$

$\left(\partial y_{t+j} / \partial \varepsilon_t\right)$表示，当第 t 期的扰动项ε_t变化 1 单位时(其他期扰动项均不变)，对相隔 j 期的y_{t+j}的影响，称为“动态乘子”(dynamic multiplier)。

动态乘子与绝对时间 t 无关，是时间间隔 j 的函数。

将$\left(\partial y_{t+j} / \partial \varepsilon_t\right)$视为$j$的函数，称为“脉冲响应函数”(Impulse Response Function，简记 IRF)，它刻画y_{t+j}对ε_t的 1 单位脉冲(impulse)的响应(response)。

将$\left(j,\ \partial y_{t+j}/\partial\varepsilon_t\right)$画图，可得到对 IRF 的直观认识。

考察ε_t从第 0 期(即当前期)到第 k 期对y_{t+j}的累积效应：

$$CIRF(k) \equiv \sum_{j=0}^{k} \frac{\partial y_{t+j}}{\partial \varepsilon_t}$$

称为“累积脉冲响应函数”(Cumulative Impulse Response Function，简记 CIRF)。

AR(p)也是MA(∞)：

将“$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t$”写为“$\beta(L) y_t = \beta_0 + \varepsilon_t$”，

如果$\beta(L)^{-1}$为 AS，则

$$y_t = \beta(L)^{-1}\beta_0 + \beta(L)^{-1}\varepsilon_t$$

这是MA(∞)的形式。对于 AR(p)的 IRF，可通过计算$\beta(L)^{-1}$得到。

另一方法，直接对 AR(p)进行模拟：

令$y_{-1} = y_{-2} = \cdots = y_{-p} = 0$，$\varepsilon_0 = 1$，而$\varepsilon_t = 0\ (\forall t \neq 0)$，使用 AR($p$)方程“$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t$”(令$t = 0$)可得$y_0$。

将$\left\{y_0, y_{-1}, \cdots, y_{-p+1}\right\}$代入 AR($p$)方程(令$t = 1$)可得$y_1$。

以此类推，可计算出y_j，就是相隔j期的动态乘子$\left(\partial y_{t+j} / \partial \varepsilon_t\right)$。

ARMA(p, q)也是 MA(∞)：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$$y_t - \beta_1 L y_t - \cdots - \beta_p L^p y_t = \beta_0 + \varepsilon_t + \theta_1 L \varepsilon_t + \cdots + \theta_q L^q \varepsilon_t$$

$$\beta(L) y_t = \beta_0 + \theta(L)\varepsilon_t$$

中，$\theta(L) \equiv 1 + \theta_1 L + \cdots + \theta_q L^q$。假设$\beta(L)^{-1}$为 AS，则

$$y_t = \beta(L)^{-1}\beta_0 + \beta(L)^{-1}\theta(L)\varepsilon_t$$

还是MA(∞)的形式。

20.9 向量自回归过程

同时关心几个经济变量的预测。将这些变量放在一起，作为一个系统来预测，称为“多变量时间序列”(multivariate time series)。

Sims(1980)提倡“向量自回归”(Vector Autoregression，简记 VAR)。假设有两个时间序列变量$\{y_{1t}, y_{2t}\}$，分别作为两个回归方程的被解释变量。

Tom Sargent.

解释变量为这两个变量的 p 阶滞后值，构成二元 VAR(p)系统：

$$\begin{cases} y_{1t} = \beta_{10} + \beta_{11} y_{1,t-1} + \cdots + \beta_{1p} y_{1,t-p} + \gamma_{11} y_{2,t-1} + \cdots + \gamma_{1p} y_{2,t-p} + \varepsilon_{1t} \\ y_{2t} = \beta_{20} + \beta_{21} y_{1,t-1} + \cdots + \beta_{2p} y_{1,t-p} + \gamma_{21} y_{2,t-1} + \cdots + \gamma_{2p} y_{2,t-p} + \varepsilon_{2t} \end{cases}$$

$\{\varepsilon_{1t}\}$与$\{\varepsilon_{2t}\}$均为白噪声过程(不存在自相关)，但允许两个方程的扰动项之间存在“同期相关性”(contemporaneous correlation)：

$$\mathrm{Cov}(\varepsilon_{1t}, \varepsilon_{2s}) = \begin{cases} \sigma_{12} & 若 t = s \\ 0 & 其他 \end{cases}$$

两个方程的解释变量完全一样。将两个方程写在一起：

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix} y_{1,\, t-1} + \cdots + \begin{pmatrix} \beta_{1p} \\ \beta_{2p} \end{pmatrix} y_{1,\, t-p} + \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \end{pmatrix} y_{2,\, t-1} + \cdots + \begin{pmatrix} \gamma_{1p} \\ \gamma_{2p} \end{pmatrix} y_{2,\, t-p} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

将同期变量写成列向量，把相应系数合并为矩阵，

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \gamma_{11} \\ \beta_{21} & \gamma_{21} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \beta_{1p} & \gamma_{1p} \\ \beta_{2p} & \gamma_{2p} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

记$\boldsymbol{y}_t \equiv \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$，$\boldsymbol{\varepsilon}_t \equiv \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$，则

$$\boldsymbol{y}_t = \underbrace{\begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix}}_{\boldsymbol{\Gamma}_0} + \underbrace{\begin{pmatrix} \beta_{11} & \gamma_{11} \\ \beta_{21} & \gamma_{21} \end{pmatrix}}_{\boldsymbol{\Gamma}_1} \boldsymbol{y}_{t-1} + \cdots + \underbrace{\begin{pmatrix} \beta_{1p} & \gamma_{1p} \\ \beta_{2p} & \gamma_{2p} \end{pmatrix}}_{\boldsymbol{\Gamma}_p} \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

定义相应的系数矩阵为$\boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \cdots, \boldsymbol{\Gamma}_p$，可得

$$\boldsymbol{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

此形式与 AR(p)相似，故名“VAR(p)”。

$\{\boldsymbol{\varepsilon}_t\}$为“向量白噪声过程”(vector white noise process)。

定义　向量白噪声过程$\{\boldsymbol{\varepsilon}_t\}$是弱平稳的随机过程，满足以下条件：

2x2

$$\mathrm{E}(\boldsymbol{\varepsilon}_t) = \boldsymbol{0},\ \ \mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma},\ \ \mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \boldsymbol{0}\ \ (t \neq s)$$

其中，$\boldsymbol{\Sigma}$为正定矩阵。不要求$\boldsymbol{\Sigma}$为对角矩阵，故向量白噪声过程的各个分量间可存在同期相关。

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \leftarrow y_1, y_2 \text{ 同期协方差.}$$

也称白噪声过程为“新息过程”(innovation process)。

由于 VAR(p)的解释变量$\{\boldsymbol{y}_{t-1},\cdots,\boldsymbol{y}_{t-p}\}$依赖于$\{\boldsymbol{\varepsilon}_{t-1},\boldsymbol{\varepsilon}_{t-2},\cdots\}$，而$\boldsymbol{\varepsilon}_t$与$\{\boldsymbol{\varepsilon}_{t-1},\boldsymbol{\varepsilon}_{t-2},\cdots\}$不相关，故可视所有解释变量为前定变量(predetermined)，与当期扰动项$\boldsymbol{\varepsilon}_t$不相关，故可用 OLS 对每个方程分别进行估计。

如果假设扰动项服从正态分布，可推导出条件 MLE 估计量。

条件 MLE 等价于对每个方程分别用 OLS 进行估计。

方程右边的解释变量均为滞后变量$\{\boldsymbol{y}_{t-1},\cdots,\boldsymbol{y}_{t-p}\}$，不包含同期变量$\boldsymbol{y}_t$，也称“简化式 VAR”(reduced-form VAR)。

如果解释变量包括同期变量$\boldsymbol{y}_t$，则称为“结构 VAR”(structural VAR)，参见第 24 章。

滞后阶数的选择

方法之一是使用信息准则。根据残差$\hat{\boldsymbol{\varepsilon}}_t$可以估计协方差矩阵$\boldsymbol{\Sigma}$，记为$\hat{\boldsymbol{\Sigma}}$。矩阵$\hat{\boldsymbol{\Sigma}}$的$(i, j)$元素为$\hat{\boldsymbol{\Sigma}}_{ij} \equiv \dfrac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{jt}$，其中 T 为样本容量。

VAR 模型的 AIC 与 BIC 分别为 $e'e$

$$\mathrm{AIC}(p) \equiv \ln\left|\hat{\boldsymbol{\Sigma}}\right| + n(np+1)\frac{2}{T}$$

$$\mathrm{BIC}(p) \equiv \ln\left|\hat{\boldsymbol{\Sigma}}\right| + n(np+1)\frac{\ln T}{T}$$

其中，n为 VAR 系统中变量的个数，p为滞后阶数，$\left|\hat{\boldsymbol{\Sigma}}\right|$为$\hat{\boldsymbol{\Sigma}}$的行

列式，而$n(np+1)$为 VAR 模型中待估系数之总数(每个方程共有$(np+1)$个系数，共有 n 个方程)。

方法之二是检验最后一阶系数的显著性。

假设要确定使用 VAR(p) 还是 VAR(p–1)，可检验原假设“$H_0: \beta_{1p} = \beta_{2p} = \gamma_{1p} = \gamma_{2p} = 0$”。

方法之三是检验 VAR 模型的残差是否为白噪声，即是否存在自相关。

如存在自相关，则应在解释变量中加入更高阶的滞后变量。

N小，T大

VAR 变量个数的选择

VAR 系统中包含的变量个数越多，则待估系数越多。

长面板.

假设有 5 个变量，滞后 4 期，则每个方程中共有 21 个待估系数(含截距项)，而整个 VAR 系统共有 105 个待估系数！

待估系数过多使得样本容量过小，增大估计误差，降低预测精度。

VAR 模型通常仅包含为数不多的几个变量。

另一方面，如果 VAR 模型太小，则可能存在遗漏变量偏差。

例 早期的 VAR 模型常发现一个奇怪结果，即紧缩货币政策之后通货膨胀反而上升，称为“物价之谜”(price puzzle)。Sims (1992)认为，美联储在调整基准利率时，会考虑许多因素，而有些因素未包括在 VAR 模型中。如果某遗漏变量的变化预示着通货膨胀加剧，而美联储据此提高利率，则所谓的货币冲击(monetary shocks)不过是预期通胀的反映。此后，在 VAR 模型常使用“大宗商品价格”(commodity price)来控制未来的预期通胀。

在设定 VAR 模型时，主要应根据经济理论来选择变量。比如，经济理论告诉我们，通货膨胀率、失业率、短期利息率互相关联，可构成三变量的 VAR 模型。

如果 VAR 模型包含不相关的变量，则会增大估计量方差，降低预测能力。也可在 VAR 系统中引入其他外生变量。

20.10 VAR 的脉冲响应函数

VAR 模型参数众多，参数的经济意义很难解释，故关注 IRF。

既然有向量自回归，也可以有向量移动平均过程。

将 MA(∞)向多维推广，定义 n 维“无穷阶向量移动平均过程”(Vector Moving Average，简记 VMA(∞))：

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\psi}_0 \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots = \boldsymbol{\alpha} + \sum_{j=0}^{\infty} \boldsymbol{\psi}_j \boldsymbol{\varepsilon}_{t-j}$$

其中$\boldsymbol{\psi}_0 = \boldsymbol{I}_n$，$\boldsymbol{\psi}_j$皆为 n 维方阵。

定义“多维滤波”(multivariate filter)为(无穷项)“滞后矩阵多项式”

(lag matrix polynomial)：

$$\boldsymbol{\psi}(L) \equiv \boldsymbol{\psi}_0 + \boldsymbol{\psi}_1 L + \boldsymbol{\psi}_2 L^2 + \cdots$$

可将 VMA(∞)简洁地写为$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\psi}(L)\boldsymbol{\varepsilon}_t$。对于两个多维滤波，可以定义其乘积，以及多维滤波$\boldsymbol{\psi}(L)$的“逆”(inverse)$\boldsymbol{\psi}(L)^{-1}$。

把上述 VAR(p) 系统“$\boldsymbol{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t$”写成 VMA(∞)的形式：

$$(\boldsymbol{I} - \boldsymbol{\Gamma}_1 L - \cdots - \boldsymbol{\Gamma}_p L^p)\boldsymbol{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\Gamma}(L)\boldsymbol{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\varepsilon}_t$$

其中，$\boldsymbol{\Gamma}(L) \equiv \boldsymbol{I} - \boldsymbol{\Gamma}_1 L - \cdots - \boldsymbol{\Gamma}_p L^p$。在方程两边左乘$\boldsymbol{\Gamma}(L)^{-1}$：

$$\boldsymbol{y}_t = \boldsymbol{\Gamma}(L)^{-1}\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}(L)^{-1}\boldsymbol{\varepsilon}_t$$

记$\boldsymbol{\Gamma}(L)^{-1} \equiv \boldsymbol{\psi}(L) = \boldsymbol{\psi}_0 + \boldsymbol{\psi}_1 L + \boldsymbol{\psi}_2 L^2 + \cdots$，$\boldsymbol{\Gamma}(L)^{-1}\boldsymbol{\Gamma}_0 \equiv \boldsymbol{\alpha}$，可得 VAR 模型的 VMA 表示法(Vector Moving Average Representation)：

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\psi}_0\boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\psi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots = \boldsymbol{\alpha} + \sum\nolimits_{i=0}^{\infty}\boldsymbol{\psi}_i\boldsymbol{\varepsilon}_{t-i}$$

可以证明，$\boldsymbol{\psi}_0 \equiv \boldsymbol{I}_n$，而其余的$\boldsymbol{\psi}_i$可通过递推公式$\boldsymbol{\psi}_i = \sum\nolimits_{j=1}^{i}\boldsymbol{\psi}_{i-j}\boldsymbol{\Gamma}_j$来确定。

M

根据向量微分法则：

$$\frac{\partial \boldsymbol{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'} = \boldsymbol{\psi}_s \qquad n$$

其中，$\left(\partial \boldsymbol{y}_{t+s} / \partial \boldsymbol{\varepsilon}_t'\right)$为 n 维列向量$\boldsymbol{y}_{t+s}$对 n 维行向量$\boldsymbol{\varepsilon}_t'$求偏导数，故得到$n \times n$矩阵$\boldsymbol{\psi}_s$。

矩阵$\boldsymbol{\psi}_s$是一维情形下相隔 s 期的动态乘子向多维的推广，其第 i 行、第 j 列元素等于$\left(\partial y_{i,t+s} / \partial \varepsilon_{jt}\right)$。

$\left(\partial y_{i,t+s} / \partial \varepsilon_{jt}\right)$表示，当第 j 个变量在第 t 期的扰动项ε_{jt}增加 1 单位时(而其他变量与其他期的扰动项均不变)，对第 i 个变量在第$(t+s)$

期的取值$y_{i,t+s}$的影响。

将$\left(\partial y_{i,t+s}/\partial \varepsilon_{jt}\right)$视为时间间隔 s 的函数，就是 IRF。

考察累积脉冲响应函数，即经过 k 期后的累积效应：

$$\sum_{s=0}^{k}\frac{\partial \boldsymbol{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'}=\sum_{s=0}^{k}\boldsymbol{\psi}_s$$

让$k\to\infty$，则得到长期效应：

$$\sum_{s=0}^{\infty}\frac{\partial \boldsymbol{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'}=\sum_{s=0}^{\infty}\boldsymbol{\psi}_s=\boldsymbol{\psi}_0+\boldsymbol{\psi}_1+\boldsymbol{\psi}_2+\cdots=\boldsymbol{\psi}(1)$$

ε_{it}, ε_{jt} 有相关

根据定义，$\boldsymbol{\Gamma}(L)\boldsymbol{\psi}(L)=\boldsymbol{I}$。令$L=1$可得，$\boldsymbol{\Gamma}(1)\boldsymbol{\psi}(1)=\boldsymbol{I}$， 故

$$\boldsymbol{\psi}(1)=\boldsymbol{\Gamma}(1)^{-1}=(\boldsymbol{I}-\boldsymbol{\Gamma}_1-\cdots-\boldsymbol{\Gamma}_p)^{-1}$$

IRF 的缺点是，假定在计算$\left(\partial y_{i,t+s}/\partial\varepsilon_{jt}\right)$时，只让$\varepsilon_{jt}$变动，而所有其他同期扰动项均不变。

此假定只有当$\boldsymbol{\Sigma}\equiv \mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$为对角矩阵时才成立(同期扰动项正交)。否则，当$\varepsilon_{jt}$变动时，必然伴随其他方程的同期扰动项相应地变动。

考虑“正交化的脉冲响应函数”(Orthogonalized Impulse Response Function，简记 OIRF)。

从扰动项$\boldsymbol{\varepsilon}_t$中分离出相互正交的部分，记为$\boldsymbol{u}_t$，然后计算当$\boldsymbol{u}_t$中的

某个分量变动时，对各变量在不同时期的影响。

可以证明，对任何对称正定矩阵$\boldsymbol{\Sigma}_{n\times n}$均可进行“三角分解”：

$$\boldsymbol{\Sigma} = \boldsymbol{ADA}'$$

$\boldsymbol{D}$为 n 级对角矩阵，其所有元素均为正数；$\boldsymbol{A}$为 n 阶下三角矩阵，其主对角线元素均为 1；矩阵$\boldsymbol{A}$与$\boldsymbol{D}$唯一确定。

对扰动项的协方差矩阵$\boldsymbol{\Sigma}$进行三角分解，定义“正交化的扰动项”：

$$\boldsymbol{u}_t \equiv \boldsymbol{A}^{-1}\boldsymbol{\varepsilon}_t$$

可以证明，新扰动项$\boldsymbol{u}_t$不再存在同期相关：

$$\begin{aligned}\mathrm{E}(\boldsymbol{u}_t\boldsymbol{u}_t') &= \mathrm{E}(\boldsymbol{A}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{A}^{-1\prime}) = \boldsymbol{A}^{-1}\,\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\boldsymbol{A}^{-1\prime} \\ &= \boldsymbol{A}^{-1}\boldsymbol{\Sigma}\boldsymbol{A}'^{-1} = \underbrace{\cancel{\boldsymbol{A}^{-1}\boldsymbol{A}}}_{=\boldsymbol{I}}\boldsymbol{D}\underbrace{\cancel{\boldsymbol{A}'}\boldsymbol{A}'^{-1}}_{=\boldsymbol{I}} = \boldsymbol{D}\end{aligned}$$

$\boldsymbol{D}$主对角线上的每一元素即为新扰动项$\boldsymbol{u}_t$相应分量的方差。

实践中，常使用与三角分解相关联的“乔利斯基分解”(Cholesky decomposition)。

由于$\boldsymbol{\Sigma} = \boldsymbol{ADA}'$，而$\boldsymbol{D}$为对角矩阵且所有元素均为正数，故

$$\boldsymbol{\Sigma} = \boldsymbol{AD}^{1/2}\boldsymbol{D}^{1/2}\boldsymbol{A}' \equiv \boldsymbol{PP}'$$

其中，$\boldsymbol{P} \equiv \boldsymbol{A}\boldsymbol{D}^{1/2}$。由于$\boldsymbol{A}$为下三角矩阵，$\boldsymbol{D}^{1/2}$为对角矩阵，故$\boldsymbol{P}$依然为下三角矩阵；而$\boldsymbol{P}'$依然为上三角矩阵。

定义经乔利斯基分解所得的正交化扰动项为

$$\boldsymbol{v}_t \equiv \boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_t$$

新扰动项$\boldsymbol{v}_t$的协方差矩阵为单位矩阵：

$$\begin{aligned}\mathrm{E}(\boldsymbol{v}_t\boldsymbol{v}_t') &= \mathrm{E}(\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{P}^{-1'}) = \boldsymbol{P}^{-1}\,\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\boldsymbol{P}^{-1'} \\ &= \boldsymbol{P}^{-1}\boldsymbol{\Sigma}\boldsymbol{P}'^{-1} = \boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{P}'\boldsymbol{P}'^{-1} = \boldsymbol{I}\end{aligned}$$

新扰动项$\boldsymbol{v}_t$各分量正交，且方差均被标准化为 1。故新扰动项$\boldsymbol{v}_t$的

一个分量变化一单位，也就是变化一个标准差。

乔利斯基分解将扰动项正交化，且标准化。

利用乔利斯基分解，可将 VAR 模型的 VMA 表示法写为：

$$\boldsymbol{y}_t=\boldsymbol{\alpha}+\sum_{i=0}^{\infty}\boldsymbol{\psi}_i\boldsymbol{\varepsilon}_{t-i}=\boldsymbol{\alpha}+\sum_{i=0}^{\infty}\underbrace{\boldsymbol{\psi}_i\boldsymbol{P}}\,\underbrace{\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_{t-i}}=\boldsymbol{\alpha}+\sum_{i=0}^{\infty}\boldsymbol{\Phi}_i\boldsymbol{v}_{t-i}$$

其中，$\boldsymbol{\Phi}_i\equiv\boldsymbol{\psi}_i\boldsymbol{P}$。由向量微分法则可知：

$$\frac{\partial\boldsymbol{y}_{t+s}}{\partial\boldsymbol{v}_t'}=\boldsymbol{\Phi}_s$$

矩阵$\boldsymbol{\Phi}_s$的第 i 行、第 j 列元素等于$\left(\partial y_{i,t+s}/\partial v_{jt}\right)$，即正交化的脉冲

响应函数(OIRF)。

在估计过程中，可根据残差$\hat{\boldsymbol{\varepsilon}}_t$，估计$\hat{\boldsymbol{\Sigma}}$，对$\hat{\boldsymbol{\Sigma}}$进行乔利斯基分解得到$\hat{\boldsymbol{P}}$以及$\hat{\boldsymbol{v}}_t = \hat{\boldsymbol{P}}^{-1}\hat{\boldsymbol{\varepsilon}}_t$。

OIRF 依然有缺点。首先，正交化冲击(orthogonalized shocks)$\boldsymbol{v}_t$的经济含义不易解释。

其次，由于乔利斯基分解所用矩阵$\boldsymbol{P}$为下三角矩阵，故正交化的脉冲响应函数依赖于变量的次序(order of variables)；如改变变量次序，可能得到很不相同的结果。

按照一定变量顺序进行乔利斯基分解，意味着在 VAR 模型之上附加了一种“递归结构”(recursive structure)，故也称为“递归 VAR” (recursive VAR)。

OIRF 虽然使得因果关系更清楚，但代价是需对变量起作用的次序作较强的先验假设，而经济理论通常无法给出明确指南。

在实践中，可借助三种方法来确定 OIRF 的变量次序。

(1) 考察两个变量之间的“交叉相关图”(cross correlogram)。对于变量 y_t 与 x_t，定义其“交叉相关函数”(cross-correlation function)

$$\rho_{yx}(k) \equiv \operatorname{corr}(y_t, x_{t+k})$$

将 $\rho_{yx}(k)$ 视为 k 的函数，并画图，即得到交叉相关图。

$\rho_{yx}(k)$ 并不关于原点对称，即一般 $\rho_{yx}(k) \neq \rho_{yx}(-k)$；但 $\rho_{yx}(k) \equiv \mathrm{corr}(y_t, x_{t+k}) = \mathrm{corr}(x_t, y_{t-k}) \equiv \rho_{xy}(-k)$。

通过交叉相关图，可看出使得 $\left|\rho_{yx}(k)\right|$(绝对值)最大的 k 值，记为 k^*。

如果 $k^* > 0$，则意味着 y_t 与未来的 x_{t+k^*} 最相关(变量排序应为 y_t, x_t)；

如果 $k^* < 0$，则意味着 y_t 与过去的 x_{t+k^*} 最相关(变量排序应为 x_t, y_t)。

(2) 借助格兰杰因果检验来确定两个变量间的排序，参见下文。

(3) 稳健性分析：对于不同的变量排序，比较其 OIRF。

20.11 预测误差的方差分解

对于 VAR(p)模型 $\boldsymbol{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t$，得到参数估计后，进行向前一期的预测：

$$\hat{\boldsymbol{y}}_{t+1} = \hat{\boldsymbol{\Gamma}}_0 + \hat{\boldsymbol{\Gamma}}_1 \boldsymbol{y}_t + \cdots + \hat{\boldsymbol{\Gamma}}_p \boldsymbol{y}_{t-p+1}$$

其中，假设 $\boldsymbol{y}_t$ 为已知，即当前时间为第 t 期。

为了预测 $\boldsymbol{y}_{t+2}$，则可将 $\hat{\boldsymbol{y}}_{t+1}$ 视为 $\boldsymbol{y}_{t+1}$，即

预测　看得到.

$$\hat{\boldsymbol{y}}_{t+2} = \hat{\boldsymbol{\Gamma}}_0 + \hat{\boldsymbol{\Gamma}}_1 \hat{\boldsymbol{y}}_{t+1} + \hat{\boldsymbol{\Gamma}}_2 \boldsymbol{y}_t + \cdots + \hat{\boldsymbol{\Gamma}}_p \boldsymbol{y}_{t-p+2}$$

类似地，可计算向前h期的预测$\hat{\boldsymbol{y}}_{t+h}$。

利用VAR模型的VMA表示法，可将向前h期的预测误差(h-step forecast error)写为：

$$\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h} = \boldsymbol{\psi}_0 \boldsymbol{\varepsilon}_{t+h} + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t+h-1} + \cdots + \boldsymbol{\psi}_{h-1} \boldsymbol{\varepsilon}_{t+1} = \sum_{i=0}^{h-1} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t+h-i}$$

希望度量各方程的扰动项对预测误差的单独贡献。由于$\boldsymbol{\varepsilon}_{t+h-i}$的各分量存在同期相关，故利用乔利斯基分解可得：

$n \times 1$

$$\boldsymbol{y}_{t+h}-\hat{\boldsymbol{y}}_{t+h}=\sum_{i=0}^{h-1}\boldsymbol{\psi}_i\boldsymbol{\varepsilon}_{t+h-i}=\sum_{i=0}^{h-1}\underbrace{\boldsymbol{\psi}_i\boldsymbol{P}}\,\underbrace{\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_{t+h-i}}=\sum_{i=0}^{h-1}\boldsymbol{\Phi}_i\boldsymbol{v}_{t+h-i}$$

其中，$\boldsymbol{v}_{t+h-i}$为正交化冲击。记矩阵$\boldsymbol{\Phi}_i$的(m, k)元素为$\phi_{i,\,mk}$，向量$\boldsymbol{v}_{t+h-i}$的第 k 个元素为$v_{k,t+h-i}$，可$\boldsymbol{y}_{t+h}$的第 j 个分量$y_{j,\,t+h}$的预测误差为：

第j行 → Φ_i v_{t+h-i} 向量.

第j个方程

$$\begin{aligned}y_{j,\,t+h}-\hat{y}_{j,\,t+h}&=\sum_{i=0}^{h-1}(\phi_{i,\,j1}v_{1,t+h-i}+\cdots+\phi_{i,\,jn}v_{n,t+h-i})\\&=\sum_{i=0}^{h-1}\sum_{k=1}^{n}\phi_{i,\,jk}v_{k,t+h-i}\\&=\sum_{k=1}^{n}\sum_{i=0}^{h-1}\phi_{i,\,jk}v_{k,t+h-i}\\&=\sum_{k=1}^{n}(\phi_{0,\,jk}v_{k,t+h}+\cdots+\phi_{h-1,\,jk}v_{k,t+1})\end{aligned}$$

里面：预测误差累积

外面：方程预测误差.

在上式中交换了求和的次序。由于$\boldsymbol{v}_{t+h-i}$的每个正交化分量均不相关，且方差为 1，故可计算对$y_{j,t+h}$预测的均方误差(MSE)：

$$\mathrm{MSE}(\hat{y}_{j,t+h}) \equiv \mathrm{E}\left[(y_{j,t+h}-\hat{y}_{j,t+h})^2\right]=\sum_{k=1}^{n}(\phi_{0,jk}^2+\cdots+\phi_{h-1,jk}^2)$$

可计算第l个变量的正交化冲击对$y_{j,t+h}$预测均方误差的贡献比例：

其中第l个方程.

$$\frac{\phi_{0,jl}^2+\cdots+\phi_{h-1,jl}^2}{\sum_{k=1}^{n}(\phi_{0,jk}^2+\cdots+\phi_{h-1,jk}^2)}$$

这是度量第 l 个变量对第 j 个变量影响的又一方法。

由上式计算的贡献比例是预测时期 h 的函数，可将其列表或画图。所有变量对预测误差均方误差的贡献比例之和为 1。

此性质称为“预测误差的方差分解”(Forecast-error Variance Decomposition，简记 FEVD)。

它将预测误差的来源分别归因于各变量的正交化新息，也称为“新息核算”(innovation accounting)。

在推导过程中使用了乔利斯基分解，故 FEVD 的结果也依赖于变量排序。这是 FEVD 与 IRF 的共同缺陷。

20.12 格兰杰因果检验

经济学中常常需要确定因果关系究竟是从 x 到 y，还是从 y 到 x，抑或双向因果关系。

格兰杰[Granger(1969)]因果检验的思想：如果 x 是 y 的因，但 y 不是 x 的因，则 x 的过去值可帮助预测 y 的未来值，但 y 的过去值却不能帮助预测 x 的未来值。

考虑以下时间序列模型：

$$y_t = \gamma + \sum_{m=1}^{p} \alpha_m y_{t-m} + \sum_{m=1}^{p} \beta_m x_{t-m} + \varepsilon_t$$

滞后阶数 p 可根据信息准则或序贯 t 规则确定。

检验原假设“$H_0: \beta_1 = \cdots = \beta_p = 0$”，即 x 的过去值对预测 y 的未来值没有帮助。

如果拒绝 H_0，则称 x 是 y 的“格兰杰因”(Granger cause)。

将 x 与 y 的位置互换，可检验 y 是否为 x 的格兰杰因。

格兰杰因果关系并非真正意义上的因果关系，充其量只是一种动态相关关系，表明的是一个变量是否对另一变量有“预测能力”(predictability)。

它顶多是因果关系的必要条件(如果不考虑非线性的因果关系)。

格兰杰因果关系也可由第三个变量所引起。

格兰杰因果检验仅适用于平稳序列，或者有协整关系的单位根过程(参见第 21 章)。对于不存在协整关系的单位根变量，则只能先差分，得到平稳序列后再进行格兰杰因果检验。

20.13　面板格兰杰因果检验

对于面板数据，可进行面板格兰杰因果检验。考虑以下面板模型：

$$y_{it} = \gamma + \sum_{m=1}^{p} \alpha_m y_{i,\,t-m} + \sum_{m=1}^{p} \beta_m x_{i,\,t-m} + u_i + \varepsilon_{it} \quad (i = 1,\cdots,n;\, t = 1,\cdots,T)$$

u_i为个体异质性，可能与解释变量相关。

这是动态面板。如果时间维度较短，应使用差分 GMM 或系统 GMM 进行估计。

反之，如果时间维度较长，可考虑偏差校正 LSDV 法。

20.14 VAR 的 Stata 命令及实例

20.15 季节调整

1. 季节效应

对于月度或季度时间序列，常需对其进行“季节调整”(seasonal adjustment)，去掉“季节效应”后才能使用。

年度数据不需要进行季节调整。

例 考察中国的季度 GDP 数据。第一季度包含春节，故通常第一季度的 GDP 偏低。如直接以第二季度 GDP 除以第一季度 GDP 来

计算环比增长率，则会高估第二季度的 GDP 增长率。包含季节效应的时间序列不能直接计算环比增长率，只能计算同比增长率。

可能导致季节效应的因素包括：

(1) 天气因素：比如，在冬季由于取暖而增加能源消耗。
(2) 行政因素：比如，学校开学与放假的日期对交通量的影响。
(3) 固定假日：比如，十一国庆节对旅游与交通的影响。
(4) 移动假日(moving holiday)：比如，春节期间，铁路运输量增加而 GDP 下降。
(5) 日历因素：比如，闰年与闰月的影响。
(6) 交易日效应：比如，五金店销售额在有五个周末的月份高于只有四个周末的月份。

所有这些季节因素共同构成时间序列的“季节要素”(seasonal component)。

该时间序列的长期走势与中期周期被称为“趋势循环要素”(trend cycle component)，有时简称“趋势要素”(trend component)。

其他不可预测的随机扰动则为该序列的“不规则要素”(irregular component)。

2．季节调整的原理

季节调整通常通过估计“季节因子”(seasonal factor)来进行。

季节因子主要分为两种，即“加法季节因子”(additive seasonal factor)与“乘法季节因子”(multiplicative seasonal factor)。

“加法季节因子”意味着对所有第 1 月(或第 1 季)加上相同的季节因子，以此类推。

“加法模型”(additive model)的表达式：节假日 . . . —

$$Y_t = TC_t + S_t + I_t$$

其中，Y_t为原序列，TC_t为趋势循环要素，S_t为季节要素，而I_t为不规则要素。

“乘法季节因子”则意味着对所有第 1 月(或第 1 季)乘以相同的季节因子，以此类推。“乘法模型” (multiplicative model)的表达式：

$$Y_t = TC_t \times S_t \times I_t$$

使用乘法模型要求Y_t序列中不包含零或负数。两边取对数可得：

$$\ln Y_t = \ln TC_t + \ln S_t + \ln I_t$$

此方程在形式上与加法模型相同，故称为“对数加法模型”。

季节调整的目标就是将原序列Y_t分解为趋势循环要素、季节要素

与不规则要素，然后去掉季节要素 S_t，得到季节调整序列(seasonally adjusted series)。

季节调整的方法有多种，使用不同方法，会得到不同的季节调整序列，带有一定的主观性。国外的统计部门一般同时公布原序列与季节调整序列。

以加法模型为例，介绍常用的回归法、移动平均比率法与 X12。

1．回归法

首先生成月度(或季度)虚拟变量，把时间序列对这些虚拟变量进行 OLS 回归，所得残差就是经季节调整后的序列。

例 Kennan (1985)在使用美国制造业的月度产量数据时，首先取对数，然后将其对时间、时间平方以及月度虚拟变量进行回归，所得残差即为经过季节调整并去掉时间趋势的产量对数。

回归法的缺点是，它假设每年的季节调整因子均不变(可以理解为季节固定效应)，但对于移动假日(比如，春节有时落在 1 月，有时落在 2 月)，每年的季节调整因子可能不尽相同。

2．移动平均比率法

移动平均比率法的基础是移动平均(Moving Average)。奇数项移动平均很容易进行。考虑$\{y_1,\cdots,y_T\}$，则时点 t 的 3 项移动平均为

$$MA_t=\frac{y_{t-1}+y_t+y_{t+1}}{3}\quad(t=2,\cdots,T-1)$$

进行 3 项移动平均后的序列MA_t缺失第一项与最后一项，可通过插值来补齐，比如，$MA_1=(2y_1+y_2)/3$，而$MA_T=(2y_T+y_{T-1})/3$。

消除季节影响的简单方法为，对季度数据进行 4 项移动平均；对

月度数据进行 12 项移动平均。

由于 4 与 12 均为偶数，故进行“中心化移动平均”。以季度数据为例，求$t=3$时点上的 4 项平均。

$$MA_{2.5}=\frac{y_1+y_2+y_3+y_4}{4}$$

$$MA_{3.5}=\frac{y_2+y_3+y_4+y_5}{4}$$

定义相应的中心化移动平均值为

$$MA_3 = \frac{1}{2}(MA_{2.5} + MA_{3.5}) = \frac{1}{2}\left(\frac{y_1 + y_2 + y_3 + y_4}{4} + \frac{y_2 + y_3 + y_4 + y_5}{4}\right)$$

中心化移动平均的一般公式为

$$MA_t = \frac{1}{2}\left(\frac{y_{t-2} + y_{t-1} + y_t + y_{t+1}}{4} + \frac{y_{t-1} + y_t + y_{t+1} + y_{t+2}}{4}\right)$$

$$= \frac{1}{4}\left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right)$$

中心化移动平均可视为加权平均，各期权重不完全相同。

类似地，可以计算适用于月度数据的 12 项中心化移动平均。

以加法模型为例，说明季节调整的“移动平均比率法”(ratio to moving average)。该法分为以下五个步骤：

(1) 对季度(月度)数据Y_t进行 4 项(12 项)中心化移动平均，得到趋势循环序列TC_t。

原始

(2) 计算季节要素S_t与不规则要素I_t之和：$SI_t \equiv S_t + I_t = Y_t - TC_t$。注：如果使用乘法模型，则$SI_t \equiv S_t \times I_t = Y_t / TC_t$。

I_t 平均为 0

(3) 对于季(月)度数据的SI_t，分别计算其第 j 季(月)的季度平均值，得到季节因子s_t。例如，如果 $j = 2$，则计算SI_t序列在整个数

据期间所有第 2 季(月)数据的平均值。

(4) 调整季节因子，使得它们的和为 0，得到标准化的季节因子 $S_t \equiv s_t - \frac{1}{k}\sum_{j=1}^{k} s_j$，其中 $k = 4$ (季度数据)，或 $k = 12$ (月度数据)。

注：如果使用乘法模型，则$S_t \equiv \dfrac{s_t}{\sqrt[k]{s_1 \times \cdots \times s_k}}$。

(5) 计算季节调整的最终结果：$TCI_t \equiv Y_t - S_t$。注：如果使用乘法模型，则$TCI_t \equiv Y_t / S_t$。

其中，第(3)-(4)步的目的是将季节序列S_t从序列$SI_t \equiv S_t + I_t$中分离出来。

3．X12 方法

1954 年美国商务部人口普查局首次开发了可在计算机上运行的季节调整程序，称为 X1。

1965 年推出基于移动平均法的 X11，成为全世界统计机构进行季节调整的标准方法。

X11 的基本方法为通过多次移动平均来分解原序列，并对极端值(outlier)进行自动调整。

对于时间序列的两端观测值，X11 进行单边移动平均(one-sided moving average)，但非对称滤波法(asymmetric filter)的质量不高。

加拿大统计局于 1980 年提出 X11-ARIMA，使用 ARIMA 建模方法对原序列进行前推(forecast)与后推(backcast)，然后对加长的序列进行 X11 季节调整。

美国人口普查局于 1998 年推出 X12-ARIMA，基本方法与 X11-ARIMA 一致，主要增加了对交易日、节假日影响的调节功能，以及对各种极端值的处理。

这些极端值包括：

“单点异常值”(Additive Outlier, 简记 AO)，即发生于某时刻的异常值；

“水平移动”(Level Shift，简记 LS)，表示在瞬间变化到一个新水平并维持在此新水平上；

“暂时变化”(Temporary Change，简记 TC)，表示在瞬间变化到一个新水平，但逐渐恢复到原水平。

X12-ARIMA 分为两个模块，即 regARIMA 与 X11。

模块 regARIMA 使用线性回归方法(带 ARIMA 扰动项)来处理交易日、移动假日与极端值的影响，并对原序列进行前推与后推拓展。