第 21 章　单位根与协整

21.1　非平稳序列

称不平稳的时间序列为“非平稳序列”(non-stationary time series)。

(1) 确定性趋势(deterministic trend)。比如，

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

$\mathrm{E}(y_t) = \beta_0 + \beta_1 t$ 随时间而变，故不平稳。去掉时间趋势，即为平稳

序列，故称“趋势平稳”(trend stationary)序列。

(2) 结构变动(structural break)：如果时间序列存在结构变动，则为非平稳序列。对此，可进行邹检验(Chow test)。

(3) 随机趋势(stochastic trend)。比如，随机游走模型(random walk)：

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

其中，$\{\varepsilon_t\}$为白噪声。

由于$\Delta y_t = \varepsilon_t$，故来自$\{\varepsilon_t\}$的任何扰动对$\{y_t\}$都具有永久冲击。

如包含常数项，则为“带漂移的随机游走”(random walk with drift)：

$$y_t = \beta_0 + y_{t-1} + \varepsilon_t, \quad \beta_0 \neq 0$$

其中，β_0为每个时期的平均“漂移”(drift)。

对于随机游走进行一阶差分，可得平稳序列，故称“差分平稳”(difference stationary)序列。

定义 称平稳的时间序列为“零阶单整”(Integrated of order zero)，记为 I(0)。

如果时间序列的一阶差分为平稳过程，则称为“一阶单整”(Integrated of order one)，记为 I(1)，也称“单位根过程”(unit root process)。

更一般地，如果时间序列的 d 阶差分为平稳过程，则称为“d 阶单整”(Integrated of order d)，记为 I(d)。

对于平稳的 I(0)序列，长期而言有回到期望值的趋势，称为“均值回复”(mean-reverting)。

非平稳的 I(1)序列则“到处乱跑”(wander widely)，无此性质。

I(0)序列对过去行为只有有限记忆，过去扰动项对未来的影响随时间而衰减。

I(1)序列对过去行为有无限记忆，过去冲击将永久改变未来的整个序列。

定义　如果时间序列$\{y_t\}$的 d 阶差分为平稳的 ARMA(p, q)过程，则称$\{y_t\}$为 ARIMA(p, d, q)过程。

最常见的为 ARIMA(p, 1, q)，经过一次差分得到平稳 ARMA(p, q)。

21.2　ARMA 的平稳性

在什么情况下，ARMA(p, q)才平稳？

MA(q)平稳，因为是有限白噪声的线性组合。

故 ARMA(p, q)的平稳性取决于 AR(p)部分。

对于 AR(1)，$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$，如$\varepsilon_t$平稳且$|\beta_1| < 1$，则为平稳过程。

AR(1)其实是一阶随机差分方程，其稳定性与确定性差分方程“$y_t = \beta_0 + \beta_1 y_{t-1}$”一样。

而非齐次差分方程“$y_t = \beta_0 + \beta_1 y_{t-1}$”的解取决于对应的齐次差分方程“$y_t = \beta_1 y_{t-1}$”的通解，$y_t = y_0 \beta_1^t$，故稳定条件为$|\beta_1| < 1$。

考虑 AR(p)的平稳性，即$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t$。

对应的确定性齐次差分方程：$y_t = \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p}$。

假设解为$y_t = z^{-t} = (1/z)^t$，其中 z 待定。代入差分方程可得

$$z^{-t}-\beta_1 z^{-(t-1)}-\cdots-\beta_p z^{-(t-p)}=0$$

两边同乘以 z^t 可得特征方程：

$$\phi(z)\equiv 1-\beta_1 z-\cdots-\beta_p z^p=0$$

此多项式方程在复数域中一定有 p 个根(包括重根)。

对应的齐次差分方程也有 p 个形如$(1/z)^t$的解，而通解是这 p 个解的线性组合。

如果要求$\{y_t\}$收敛于某稳定值，则特征方程所有解的范数$\|z\|$(即在复平面上 z 离原点的距离)都须大于 1，故所有解必须都落在复平面上的单位圆之外，参见图 21.1。

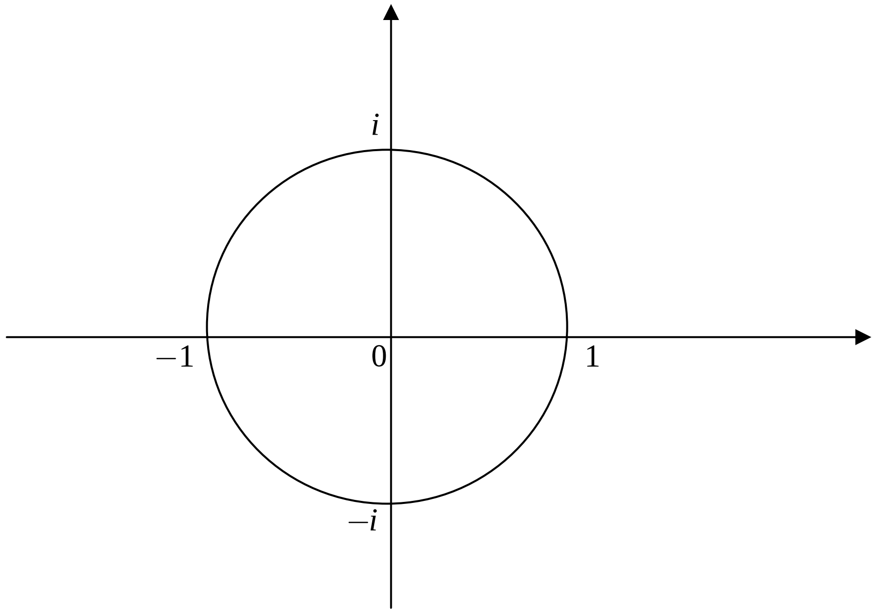

图 21.1　复平面上的单位圆

如果某个根正好落在单位圆之上，则称为“单位根”(unit root)，比如随机游走的情形。

如果特征方程的某个根落在单位圆之内，则为爆炸式(explosive)增长的非平稳过程。

例 对于 AR(1)，其特征方程为$1-\beta_1 z=0$，故$z=1/\beta_1$。

$\|z\|=|z|>1 \Leftrightarrow |\beta_1|<1$。故 AR($p$)稳定性的结论是对 AR(1)的推广。

21.3　VAR 的平稳性

AR(p)的平稳性条件可推广到 VAR(p)：

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

其中，$\{\boldsymbol{\varepsilon}_t\}$为向量白噪声过程。如果对于复数 z，特征方程

$$\left| \boldsymbol{I}_n - \boldsymbol{\Gamma}_1 z - \cdots - \boldsymbol{\Gamma}_p z^p \right| = 0$$

的所有根都落在复平面的单位圆之外(即$\|z\| > 1$)，则此 VAR(p)为平稳过程，其中，$|\cdot|$表示行列式。

当 $p=1$ 时， VAR(1)的平稳性要求 $\left|\boldsymbol{I}_n-\boldsymbol{\Gamma}_1 z\right|=0$ 的所有根都满足 $\|z\|>1$，即 $\|1/z\|<1$。

由于 $\left|\boldsymbol{I}_n-\boldsymbol{\Gamma}_1 z\right|=|z|\cdot\left|(1/z)\boldsymbol{I}_n-\boldsymbol{\Gamma}_1\right|$，故 $\left|(1/z)\boldsymbol{I}_n-\boldsymbol{\Gamma}_1\right|=0$，因此 $1/z$ 为矩阵 $\boldsymbol{\Gamma}_1$ 的特征值(根据特征值的定义)。

故 VAR(1)的平稳性要求 $\boldsymbol{\Gamma}_1$ 的所有特征值都落在单位圆之内(即 $\|1/z\|<1$)。

对于 VAR(p)，可先将其写为 VAR(1)的形式，再判断其平稳性。

首先，定义如下三个 $np\times 1$ 的列向量：

$$\tilde{\boldsymbol{y}}_t = \begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p+1} \end{pmatrix}_{np\times 1}, \quad \tilde{\boldsymbol{\Gamma}}_0 = \begin{pmatrix} \boldsymbol{\Gamma}_0 \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix}_{np\times 1}, \quad \tilde{\boldsymbol{\varepsilon}}_t = \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix}_{np\times 1}$$

其次，定义 $np \times np$ “伴随矩阵”(companion matrix)：

$$\tilde{\boldsymbol{\Gamma}} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_p \\ \boldsymbol{I}_n & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}_n & \mathbf{0} \end{pmatrix}_{np\times np}$$

将 VAR(p)模型写为 VAR(1)的形式：

$$\tilde{\boldsymbol{y}}_t = \tilde{\boldsymbol{\Gamma}}_0 + \tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{y}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t$$

故 VAR(p)平稳性要求伴随矩阵$\tilde{\boldsymbol{\Gamma}}$的所有特征值都在单位圆之内。

21.4　单位根所带来的问题

(1) 自回归系数的估计值向左偏向于 0。

假设对于 AR(1)，$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$，其中$\beta_1 = 1$。

β_1的 OLS 估计量$\hat{\beta}_1$不是渐近正态分布，也不是对称分布(即使是在大样本中)，而是向左偏于 0。

由于$\{y_t\}$不是平稳序列，中心极限定理不再适用。

虽然$\operatorname*{plim}_{n\to\infty}\hat{\beta}_1=\beta_1$，但在有限样本下可能存在较大偏差。

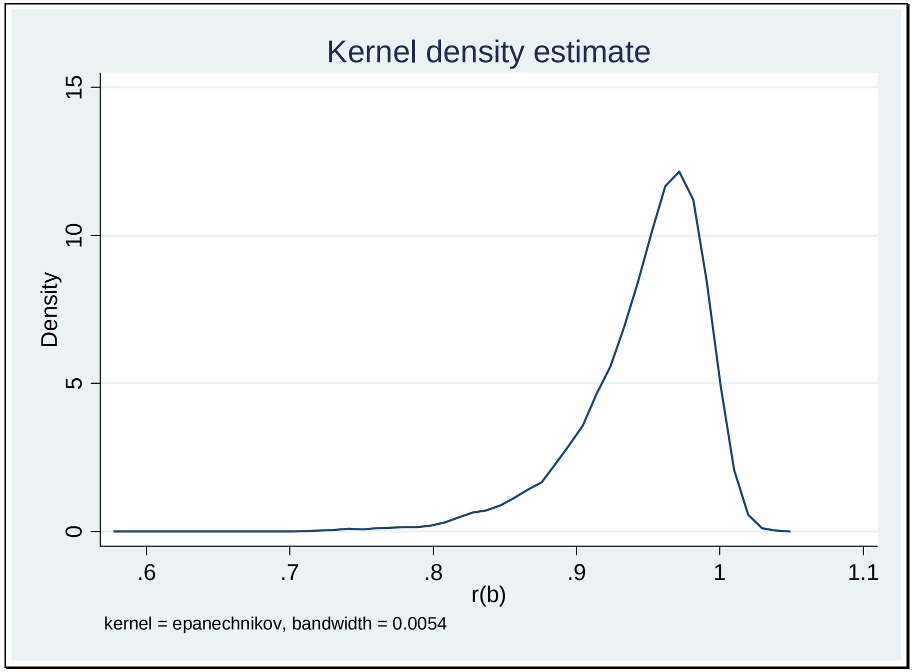

图 21.2　在单位根情况下$\hat{\beta}_1$的大样本分布($n=10,000$)

(2) **传统的 t 检验失效**：由于$\hat{\beta}_1$不是渐近正态分布，t 统计量也不服从渐近标准正态分布，传统的区间估计与假设检验无效。

(3) **两个相互独立的单位根变量可能出现伪回归**(spurious regression)**或伪相关**。

假设$y_t = y_{t-1} + u_t$，$x_t = x_{t-1} + v_t$，其中，u_t, v_t为 iid 且相互独立。故y_t与x_t相互独立。

考虑 OLS 回归，$y_t = \alpha + \beta x_t + \varepsilon_t$。由于$y_t$与$x_t$相互独立，故真实参数$\beta = 0$。

如果样本容量足够大，则$\hat{\beta} \approx 0$，$R^2 \approx 0$，但实际结果并非如此，因为扰动项$\varepsilon_t = y_t - \alpha - \beta x_t$非平稳。

此结论由 Granger and Newbold (1974)通过蒙特卡罗模拟发现。

21.5 单位根检验与平稳性检验

1．Dickey-Fuller 单位根检验(DF 检验)

定义 如果$\{\varepsilon_t\}$独立同分布，期望为 0，方差有限(finite variance)，则称$\{\varepsilon_t\}$为**独立白噪声** (independent white noise)。

考虑 AR(1)：

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

其中，γt为时间趋势(如不含时间趋势，令$\gamma = 0$)；β_0为漂移项（如不带漂移项，令$\beta_0 = 0$）；ε_t为独立白噪声。

考虑单边检验：

$$H_0:\beta_1=1 \quad vs \quad H_1:\beta_1<1$$

方程两边同减y_{t-1}：

$$\Delta y_t=\beta_0+\delta y_{t-1}+\gamma t+\varepsilon_t$$

其中，$\delta\equiv\beta_1-1$。原假设与替代假设变为

$$H_0:\delta=0 \quad vs \quad H_1:\delta<0$$

对方程作 OLS 回归，可得估计量$\hat{\delta}$及相应的 t 统计量。

此 t 统计量称为“Dickey-Fuller 统计量”(简记 DF)，Stata 记为 Z(t)。

DF 统计量不服从渐近正态，临界值须通过蒙特卡罗模拟来获得。

DF 越小(绝对值很大的负数)，则越倾向于拒绝原假设，故 DF 检验是左边单侧检验，即其拒绝域只在分布的最左边。

2．Augmented Dickey-Fuller 单位根检验(ADF 检验)

DF 检验要求扰动项$\{\varepsilon_t\}$为独立白噪声，故扰动项无自相关。

如果$\{\varepsilon_t\}$存在自相关，可引入更高阶的滞后项来控制。

假设选择适当滞后期p，使 AR(p)模型的扰动项$\{\varepsilon_t\}$为独立白噪声：

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

将上式转换为以下形式：

$$y_t = \beta_0 + \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \cdots + \gamma_{p-1} \Delta y_{t-p+1} + \gamma t + \varepsilon_t$$

其中，系数$(\rho, \gamma_1, \gamma_2, \cdots, \gamma_{p-1})$待定，$\left\{\Delta y_{t-1}, \Delta y_{t-2}, \cdots, \Delta y_{t-p+1}\right\}$称为“滞后差分项”(lagged difference terms)。

去掉差分算子，合并同类项可得：

$$y_t = \beta_0 + (\rho + \gamma_1) y_{t-1} + (\gamma_2 - \gamma_1) y_{t-2} + \cdots + (\gamma_{p-1} - \gamma_{p-2}) y_{t-p+1} - \gamma_{p-1} y_{t-p} + \gamma t + \varepsilon_t$$

将两个方程的相应系数对等起来可得

$$
\begin{cases}
\beta_1 = \rho + \gamma_1 \\
\beta_2 = \gamma_2 - \gamma_1 \\
\qquad \vdots \\
\beta_{p-1} = \gamma_{p-1} - \gamma_{p-2} \\
\beta_p = -\gamma_{p-1}
\end{cases}
$$

由最后一个方程倒推上去可得

$$
\begin{cases}
\rho = \beta_1 + \cdots + \beta_p \\
\gamma_1 = -(\beta_2 + \cdots + \beta_p) \\
\qquad \vdots \\
\gamma_{p-2} = -(\beta_{p-1} + \beta_p) \\
\gamma_{p-1} = -\beta_p
\end{cases}
$$

命题 如果$\rho = 1$，则 AR(p)有一个单位根。

证明：由于$\rho = 1$，故$\phi(1) = 1 - \beta_1 - \cdots - \beta_p = 1 - \rho = 0$。故 1 是特征方程$\phi(z) = 1 - \beta_1 z - \cdots - \beta_p z^p = 0$的根，正好落在单位圆上，故 AR($p$)有单位根。

命题 如果$\rho > 1$，则特征方程至少有一个根在单位圆之内，故 AR(p)为非平稳。

证明：首先，$\phi(0) = 1 - \beta_1 \cdot 0 - \cdots - \beta_p \cdot 0 = 1$。

其次，由于$\rho > 1$，故$\phi(1) = 1 - \beta_1 - \cdots - \beta_p = 1 - \rho < 0$。

然而$\phi(z)$为连续函数，根据中值定理，存在复数z^*，满足$0 < \left\| z^* \right\| < 1$，

使得$\phi(z^*)=0$，参见图 21.4。由于$\|z^*\|<1$，z^*落在单位圆之内，故AR(p)为非平稳。

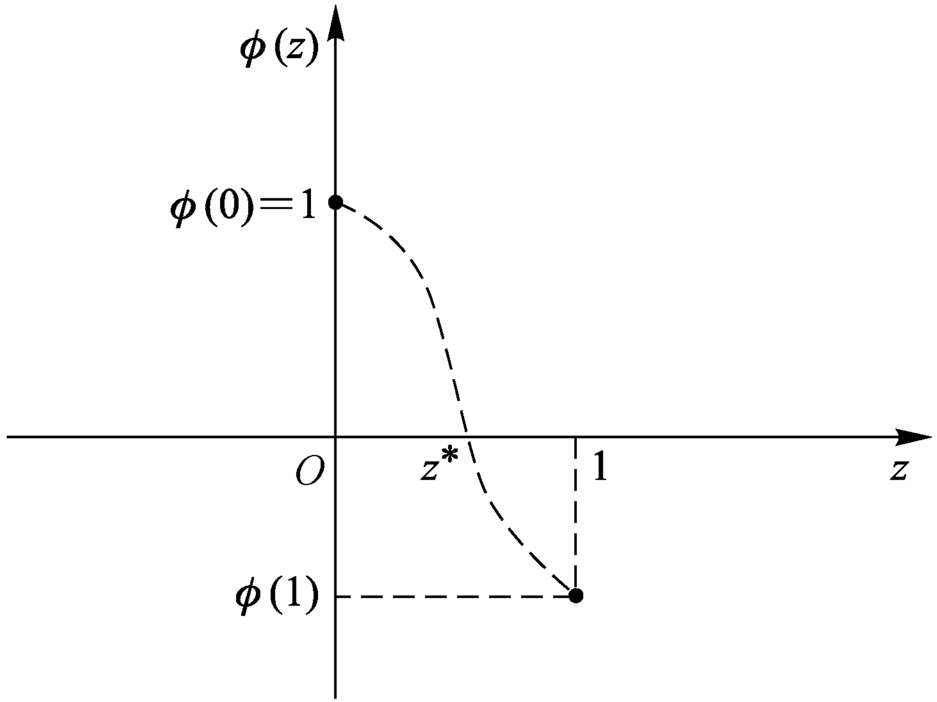

图 21.4　$\rho>1$时 AR(p)的非平稳性

为检验 AR(p)是否有单位根，可对方程进行回归，并检验

$$H_0: \rho = 1 \quad vs \quad H_1: \rho < 1$$

在方程两边同时减去y_{t-1}可得：

$$\Delta y_t = \beta_0 + \delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \cdots + \gamma_{p-1} \Delta y_{t-p+1} + \gamma t + \varepsilon_t$$

其中，$\delta \equiv \rho - 1$。则原假设与替代假设变为

$$H_0: \delta = 0 \quad vs \quad H_1: \delta < 0$$

对方程使用 OLS 可得估计量$\hat{\delta}$及相应的 t 统计量。

此 t 统计量称为“Augmented Dickey-Fuller 统计量”(简记 ADF)。

ADF 检验也是左边单侧检验，其拒绝域只在分布的最左边。

ADF 统计量的临界值也要通过蒙特卡罗模拟得到。

ADF 统计量的临界值取决于真实模型(H_0)是否带漂移项，以及回归方程是否包含常数项或时间趋势。

关于常数项与时间趋势项

是否应带常数项或时间趋势项，主要从理论上考虑。

比如，考察 GDP 之对数是否有单位根，一般应包含时间趋势项；而利率、汇率等则不应有时间趋势项。

也可通过画变量的时间序列图来大致判断有无长期增长趋势。

在作 ADF 检验时，使用选择项“`regress`”，可看到常数项或时间趋势项是否显著。

如无从判断，可把各种情况都检验，将结果以(c, t, p)格式列表，其中“$c=1$”表示带常数项，“$c=0$”表示不带常数项；“$t=1$”表示带趋势项，“$t=0$”表示不带趋势项；p 表示滞后期数。

关于滞后阶数 p 的确定

ADF 检验的结果常对滞后阶数 p 敏感。如果 p 太小，则扰动项$\{\varepsilon_t\}$可能存在自相关，使检验出现偏差。

如果 p 太大，则会降低检验的功效(power)。

Schwert (1989)建议 $p_{\max} = \left[12 \cdot (T/100)^{1/4}\right]$，其中 T 为样本容量，$[\cdot]$ 表示整数部分，然后使用由大到小的序贯 t 规则。

也可使用信息准则，比如 AIC 或 BIC。

3．Phillips-Perron 单位根检验(PP 检验)

ADF 检验通过引入高阶滞后项来保证扰动项$\{\varepsilon_t\}$没有自相关。

“Phillips-Perron 检验”(Phillips and Perron, 1988，简记 PP)仍使用一阶自回归，但使用异方差自相关稳健的标准误修正 DF 统计量：

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

其中，$\{\varepsilon_t\}$可以存在异方差或自相关。

经修正的 Z(t)统计量，其渐近分布与 DF 统计量相同，故临界值也相同，也是左边单侧检验。

Phillips and Perron (1988)还提供了另一检验统计量$Z(\rho)$。

使用 PP 检验须指定用于计算 Newey-West 标准误的滞后阶数(Newey-West lags)。

Stata 默认的 Newey-West 滞后阶数为$\left[4(T/100)^{2/9}\right]$，其中 T 为样本容量，$[\cdot]$表示取整数。

金融变量常存在异方差与自相关，PP 检验在金融数据中应用多。

PP 检验的另一优点是，不必指定差分滞后项的滞后阶数。

4．DF-GLS 单位根检验

ADF 检验与 PP 检验的共同缺点是，检验功效较低(犯第 II 类错误的概率很大)，尤其当样本容量不大，或真实模型接近于单位根的情形。

Elliot, Rothenberg and Stock (1996)提出以下的两步检验。

第一步，用 GLS 估计原序列$\{y_t\}$的常数项与时间趋势项$\hat{\delta}_0 + \hat{\delta}_1 t$，

计算去势后(detrended)的序列$\left\{y_t^d \equiv y_t - \hat{\delta}_0 - \hat{\delta}_1 t\right\}$。

第二步，对$\left\{y_t^d\right\}$使用 ADF 检验。

此检验称为“DF-GLS 检验”，是目前最有功效的单位根检验。

5．KPSS 平稳性检验

Kwiatkowski, Phillips, Schmidt and Shin (1992)提出平稳性检验(KPSS)将原假设改为“H_0 :时间序列为平稳”，而替代假设变为“H_1 :有单位根”。

假设时间序列y_t可分解为时间趋势、随机游走与平稳过程之和：

$$y_t = \beta t + u_t + \varepsilon_t$$

$$u_t = u_{t-1} + v_t,\quad v_t \sim WN(0, \sigma_v^2)$$

其中，βt为时间趋势，u_t是随机游走，ε_t为平稳过程(允许存在异方差与自相关)；v_t为白噪声，其方差为σ_v^2。

“y_t为趋势平稳”(trend stationary)的原假设等价于“$H_0 : \sigma_v^2 = 0$”(即$v_t \equiv 0$，故u_t为常数)，而替代假设为“$H_1 : \sigma_v^2 > 0$”。

如不含时间趋势($\beta = 0$)，则原假设为“y_t为平稳过程”(level stationary)。

对此原假设进行 LM 检验，可得 KPSS 统计量。

KPSS 检验是单边右侧检验，其临界值须通过蒙特卡罗模拟得到。

6．单整阶数(order of integration)的确定

进行单位根检验后，如认为$\{y_t\}$非平稳，要进一步判断其为 I(1)或 I(2)。

对一阶差分$\{\Delta y_t\}$进行单位根检验，如$\{\Delta y_t\}$为平稳，则$\{y_t\}$是 I(1)。

否则，继续对二阶差分$\{\Delta^2 y_t\}$进行单位根检验。如$\{\Delta^2 y_t\}$平稳，则$\{y_t\}$为 I(2)，以此类推。

21.6 单位根检验的 Stata 实例

21.7 面板单位根检验

如样本容量较小，对单个变量进行单位根检验的功效可能很弱。

如有面板数据，可找到更有效的检验方法。

例 同时检验多个国家的实际汇率是否含有单位根。

面板单位根检验，可通过 Stata 命令 `xtunitroot` 来实现。

为检验$\{y_{it}\}$是否包含单位根，考虑如下面板自回归模型：

$$y_{it} = \rho_i y_{i,\,t-1} + \boldsymbol{z}_{it}'\boldsymbol{\gamma}_i + \varepsilon_{it}$$

其中，$i = 1, \cdots, n$为横截面单位，$t = 1, \cdots, T_i$为时间。

根据命令 `xtunitroot` 的默认设置，$\boldsymbol{z}_{it}'\boldsymbol{\gamma}_i$表示个体固定效应(即$\boldsymbol{z}_{it} = 1$)；如果加上选择项“`trend`”，则$\boldsymbol{z}_{it}'\boldsymbol{\gamma}_i$表示个体固定效应与时间趋势，即$\boldsymbol{z}_{it}' = (1, t)$；如果加上选择项“`noconstant`”，则忽略$\boldsymbol{z}_{it}'\boldsymbol{\gamma}_i$。

下文的 IPS 检验、费雪式检验与 Hadri LM 检验允许非平衡面板；其他检验则要求平衡面板，即$T_i = T, \forall i$。

面板单位根检验的原假设为“$H_0 : \rho_i = 1, \forall i$”，而替代假设为“$H_1 : \rho_i < 1$”。方程可写为等价形式：

$$\Delta y_{it} = \delta_i y_{i,\,t-1} + \boldsymbol{z}_{it}'\boldsymbol{\gamma}_i + \varepsilon_{it}$$

其中，$\delta_i \equiv \rho_i - 1$。相应的原假设与替代假设变为

$$H_0: \delta_i = 0, \forall i \quad vs \quad H_1: \delta_i < 0$$

有些面板单位根检验(LLC 检验、HT 检验与 Breitung 检验)，假设各面板单位的自回归系数均相同，称为“共同根” (common root)，即 $\rho_i = \rho, \forall i$。

其他检验则允许各面板单位的自回归系数不同。

为导出统计量的大样本分布，这些检验对于横截面维度 n 或时间维度 T 是否固定，或趋于无穷的速度所作的渐近假定不尽相同。

例 基于$n/T \to 0$的检验，要求时间维度 T 增长速度快于横截面维度 n，故适用于长面板。

例 基于 T 固定而$n \to \infty$的检验适用于短面板。

Stata 手册将这些检验分类总结如表 21.2：

表 21.2　面板单位根检验的特征

检验	Stata 选择项	适用的渐近理论	允许不同的自回归系数	允许非平衡面板
LLC	`noconstant`	$\sqrt{n}/T \to 0$	否	否
LLC		$n/T \to 0$	否	否
LLC	`trend`	$n/T \to 0$	否	否
HT	`noconstant`	$n \to \infty$，T 固定	否	否
HT		$n \to \infty$，T 固定	否	否
HT	`trend`	$n \to \infty$，T 固定	否	否
Breitung	`noconstant`	$(T, n) \to_{\text{seq}} \infty$	否	否
Breitung		$(T, n) \to_{\text{seq}} \infty$	否	否
Breitung	`trend`	$(T, n) \to_{\text{seq}} \infty$	否	否

IPS		$n \to \infty$，T 固定； 或 n, T 都固定	是	是
IPS	`trend`	$n \to \infty$，T 固定； 或 n, T 都固定	是	是
IPS	`lags()`	$(T, n) \to_{\text{seq}} \infty$	是	是
IPS	`trend` `lags()`	$(T, n) \to_{\text{seq}} \infty$	是	是
费雪式		$T \to \infty$， n 有限或趋无穷	是	是
Hadri LM		$(T, n) \to_{\text{seq}} \infty$	-	否
Hadri LM	`trend`	$(T, n) \to_{\text{seq}} \infty$	-	否

其中，$(T,n)\to_{\text{seq}}\infty$表示“序贯极限”(sequential limit)，即首先给定 n，让$T\to\infty$，然后再让$n\to\infty$。实践中，这要求 T 较大(large)，而且 n 也不能太小(at least moderate)。

由于 Hadri LM 检验为面板平稳性检验(原假设为平稳过程)，故不存在是否“允许不同的自回归系数”的问题。

1. LLC 检验

由于扰动项可能存在自相关，Levin, Lin and Chu (2002)(简记 LLC)引入高阶差分滞后项：

$$\Delta y_{it} = \delta y_{i,\,t-1} + \boldsymbol{z}'_{it}\boldsymbol{\gamma}_i + \sum_{j=1}^{p_i} \theta_{ij} \Delta y_{i,\,t-j} + \varepsilon_{it}$$

其中，δ为共同的自回归系数(共同根)；

不同个体的滞后阶数p_i可以不同；

$\{\varepsilon_{it}\}$为平稳的 ARMA 过程；

不同个体的ε_{it}相互独立(不存在截面相关)，但允许异方差。

引入足够高阶的差分滞后项，可保证ε_{it}为白噪声。

由于此方程为动态面板，如直接进行 OLS 回归，估计量$\hat{\delta}$及 t 统计量将存在动态偏差，且不服从渐近正态。

Levin, Lin and Chu (2002) 提出“偏差校正 t 统计量”(bias-adjusted t statistic)，记为t_{δ}^{*}，在大样本下服从标准正态分布。

与 ADF 检验类似，LLC 检验也是左边单侧检验，即拒绝域仅在分布的最左边。

LLC 检验假设不存在截面相关。

如果此假设不成立，则 LLC 检验将存在“显著性水平扭曲”。

为了缓解可能存在的截面相关，建议先将面板数据减去各截面单位的均值(cross-sectional means)，再进行 LLC 检验。

2. HT 检验

LLC检验仅适用长面板，而许多微观面板数据的时间维度T较小。

Harris and Tzavalis (1999) (简记HT)提出了基于T固定而$n \to \infty$的检验。令方程中的自回归系数均相等可得：

$$y_{it} = \rho y_{i,\,t-1} + \boldsymbol{z}_{it}'\boldsymbol{\gamma}_i + \varepsilon_{it}$$

其中，ρ为共同根；ε_{it}服从 iid 正态分布，故为同方差。

在$H_0: \rho = 1$成立的情况下，Harris and Tzavalis (1999)导出 OLS 估计量$\hat{\rho}$的期望μ与方差σ^2的表达式(为 T 的函数)，并证明当 T 固定

而$n \to \infty$时，$z \equiv \dfrac{\hat{\rho} - \mu}{\sqrt{\sigma^2 / n}} \xrightarrow{d} N(0,1)$。

基于此大样本分布，然后进行左边单侧检验。

. Breitung 检验

Breitung 检验(Breitung 2000)的基本思路与 LLC 检验类似。

主要区别在于，首先对数据进行“向前正交变换”(forward orthogonalization)，即减去未来各期的平均值，再进行回归，使得回归后不再需要偏差校正。

所得检验统计量记为λ，服从渐近标准正态分布，然后进行左边单侧检验。

Breitung 检验假设数据生成过程为 AR(1)。

如果存在更高阶的自回归项，应先进行“预白噪声化”(prewhitening)，以消除原序列的自相关，即分别把Δy_{it}与$y_{i,t-1}$对$(\Delta y_{i,t-1},\cdots,\Delta y_{i,t-p})$进行回归，然后以这两个回归的残差来替代$\Delta y_{it}$与$y_{i,t-1}$进行 Breitung 检验。

Breitung (2000)假设不同个体的扰动项不存在截面相关，而

Breitung and Das (2005)则提出在截面相关情况下也成立的检验。

4. IPS 检验

LLC 检验、HT 检验与 Breitung 检验的共同局限在于，它要求每位个体的自回归系数δ都相等，此共同根假设在实践中可能过强。

Im, Pesaran and Shin (2003)(简记 IPS)假设面板数据中共有 n 个相互独立的个体，对每位个体分别进行如下 DF 式回归：

$$\Delta y_{it} = \delta_i y_{i,\,t-1} + \boldsymbol{z}_{it}'\boldsymbol{\gamma}_i + \varepsilon_{it}$$

其中，δ_i为个体 i 的自回归系数；ε_{it}服从相互独立的正态分布(扰

动项无自相关)，但允许异方差。

假设 T 固定，而$n \to \infty$或固定。

面板单位根的原假设为“$H_0: \delta_i = 0, \forall i$”。

替代假设为“服从平稳过程的个体比例大于零”，即当$n \to \infty$时，n_1/n收敛至某非零正数，其中n_1为服从平稳过程的个体数。

记个体 i 的 t 统计量(即 ADF 统计量)为t_i，计算所有个体 t 统计量的样本均值$\bar{t} \equiv \dfrac{1}{n}\sum_{i=1}^{n} t_i$ (Stata 称为 t-bar)。

Im, Pesaran and Shin (2003)给出了$\bar{t}$分布的临界值。

Stata 还汇报另一统计量$\bar{\tilde{t}}$(Stata 称为 t-tilde-bar)，它与$\bar{t}$的区别仅在对扰动项的方差估计量不同。

将$\bar{\tilde{t}}$标准化，构造如下统计量$Z_{\bar{\tilde{t}}}$(Stata 称为 Z-t-tilde-bar)：

$$Z_{\bar{\tilde{t}}} \equiv \frac{\bar{\tilde{t}} - \mathrm{E}(\bar{\tilde{t}})}{\sqrt{\mathrm{Var}(\bar{\tilde{t}})/n}} \xrightarrow{d} N(0,1)$$

其中，$\mathrm{E}(\bar{\tilde{t}})$与$\mathrm{Var}(\bar{\tilde{t}})$为$\bar{t}$的理论均值与方差，可通过查表获得。

由于假设这 n 个时间序列相互独立，故适用中心极限定理，因此$Z_{\bar{\tilde{t}}}$的渐近分布为标准正态。

IPS 检验也是左边单侧检验。

如果扰动项ε_{it}存在自相关，可引入差分滞后项来消除，即对每位个体分别进行如下 ADF 式回归：

$$\Delta y_{it} = \delta_i y_{i,\,t-1} + \boldsymbol{z}_{it}'\boldsymbol{\gamma}_i + \sum_{j=1}^{p_i} \theta_{ij} \Delta y_{i,\,t-j} + \varepsilon_{it}$$

其中，不同个体的滞后阶数p_i可以不同(可通过信息准则来确定)，且假设$(T, n) \to_{\text{seq}} \infty$。其余检验步骤与扰动项无自相关的情形类似，记其统计量为$W_{\bar{t}}$ (Stata 称为 W-t-bar)，对应于上文的$Z_{\tilde{\bar{t}}}$统计量。

5. 费雪式检验

费雪式检验的大思路类似于 IPS 检验，即对每位个体分别进行检验，然后再将这些信息综合起来。

对面板数据中的每位个体分别进行单位根检验 (ADF 检验或 PP 检验)，得到 n 个检验统计量及相应的 p 值$\{p_1, \cdots, p_n\}$。

Choi (2001)提出以下四种方法将这些 p 值综合成“费雪式” (Fisher type)统计量。

方法一为“逆卡方变换”(inverse chi-squared transformation)：

$$P \equiv -2\sum\nolimits_{i=1}^{n} \ln p_i \xrightarrow{d} \chi^2(2n) \quad (T_i \to \infty)$$

其中，T_i为个体 i 的时间维度(因个体而异，允许非平衡面板)。由于取负号，故是单边右侧检验，即统计量P越大，则越倾向于拒绝“面板单位根”的原假设。

方法二为“逆正态变换”(inverse normal transformation)：

$$Z \equiv \frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n} \Phi^{-1}(p_i) \xrightarrow{d} N(0,1) \quad (T_i \to \infty)$$

其中，$\Phi^{-1}(\cdot)$为标准正态累积分布函数的逆函数，故名。如果使用方法二，则为单边左侧检验。

方法三为“逆逻辑变换”(inverse logit transformation)：

$$L^* \equiv \sqrt{k}\sum_{i=1}^{n}\ln\left(\frac{p_i}{1-p_i}\right) \xrightarrow{d} t(5n+4) \quad (T_i \to \infty)$$

其中，$k \equiv \dfrac{3(5n+4)}{\pi^2 n(5n+2)}$。方法三也是单边左侧检验。

如果面板中的个体数 n 很大，可使用“修正逆卡方变换”(modified

inverse chi-squared transformation)：

$$P_m \equiv -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\ln p_i + 1)\xrightarrow{d} N(0,1) \quad (T_i, n \to \infty)$$

6. Hadri LM 检验

Hadri (2000)把 KPSS 平稳性检验推广到面板数据，提出了检验面板平稳性的 LM 检验(原假设为平稳过程)。

考虑以下面板形式的 KPSS 检验模型：

$$y_{it} = \beta_i t + u_{it} + \varepsilon_{it}$$
$$u_{it} = u_{i,\,t-1} + v_{it}$$

其中，$\beta_i t$ 为个体 i 的时间趋势(panel-specific time trend)，而扰动项 ε_{it} 与 v_{it} 均服从 iid 正态分布，方差分别为 σ_{ε}^2 与 σ_v^2。

面板平稳性的原假设等价于“$H_0 : \lambda = \dfrac{\sigma_v^2}{\sigma_{\varepsilon}^2} = 0$”，而替代假设为“$H_1 : \lambda > 0$”。

21.8　协整的思想与初步检验

对于有单位根的变量，传统的处理方法是进行一阶差分。

但一阶差分后变量的经济含义与原序列并不相同。

如果多个单位根变量之间由于某种经济力量而存在“长期均衡关系”(long-run equilibrium)，则有可能进行这种回归。

基本思想：如果多个单位根序列有“共同的随机趋势”(common stochastic trend)，则可对这些变量作线性组合而消去此随机趋势。

例　短期利率与长期利率可能都是单位根过程，而二者的走势也很相似。从经济理论上来看，长期利率是未来预期短期利率的平均值与“风险溢价”(risk premium)之和，故存在长期均衡关系。

例 (非正式)　当你遛狗时，假设你与狗的每一步位置为随机游走过程(带漂移项)，故均为单位根过程。由于你与狗之间有一根皮带相连(“长期均衡关系”)，故你与狗的位置之间不会相离太远(尽管二者都是单位根过程)。

假设两个 I(1)过程$\{y_t\},\{x_t\}$可以分别表示为

$$\begin{cases} y_t = \alpha + \beta w_t + \varepsilon_t \\ x_t = \gamma + \delta w_t + u_t \end{cases}$$

其中，w_t为随机游走，$w_t = w_{t-1} + v_t$；而ε_t, u_t, v_t均为白噪声。

由于$\{y_t\}$与$\{x_t\}$有共同随机趋势w_t，故二者的如下线性组合为平稳过程：

$$\delta y_t - \beta x_t = (\alpha\delta - \beta\gamma) + (\delta\varepsilon_t - \beta u_t)$$

在这种情况下，称$\{y_t\}$与$\{x_t\}$是“协整的”(cointegrated)，而称向量$(\delta, -\beta)$为“协整向量”(cointegrating vector)或“协整系数”。

可以把$(\delta, -\beta)$标准化为$(1, -\beta/\delta)$。

对于两个 I(1)变量，只可能存在一个协整关系。

对于 n 个 I(1)变量，则最多可能存在$(n-1)$个协整关系。

一组 I(1)变量之间协整关系的个数称为“协整秩”(cointegration rank)，即线性无关的协整向量的个数。

如何判断一组 I(1)变量间是否存在协整关系呢？

首先，这些变量必须在理论上可能存在长期均衡关系。

其次，如果只有两个变量，则可以直接画图，但不严格。

Engle and Granger (1987)提出如下的“EG-ADF 检验”。

原假设为$\{y_t, x_t\}$存在协整关系，且协整系数为$\{1, -\theta\}$，则$\{z_t \equiv y_t - \theta x_t\}$为平稳过程。

如果θ已知，可用 ADF 检验来确定$\{z_t\}$是否平稳。

如果接受“$\{z_t\}$为平稳”，则认为$\{y_t, x_t\}$存在协整关系。

通常并不知道θ，故“EG-ADF 方法”分两步进行。

第一步　用 OLS 估计协整系数θ，即$y_t = \phi + \theta x_t + z_t$。

在“$\{y_t, x_t\}$存在协整关系”的原假设下，虽然$\{y_t, x_t\}$非平稳，但$\{z_t\}$为平稳过程，OLS 的估计量$\hat{\phi}$与$\hat{\theta}$都是一致估计量。

第二步　对残差序列$\left\{\hat{z}_t \equiv y_t - \hat{\phi} - \hat{\theta} x_t\right\}$进行 ADF 检验，确定其是否平稳。

由于协整系数$\hat{\theta}$是估计出来的，不一定是真实的协整系数，故 EG-ADF 统计量的临界值与普通的 ADF 检验不同，参见 Hayashi (2000, p. 646)或 Stock and Watson (2004, p. 557)。

如果检验结果认为$\{\hat{z}_t\}$平稳，则接受“$\{y_t, x_t\}$存在协整关系”的原假设。

协整关系“$y_t = \hat{\phi} + \hat{\theta} x_t$”为$\{y_t, x_t\}$之间的长期均衡关系。

EG-ADF 方法的缺点是，不能处理同时存在多个协整关系的情形。

由于 EG-ADF 方法分两步进行，第一步估计的误差会被带到第二步中，故不是最有效率的方法。

更有效率的方法是 MLE，同时估计长期与短期参数。

21.9 Beveridge-Nelson 分解公式

定义 称时间序列$\{y_t\}$为“**线性 I(0)过程**”，如果$y_t = \delta + u_t$，其中δ为常数，$u_t = \psi(L)\varepsilon_t$，$\{\varepsilon_t\}$为独立白噪声，滤波$\psi(L) \equiv \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots$，满足$\sum_{j=0}^{\infty} j\left|\psi_j\right| < \infty$(称为“一可加总”，one-summable，简记为 OS)，$\psi(1) = \psi_0 + \psi_1 + \psi_2 + \cdots \neq 0$。

OS 比 AS(绝对值可加总)的假定更强，前者是后者的充分条件。“$\psi(1) \neq 0$”是一个技术性条件，防止出现退化情形。

本章下面讨论的 I(0)皆为线性 I(0)过程。

假设序列$\{y_t\}$为 I(1)，则其差分为 I(0)，故可表示为$\Delta y_t = \delta + u_t$，其中$u_t$为线性 I(0)过程。

假设时间从$t=0$开始，则$y_1 = y_0 + \delta + u_1$，

$y_2 = \delta + \underbrace{y_0 + \delta + u_1}_{=y_1} + u_2 = y_0 + 2\delta + u_1 + u_2$，…，$y_t = y_0 + \delta t + \sum_{s=1}^{t} u_s$。

由于u_t为线性 $\boldsymbol{I}$(0)过程，故可写为$u_t = \psi(L)\varepsilon_t$，其中

$$\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots$$

将滤波$\psi(L)$分解为

$$\begin{aligned}\psi(L) &= \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots \\ &= (\psi_0 + \psi_1 + \psi_2 + \cdots) + (\psi_1 L - \psi_1) + (\psi_2 L^2 - \psi_2) + (\psi_3 L^3 - \psi_3) + \cdots\end{aligned}$$

$$
\begin{aligned}
&= \psi(1) + \left[-\psi_1(1-L) - \psi_2(1-L^2) - \psi_3(1-L^3) - \cdots \right] \\
&= \psi(1) + (1-L)\left[-\psi_1 - \psi_2(1+L) - \psi_3(1+L+L^2) - \cdots \right] \\
&= \psi(1) + (1-L)\left[-(\psi_1+\psi_2+\cdots) - (\psi_2+\psi_3+\cdots)L - (\psi_3+\psi_4+\cdots)L^2 - \cdots \right] \\
&= \psi(1) + (1-L)\left[\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots \right],
\end{aligned}
$$

其中$\alpha_j \equiv -(\psi_{j+1} + \psi_{j+2} + \cdots)$ $= \psi(1) + (1-L)\alpha(L)$,
而$\alpha(L) \equiv \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$。

由于$\psi(L)$为 OS，可证明$\alpha(L)$为 AS。

定义$\eta_t \equiv \alpha(L)\varepsilon_t$。由于$\alpha(L)$为 AS，故$\eta_t$为平稳过程。因此，

$$
u_t = \psi(L)\varepsilon_t = \left[\psi(1) + (1-L)\alpha(L)\right]\varepsilon_t = \psi(1)\varepsilon_t + (1-L)\eta_t = \psi(1)\varepsilon_t + \eta_t - \eta_{t-1}
$$

$$y_t = y_0 + \delta t + \sum_{s=1}^{t} u_s = \underbrace{\delta t}_{\text{time trend}} + \underbrace{\psi(1)\sum_{s=1}^{t}\varepsilon_s}_{\text{random walk}} + \underbrace{\eta_t}_{\text{stationary}} + \underbrace{(y_0 - \eta_0)}_{\text{initial condition}}$$

"Beveridge-Nelson 分解公式"将 I(1)过程分解为时间趋势δt、随机游走$\psi(1)\sum_{s=1}^{t}\varepsilon_s$、平稳序列$\eta_t$及初始条件$(y_0 - \eta_0)$之和。

定义 称$\{\boldsymbol{y}_t\}$为 n 维**线性向量**I(0)**过程**，如果$\boldsymbol{y}_t = \boldsymbol{\delta} + \boldsymbol{u}_t$，其中$\boldsymbol{\delta}$为常数向量，$\boldsymbol{u}_t = \boldsymbol{\psi}(L)\boldsymbol{\varepsilon}_t$，多维滤波$\boldsymbol{\psi}(L) \equiv \boldsymbol{I}_n + \boldsymbol{\psi}_1 L + \boldsymbol{\psi}_2 L^2 + \cdots$，$\{\boldsymbol{\varepsilon}_t\}$为独立同分布的，$\mathrm{E}(\boldsymbol{\varepsilon}_t) = \boldsymbol{0}$，$\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$为正定矩阵，$\{\boldsymbol{\psi}_j\}$为 OS(即矩阵的每个元素均为 OS)，而且$\boldsymbol{\psi}(1) = \boldsymbol{I}_n + \boldsymbol{\psi}_1 + \boldsymbol{\psi}_2 + \cdots \neq \boldsymbol{0}_{n\times n}$。

因此，可将 n 维向量I(1)过程$\boldsymbol{y}_t$表示为

$$\Delta \boldsymbol{y}_t = \boldsymbol{\delta} + \boldsymbol{u}_t = \boldsymbol{\delta} + \boldsymbol{\psi}(L)\boldsymbol{\varepsilon}_t$$

这称为I(1)系统的“向量移动平均表示法”(VMA Representation)。将其写成绝对水平的形式：

$$\boldsymbol{y}_t = \boldsymbol{y}_0 + \boldsymbol{\delta} t + \sum_{s=1}^{t} \boldsymbol{u}_s$$

故有向量形式的 Beveridge-Nelson 分解公式：

$$\boldsymbol{y}_t = \underbrace{\boldsymbol{\delta} t}_{\text{time trend}} + \underbrace{\boldsymbol{\psi}(1)\sum_{s=1}^{t} \boldsymbol{\varepsilon}_s}_{\text{random walk}} + \underbrace{\boldsymbol{\eta}_t}_{\text{stationary}} + \underbrace{(\boldsymbol{y}_0 - \boldsymbol{\eta}_0)}_{\text{initial condition}}$$

21.10　协整的定义与最大似然估计

一般地，由于随机趋势项$\boldsymbol{\psi}(1)\sum_{s=1}^{t}\boldsymbol{\varepsilon}_s$，上述 I(1)系统$\{\boldsymbol{y}_t\}$非平稳。

如果在方程两边同乘某$1\times n$的非零行向量$\boldsymbol{a}'$，则可能将此随机趋势项消去：

$$\boldsymbol{a}'\boldsymbol{y}_t = \boldsymbol{a}'\boldsymbol{\delta}t + \boldsymbol{a}'\boldsymbol{\psi}(1)\sum_{s=1}^{t}\boldsymbol{\varepsilon}_s + \boldsymbol{a}'\boldsymbol{\eta}_t + \boldsymbol{a}'(\boldsymbol{y}_0 - \boldsymbol{\eta}_0)$$

如果$\underbrace{\boldsymbol{a}'}_{1\times n}\underbrace{\boldsymbol{\psi}(1)}_{n\times n} = \underbrace{\boldsymbol{0}'}_{1\times n}$，则上式可简化为

$$\boldsymbol{a}'\boldsymbol{y}_t = \boldsymbol{a}'\boldsymbol{\delta}t + \boldsymbol{a}'\boldsymbol{\eta}_t + \boldsymbol{a}'(\boldsymbol{y}_0 - \boldsymbol{\eta}_0)$$

$\{\boldsymbol{a}'\boldsymbol{y}_t\}$是趋势平稳(trend stationary)的过程，即只要将时间趋势项$\boldsymbol{a}'\delta t$去掉，就是平稳过程。

此时，称$\boldsymbol{y}_t$是“协整的”(cointegrated)，而称$\boldsymbol{a}$为“协整向量”(cointegrating vector)。

是否存在 n 维非零列向量$\boldsymbol{a}$，使得$\boldsymbol{a}'\boldsymbol{\psi}(1)=\boldsymbol{0}'$，即$\boldsymbol{\psi}(1)'\boldsymbol{a}=\boldsymbol{0}$？

考虑关于$\boldsymbol{a}$的 n 元方程组$\boldsymbol{\psi}(1)'\boldsymbol{a}=\boldsymbol{0}$，称其线性无关解的个数为$\{\boldsymbol{y}_t\}$的“协整秩”(cointegration rank)，即线性无关的协整向量的个数，记为 h。

根据线性代数知识，$h=n-\operatorname{rank}[\boldsymbol{\psi}(1)]$。由于$1\le\operatorname{rank}[\boldsymbol{\psi}(1)]\le n$，故$0\le h\le n-1$。

如果$h=0$，则不存在协整关系。

如果$h=1$，则存在一个协整关系，可解释为长期均衡关系(long-run equilibrium)。

如果$h>1$，则存在多个协整关系，通常需用经济理论来剔除不合理的协整向量，而将最合理的协整向量解释为长期均衡关系。

不失一般性，假设协整向量$\boldsymbol{a}$的第一个分量不为 0，并将其标准化为 1，即$\boldsymbol{a}=\begin{pmatrix} 1 \\ \underbrace{-\boldsymbol{\gamma}}_{(n-1)\times 1} \end{pmatrix}$。将$\boldsymbol{y}_t$也同样地分块，即$\boldsymbol{y}_t=\begin{pmatrix} y_{1t} \\ \underbrace{\boldsymbol{y}_{2t}}_{(n-1)\times 1} \end{pmatrix}$。

假设$\boldsymbol{a}'\boldsymbol{\delta}=0$(不存在时间趋势项)，可将方程写成

$$y_{1t} = \alpha + \boldsymbol{\gamma}' \boldsymbol{y}_{2t} + z_t$$

其中，$\alpha \equiv (1 - \boldsymbol{\gamma}')(\boldsymbol{y}_0 - \boldsymbol{\eta}_0)$ 为初始条件，可视为截距项；而 $z_t \equiv (1 - \boldsymbol{\gamma}')\boldsymbol{\eta}_t$ 为平稳过程，可视为扰动项。此回归称为“协整回归”(cointegrating regression)。

在 $\boldsymbol{y}_t$ 存在协整关系的前提下，OLS 估计是一致的。

以上为 I(1)系统的 VMA 表示法。由于扰动项不可观测，为进行 MLE 估计，考虑以下的“向量自回归表示法”(VAR Representation)：

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\delta} t + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

此 VAR(p)何时为协整系统呢？先导出其对应的 VMA 表示法。

定义$\boldsymbol{\Phi}(L) \equiv \boldsymbol{I}_n - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 - \cdots - \boldsymbol{\Phi}_p L^p$，则

$$\boldsymbol{\Phi}(L)\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\delta} t + \boldsymbol{\varepsilon}_t$$

在方程两边左乘$\Phi(L)^{-1}$可得

$$\boldsymbol{y}_t = \boldsymbol{\Phi}(L)^{-1}\boldsymbol{\alpha} + \boldsymbol{\Phi}(L)^{-1}\boldsymbol{\delta} t + \boldsymbol{\Phi}(L)^{-1}\boldsymbol{\varepsilon}_t$$

对方程两边差分可得

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Phi}(L)^{-1}\boldsymbol{\delta}(1-L)t + \boldsymbol{\Phi}(L)^{-1}(1-L)\boldsymbol{\varepsilon}_t = \boldsymbol{\delta}^* + \boldsymbol{\psi}(L)\boldsymbol{\varepsilon}_t$$

其中，$(1-L)t = t-(t-1) = 1$，$\boldsymbol{\delta}^* \equiv \boldsymbol{\Phi}(1)^{-1}\boldsymbol{\delta}$，$\boldsymbol{\psi}(L) \equiv \boldsymbol{\Phi}(L)^{-1}(1-L)$。

如果此 VAR 系统为协整秩为 h 的协整系统，则$\boldsymbol{\psi}(L)$必须为 OS，

且$\operatorname{rank}\left[\boldsymbol{\psi}(1)\right]=n-h$。

由于$\boldsymbol{\psi}(L)\equiv\boldsymbol{\Phi}(L)^{-1}(1-L)$，故$\Phi(L)\boldsymbol{\psi}(L)=(1-L)\boldsymbol{I}_n$。令 $L=1$，则有

$$\boldsymbol{\Phi}(1)_{n\times n}\boldsymbol{\psi}(1)_{n\times n}=\mathbf{0}_{n\times n}$$

可以证明，如果$\{\boldsymbol{y}_t\}$的协整秩为 h，则$\operatorname{rank}\left[\Phi(1)\right]=h$。根据线性代数知识，可以将$\Phi(1)$分解为

$$\boldsymbol{\Phi}(1)_{n\times n}=\boldsymbol{B}_{n\times h}\left(\boldsymbol{A}'\right)_{h\times n}$$

其中，$\boldsymbol{B}, \boldsymbol{A}$为两个$n\times h$的满列秩矩阵($\boldsymbol{B}, \boldsymbol{A}$不唯一)。此条件称为“降秩条件” (reduced rank condition)。因此，

$$\boldsymbol{B}_{n\times h}\left(\boldsymbol{A}'\right)_{h\times n}\boldsymbol{\Psi}(1)_{n\times n}=\mathbf{0}_{n\times n}$$

因为$\boldsymbol{B}_{n\times h}$满列秩，故

$$\boldsymbol{A}'_{h\times n}\boldsymbol{\psi}(1)_{n\times n}=\mathbf{0}_{h\times n}$$

故矩阵$\boldsymbol{A}_{n\times h}$中的列向量都是协整向量(满足协整向量的定义)。

从 VAR 方程出发，可导出对应的“向量误差修正表示法”(VECM Representation)。

根据与推导 ADF 检验类似的方法可得

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\delta} t + (\boldsymbol{\rho} - \boldsymbol{I}_n)\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

其中，$\boldsymbol{\rho} \equiv \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 + \cdots + \boldsymbol{\Phi}_p$，$\boldsymbol{\Gamma}_s \equiv -(\boldsymbol{\Phi}_{s+1} + \cdots + \boldsymbol{\Phi}_p),\ s = 1, 2, \cdots, p-1$。

由于$\boldsymbol{I}_n - \boldsymbol{\rho} = \boldsymbol{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \cdots - \boldsymbol{\Phi}_p = \boldsymbol{\Phi}(1) = \boldsymbol{B}\boldsymbol{A}'$，故

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\delta} t - \boldsymbol{B} \underbrace{\boldsymbol{A}' \boldsymbol{y}_{t-1}}_{\boldsymbol{z}_{t-1}} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_2 \Delta \boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

其中，$\boldsymbol{z}_{t-1} \equiv \boldsymbol{A}' \boldsymbol{y}_{t-1}$为误差修正项(因为$\boldsymbol{A}$中的列向量皆为协整向量)。一个协整的 I(1)系统同时有 VMA，VAR 与 VECM 表示法，此结论称为“格兰杰表示法定理”(Granger Representation Theorem)。

定义$\boldsymbol{\Gamma}_0 \equiv -\boldsymbol{\Phi}(1) = -\boldsymbol{B}\boldsymbol{A}'$。

Johansen(1988)使用 MLE 来估计如下 VECM 模型：

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

为简化推导，假定没有时间趋势项。

假设样本容量为 $T+p$，即观测数据为 $\left\{\boldsymbol{y}_{-p+1}, \boldsymbol{y}_{-p+2}, \cdots, \boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_T\right\}$。

只有当$\boldsymbol{y}_t$存在协整关系时，此 VECM 模型才能成立；否则，方程左边平稳，而右边非平稳 ($\boldsymbol{\Gamma}_0 \boldsymbol{y}_{t-1}$不平稳)。

假设协整秩为 h，则系数矩阵$\boldsymbol{\Gamma}_0$须满足约束条件“$\operatorname{rank}(\boldsymbol{\Gamma}_0) = h$”。

在满足“$\mathrm{rank}(\boldsymbol{\Gamma}_0)=h$”以及给定$\{\boldsymbol{y}_{-p+1}, \boldsymbol{y}_{-p+2}, \cdots, \boldsymbol{y}_0\}$的条件下，最大化$\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T\}$的对数似然函数(即条件 MLE)。

假设扰动项为 n 维正态分布，即$\boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Omega})$，而且$\{\boldsymbol{\varepsilon}_t\}$为独立同分布的。$n$ 维随机向量$\boldsymbol{\varepsilon}_t$的联合密度为

$$\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{\Omega}\right|^{1/2}}\exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t\right\}$$

取对数可得

$$-\frac{n}{2}\ln 2\pi-\frac{1}{2}\ln\left|\boldsymbol{\Omega}\right|-\frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t$$

记$\boldsymbol{\Pi} \equiv (\boldsymbol{\alpha} \;\; \boldsymbol{\Gamma}_0 \;\; \boldsymbol{\Gamma}_1 \cdots \boldsymbol{\Gamma}_{p-1})'$，$\boldsymbol{x}_t \equiv (1 \; \boldsymbol{y}_{t-1} \Delta \boldsymbol{y}_{t-1} \cdots \Delta \boldsymbol{y}_{t-p+1})'$，

则$\boldsymbol{\varepsilon}_t = \Delta \boldsymbol{y}_t - \boldsymbol{\Pi}' \boldsymbol{x}_t$，故$\{\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_T\}$所对应的对数条件似然函数为

$$\begin{aligned} \max \quad & -\frac{nT}{2}\ln 2\pi - \frac{T}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\sum_{t=1}^{T}\left[(\Delta \boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)' \boldsymbol{\Omega}^{-1} (\Delta \boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)\right] \\ s.t. \quad & \operatorname{rank}(\boldsymbol{\Gamma}_0) = h \end{aligned}$$

其中，n 为该系统中变量的个数，T 为样本容量。

为求解此约束极值问题，须确定协整秩 h。当协整秩为 h 时，系数矩阵$\boldsymbol{\Gamma}_0$有 h 个自由(线性无关)的列向量。

协整秩 h 越大，则对矩阵Γ_0的约束越少，对应的似然函数最大值越大。故可进行以下似然比检验：

$$H_0: \text{rank}(\Gamma_0)=0 \quad vs \quad H_1: \text{rank}(\Gamma_0)>0$$

由于矩阵Γ_0的秩取决于其非零特征值的个数，故检验统计量被称为“迹统计量”，记为λ_{trace}。

由于“迹检验”(trace test)是似然比检验，故为单边右侧检验，即λ_{trace}越大(加上H_0约束后，似然函数的最大值下降越多)，则越倾向于拒绝原假设。

如果接受“$H_0: \text{rank}(\Gamma_0)=0$”，则认为不存在协整关系。反之，则继续检验是否存在多个协整关系：

$$H_0:\mathrm{rank}(\boldsymbol{\Gamma}_0)=1 \quad vs \quad H_1:\mathrm{rank}(\boldsymbol{\Gamma}_0)>1$$

依此顺序不断进行检验，直到接受H_0，确认协整秩 h 为止。

确认协整秩 h 后，可用条件 MLE 来估计 VECM 模型中的所有参数，包括长期参数(协整系数)与短期参数。

Johansen 还考虑了另一类检验：

$$H_0:\mathrm{rank}(\boldsymbol{\Gamma}_0)=p \quad vs \quad H_1:\mathrm{rank}(\boldsymbol{\Gamma}_0)=p+1$$

其检验统计量为“最大特征值统计量”(maximum eigenvalue statistics)，记为$\lambda_{\max}$，称为“最大特征值检验”。