

# 第十一讲 时间序列模型

- 一、平稳与非平稳
- 二、长期均衡关系与协整
- 三、协整检验
- 四、误差修正模型
- 五、案例分析

# 一、平稳与非平稳

# 1、单位根过程

为了说明单位根过程的概念，我们侧重以**AR(1)**模型进行分析：

$$Y_t = \varphi Y_{t-1} + \varepsilon_t$$

根据平稳时间序列分析的理论可知，当  $|\varphi| < 1$  时，该序列  $\{Y_t\}$  是平稳的,此模型是经典的 **Box-Jenkins**时间序列**AR(1)**模型。

当  $\varphi = 1$ ，则序列的生成过程变为如下随机游动过程  
(Random Walk Process):

$$Y_t = Y_{t-1} + \varepsilon_t$$

其中  $\{\varepsilon_t\}$  独立同分布且均值为零、方差恒定为  $\sigma^2$ 。随机游动过程的方差为:

$$\begin{aligned}\text{Var}(Y_t) &= \text{Var}(Y_{t-1} + \varepsilon_t) \\ &= \text{Var}(Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t) \\ &= \text{Var}(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{t-1} + \varepsilon_t) \\ &= t\sigma^2\end{aligned}$$

当  $t \rightarrow \infty$  时，序列的方差趋于无穷大，说明随机游动过程是非平稳的。

# 单位根过程

如果一个序列是随机游动过程，则称这个序列是一个“单位根过程”。

为什么称为“单位根过程”？

将一阶自回归模型表示成如下形式：

$$Y_t - \phi Y_{t-1} = \varepsilon_t \quad \text{或} \quad (1 - \phi L)Y_t = \varepsilon_t$$

其中， $L$  是滞后算子，即  $LY_t = Y_{t-1}$

根据模型的滞后多项式  $1 - \varphi L$ ，可以写出对应的线性方程： $1 - \varphi Z = 0$ （通常称为特征方程）

该方程的根为： $Z = 1/\varphi$ 。

当  $|\varphi| < 1$  时序列是平稳的，特征方程的根满足条件： $|Z| > 1$

当  $\varphi = 1$  时，序列的生成过程变为随机游动过程，对应特征方程的根  $Z = 1$ ，所以通常称序列含有单位根，或者说序列的生成过程为“单位根过程”。

结论：

随机游动过程是非平稳的。

因此，检验序列的非平稳性就变为检验特征方程是否有单位根，这就是单位根检验方法的由来。

从单位根过程的定义可以看出，含一个单位根的过程，其一阶差分：

$$\Delta Y_t = Y_t - Y_{t-1} = u_t$$

是一平稳过程，像这种经过一次差分后变为平稳的序列称为一阶单整序列(Integrated Process)，

记为  $\{Y_{t0}\} \sim I(1)$



有时，一个序列经一次差分后可能还是非平稳的，如果序列经过二阶差分后才变成平稳过程，则称序列  $\{Y_t\}$  为二阶单整序列，记为  $\{Y_t\} \sim I(2)$ 。

一般地，如果序列  $\{Y_t\}$  经过  $d$  次差分后平稳，而  $d-1$  次差分却不平稳，那么称为

$d$  阶单整序列，记为  $\{Y_t\} \sim I(d)$ ， $d$  称为整形阶数。特别地，若序列  $\{Y_t\}$  本身是平稳的，则称序列为零阶单整序列，记为  $\{Y_t\} \sim I(0)$ 。

## 2、Dickey-Fuller检验（DF检验）

大多数经济变量呈现出强烈的趋势特征。这些具有趋势特征的经济变量，当发生经济振荡或冲击后，一般会出现两种情形：

- 受到振荡或冲击后，经济变量逐渐又回它们的长期趋势轨迹；

- 这些经济变量没有回到原有轨迹，而呈现出随机游走的状态。

若我们研究的经济变量遵从一个非平稳过程，一个变量对其他变量的回归可能会导致伪回归结果。这是研究单位根检验的重要意义所在。

假设数据序列是由下列自回归模型生成的：

$$Y_t = \gamma Y_{t-1} + \varepsilon_t$$

其中， $\varepsilon_t$  独立同分布，期望为零，方差为  $\sigma^2$ ，我们要检验该序列是否含有单位根。检验的原假设为： $H_0 : \gamma = 1$

回归系数的OLS估计为： $\hat{\gamma} = \frac{\sum y_{t-1} y_t}{\sum y_{t-1}^2}$

检验所用的统计量为： $t = \frac{\hat{\gamma} - \gamma}{\hat{\sigma}_{\hat{\gamma}}}$

在  $H_0: \gamma = 1$  成立的条件下, t统计量为:

$$t = \frac{\hat{\gamma} - 1}{\hat{\sigma}_{\hat{\rho}}}$$

**Dickey、Fuller**通过研究发现, 在原假设成立的情况下, 该统计量不服从t分布。所以传统的t检验法失效。

但可以证明, 上述统计量的极限分布存在, 一般称其为**Dickey-Fuller**分布。根据这一分布所作的检验称为**DF**检验, 为了区别, t 统计量的值有时也称为  $\tau$  值。

**Dickey、Fuller**得到**DF**检验的临界值，并编制了**DF**检验临界值表供查。在进行**DF**检验时，比较**t**统计量值与**DF**检验临界值，就可在某个显著性水平上拒绝或接受原假设。

在实际应用中，可按如下检验步骤进行：

$$Y_t = \gamma Y_{t-1} + \varepsilon_t$$

(1) 根据观察数据，用**OLS**法估计一阶自回归模型，得到回归系数的**OLS**估计：

$$\hat{\gamma} = \frac{\sum y_{t-1} y_t}{\sum y_{t-1}^2}$$

(2) 提出假设  $H_0 : \gamma = 1$   $H_1 : \gamma \neq 1$   
检验用统计量为常规t统计量,

$$t = \frac{\hat{\gamma} - \gamma}{\hat{\sigma}_{\hat{\gamma}}}$$

(3) 计算在原假设成立的条件下t统计量值，查DF检验临界值表得临界值，然后将t统计量值与DF检验临界值比较：

若t统计量值小于DF检验临界值，则拒绝原假设，说明序列不存在单位根；

若t统计量值大于或等于DF检验临界值，则接受原假设，说明序列存在单位根。

Dickey、Fuller研究发现，DF检验的临界值同序列的数据生成过程以及回归模型的类型有关，因此他们针对如下三种方程编制了临界值表，后来Mackinnon把临界值表加以扩充，形成了目前使用广泛的临界值表，在EViews软件中使用的是Mackinnon临界值表。

这三种模型如下：

**模型I：**  $Y_t = \gamma Y_{t-1} + \varepsilon_t$

**模型II：**  $Y_t = \alpha + \gamma Y_{t-1} + \varepsilon_t$

**模型III：**  $Y_t = \alpha + \beta t + \gamma Y_{t-1} + \varepsilon_t$



### 3、Augmented Dickey-Fuller检验 (ADF检验)

DF检验存在的问题是，在检验所设定的模型时，假设随机扰动项不存在自相关。但大多数的经济数据序列是不能满足此项假设的，当随机扰动项存在自相关时，直接使用DF检验法会出现偏误，为了保证单位根检验的有效性，人们对DF检验进行拓展，从而形成了扩展的DF检验(Augmented Dickey-Fuller Test)，简称为ADF检验。

假设基本模型为如下三种类型：

模型I:  $Y_t = \gamma Y_{t-1} + \mu_t$

模型II:  $Y_t = \alpha + \gamma Y_{t-1} + \mu_t$

模型III:  $Y_t = \alpha + \beta t + \gamma Y_{t-1} + \mu_t$

其中  $\mu_t$  为随机扰动项，它可以是一个一般的平稳过程。

为了借用**DF**检验的方法，将模型变为如下式：

**模型I:** 
$$Y_t = \gamma Y_{t-1} + \sum_{i=1}^p \alpha_i \Delta Y_{t-i} + \varepsilon_t$$

**模型II:** 
$$Y_t = \alpha + \gamma Y_{t-1} + \sum_{i=1}^p \alpha_i \Delta Y_{t-i} + \mu_t$$

**模型III:** 
$$Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{i=1}^p \alpha_i \Delta Y_{t-i} + \mu_t$$

可以证明，在上述模型中检验原假设的t统计量的极限分布，与**DF**检验的极限分布相同，从而可以使用相同的临界值表，这种检验称为**ADF**检验。

# 例.1

根据《中国统计年鉴2004》，得到我国1978—2003年的GDP序列(如表10.1)，检验其是否为平稳序列。

表10.1 中国1978—2003年度GDP序列

年度↵	GDP↵	年度↵	GDP↵	年度↵	GDP↵
1978↵	3624.1↵	1987↵	11962.5↵	1996↵	67884.6↵
1979↵	4038.2↵	1988↵	14928.3↵	1997↵	74462.6↵
1980↵	4517.8↵	1989↵	16909.2↵	1998↵	79395.7↵
1981↵	4862.4↵	1990↵	18547.9↵	1999↵	82067.5↵
1982↵	5294.7↵	1991↵	21617.8↵	2000↵	89468.1↵
1983↵	5934.5↵	1992↵	26638.1↵	2001↵	97314.8↵
1984↵	7171↵	1993↵	34634.4↵	2002↵	105172.3↵
1985↵	8964.4↵	1994↵	46759.4↵	2003↵	116898.4↵
1986↵	10202.2↵	1995↵	58478.1↵	↵	↵

## 时序图见图10.1

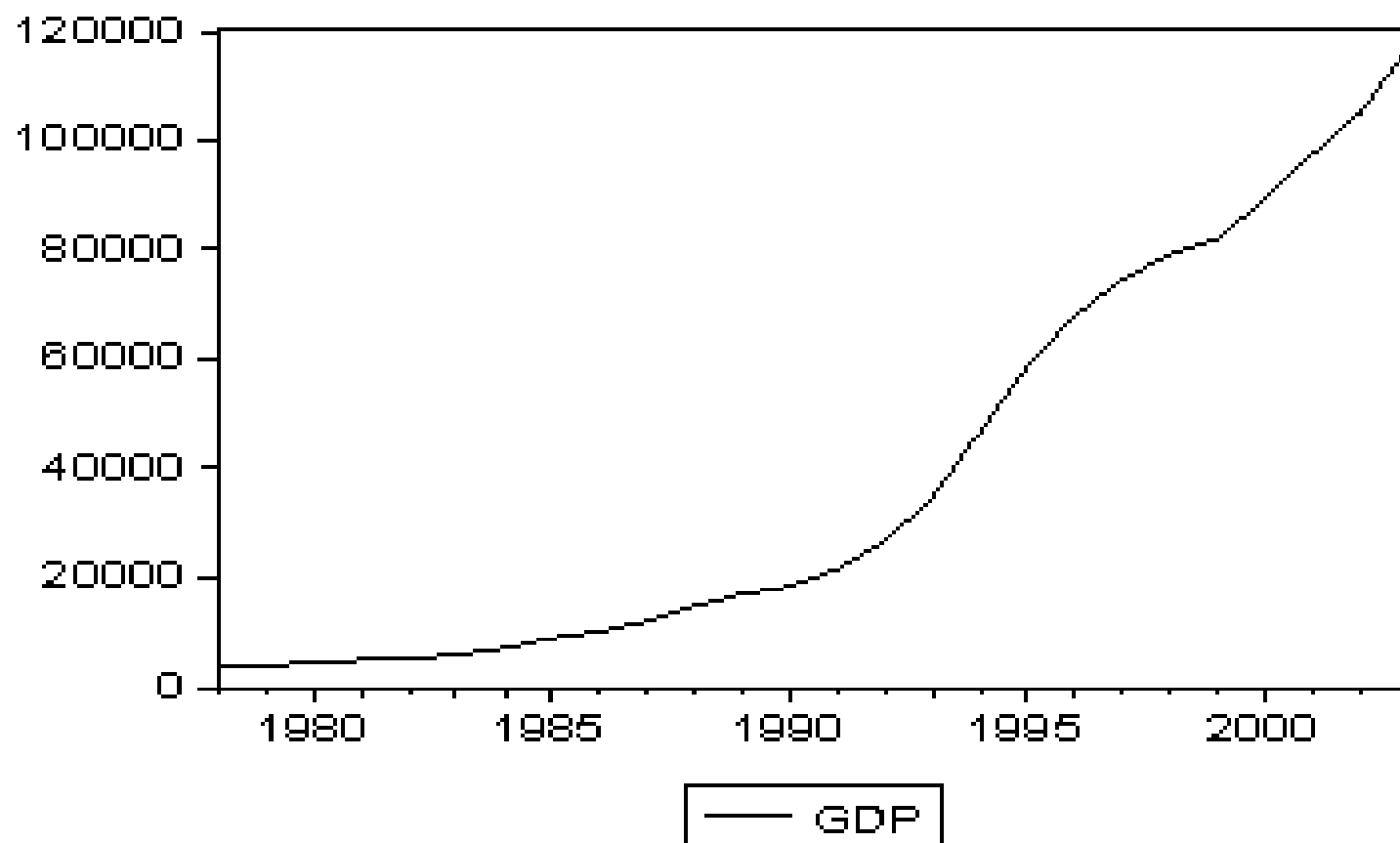


图 10.1 GDP 时序图

由GDP时序图可以看出，该序列可能存在趋势项，因此选择ADF检验的第三种模型进行检验。估计结果如下：

$$\begin{aligned}\Delta \hat{\text{GDP}}_t = & -1565.141 + 355.62 t - 0.02883 \text{GDP}_t \\ & + 1.016 \Delta \text{GDP}_t - 0.460382 \Delta \text{GDP}_t\end{aligned}$$

在原假设下，单位根的t检验统计量的值为

$$t = \frac{\hat{\gamma} - \gamma}{\hat{\sigma}_{\hat{\gamma}}} = \frac{-0.028830}{0.036679} = -0.786011$$

在1%、5%、10%三个显著性水平下，单位根检验的Mackinnon临界值分别为-4.4167、  
-3.6219、-3.2474，显然，上述t检验统计量值大于相应临界值，从而不能拒绝，表明我国1978——2003年度GDP序列存在单位根，是非平稳序列。

## 二、长期均衡关系与协整



# 0、问题的提出

- **经典回归模型（classical regression model）**是建立在稳定数据变量基础上的，对于非稳定变量，不能使用经典回归模型，否则会出现**虚假回归**等诸多问题。
- 由于许多经济变量是非稳定的，这就给经典的回归分析方法带来了很大限制。
- 但是，如果变量之间有着长期的稳定关系，**即它们之间是协整的（cointegration）**，则是可以使用经典回归模型方法建立回归模型的。
- **例如**，中国居民**人均消费水平**与**人均GDP**变量的例子中：  
因果关系回归模型要比ARMA模型有更好的预测功能，**其原因在于**，从经济理论上说，人均GDP决定着居民人均消费水平，而且它们之间有着长期的稳定关系，即它们之间是协整的（cointegration）。

# 1、长期均衡

**经济理论指出**，某些经济变量间确实存在着长期均衡关系，这种均衡关系意味着经济系统不存在破坏均衡的内在机制，如果变量在某时期受到干扰后偏离其长期均衡点，则均衡机制将会在下一期进行调整以使其重新回到均衡状态。

假设X与Y间的长期“均衡关系”由式描述

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$$

式中： $\mu_t$ 是随机扰动项。

**该均衡关系意味着**：给定X的一个值，Y相应的均衡值也随之确定为  $\alpha_0 + \alpha_1 X$ 。

在t-1期末，存在下述三种情形之一：

(1) Y等于它的均衡值： $Y_{t-1} = \alpha_0 + \alpha_1 X_t$ ；

(2) Y小于它的均衡值： $Y_{t-1} < \alpha_0 + \alpha_1 X_t$ ；

(3) Y大于它的均衡值： $Y_{t-1} > \alpha_0 + \alpha_1 X_t$ ；

在时期t，假设X有一个变化量 $\Delta X_t$ ，如果变量X与Y在时期t与t-1末期仍满足它们间的长期均衡关系，则Y的相应变化量由式给出：

$$\Delta Y_t = \alpha_1 \Delta X_t + v_t$$

式中， $v_t = \mu_t - \mu_{t-1}$ 。

## 实际情况往往并非如此

如果 $t-1$ 期末，发生了上述第二种情况，即 $Y$ 的值小于其均衡值，则 $Y$ 的变化往往会比第一种情形下 $Y$ 的变化 $\Delta Y_t$ 大一些；

反之，如果 $Y$ 的值大于其均衡值，则 $Y$ 的变化往往会小于第一种情形下的 $\Delta Y_t$ 。

可见，如果 $Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$ 正确地提示了 $X$ 与 $Y$ 间的长期稳定的“均衡关系”，则意味着 $Y$ 对其均衡点的偏离从本质上说是“临时性”的。

因此，一个重要的假设就是：随机扰动项 $\mu_t$ 必须是平稳序列。

显然，如果 $\mu_t$ 有随机性趋势（上升或下降），则会导致 $Y$ 对其均衡点的任何偏离都会被长期累积下来而不能被消除。

式 $\mathbf{Y_t}=\alpha_0+\alpha_1\mathbf{X_t}+\mu_t$ 中的随机扰动项也被称为**非均衡误差**（**disequilibrium error**），它是变量X与Y的一个线性组合：

$$\mu_t = Y_t - \alpha_0 - \alpha_1 X_t \quad (*)$$

因此，如果 $\mathbf{Y_t}=\alpha_0+\alpha_1\mathbf{X_t}+\mu_t$ 式所示的X与Y间的长期均衡关系正确的话，（\*）式表述的非均衡误差应是一平稳时间序列，并且具有零期望值，即是具有0均值的I(0)序列。

**从这里已看到**，非稳定的时间序列，它们的线性组合也可能成为平稳的。

例如：假设 $\mathbf{Y_t}=\alpha_0+\alpha_1\mathbf{X_t}+\mu_t$ 式中的X与Y是I(1)序列，如果该式所表述的它们间的长期均衡关系成立的话，则意味着由非均衡误差（\*）式给出的线性组合是I(0)序列。这时我们**称**变量X与Y是协整的（**cointegrated**）。

## 2.协整

如果序列  $\{X_{1t}, X_{2t}, \dots, X_{kt}\}$  都是  $d$  阶单整, 存在向量  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ , 使得

$$Z_t = \alpha X^T \sim I(d-b)$$

其中,  $b > 0$ ,  $X = (X_{1t}, X_{2t}, \dots, X_{kt})^T$ , 则认为序列  $\{X_{1t}, X_{2t}, \dots, X_{kt}\}$  是  $(d, b)$  阶协整, 记为  $X_t \sim CI(d, b)$ ,  $\alpha$  为协整向量 (cointegrated vector)。

在中国居民人均消费与人均GDP的例中, 该两序列都是2阶单整序列, 而且可以证明它们有一个线性组合构成的新序列为0阶单整序列, 于是认为该两序列是(2,2)阶协整。

**由此可见:** 如果两个变量都是单整变量, 只有当它们的单整阶数相同时, 才可能协整; 如果它们的单整阶数不相同, 就不可能协整。

三个以上的变量，如果具有不同的单整阶数，有可能经过线性组合构成低阶单整变量。

例如，如果存在：

$$W_t \sim I(1), V_t \sim I(2), U_t \sim I(2)$$

并且

$$P_t = aV_t + bU_t \sim I(1)$$

$$Q_t = cW_t + eP_t \sim I(0)$$

那么认为：

$$V_t, U_t \sim CI(2,1)$$

$$W_t, P_t \sim CI(1,1)$$

从协整的定义可以看出：

(d,d) 阶协整是一类非常重要的协整关系，**它的经济意义在于：**两个变量，虽然它们具有各自的长期波动规律，但是如果它们是 (d, d) 阶协整的，则它们之间存在着一个长期稳定的比例关系。

**例如：**前面提到的中国CPC和GDPPC，它们各自**都是2阶单整**，**并且将会看到，它们是(2, 2)阶协整**，说明它们之间存在着一个长期稳定的比例关系，从计量经济学模型的含义上讲，建立如下居民人均消费函数模型

$$CPC_t = \alpha_0 + \alpha_1 GDPPC_t + \mu_t$$

变量选择是合理的，随机误差项一定是“白噪声”（即均值为0，方差不变的稳定随机序列），模型参数有合理的经济解释。

这也解释了尽管这两时间序列是非稳定的，但却可以用经典的回归分析方法建立回归模型的原因。



- 从这里，我们已经初步认识到：检验变量之间的协整关系，在建立计量经济学模型中是非常重要的。

而且，从变量之间是否具有协整关系出发选择模型的变量，其数据基础是牢固的，其统计性质是优良的。

### 三、协整检验

# 1、两变量的Engle-Granger检验

为了检验两变量 $Y_t, X_t$ 是否为协整，Engle和Granger于1987年提出两步检验法，也称为EG检验。

**第一步**，用OLS方法估计方程  $Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$

并计算非均衡误差，得到：

$$\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_t$$

$$\hat{e}_t = Y_t - \hat{Y}_t$$

称为**协整回归(cointegrating)**或**静态回归(static regression)**。

**第二步**，检验 $\hat{e}_t$ 的单整性。如果 $\hat{e}_t$ 为稳定序列，则认为变量 $Y_t, X_t$ 为(1,1)阶协整；如果 $\hat{e}_t$ 为1阶单整，则认为变量 $Y_t, X_t$ 为(2,1)阶协整；...

$\hat{e}_t$ 的单整性的检验方法仍然是DF检验或者ADF检验。

由于协整回归中已含有截距项，则检验模型中无需再用截距项。如使用模型1

$$\Delta e_t = \delta e_{t-1} + \sum_{i=1}^p \theta_i \Delta e_{t-i} + \varepsilon_t$$

进行检验时，拒绝零假设 $H_0: \delta=0$ ，意味着误差项 $e_t$ 是平稳序列，从而说明X与Y间是协整的。

**需要注意的是**，这里的DF或ADF检验是针对协整回归计算出的误差项  $\hat{e}_t$ 而非真正的非均衡误差 $\mu_t$ 进行的。

而OLS法采用了残差最小平方和原理，因此估计量 $\delta$ 是向下偏倚的，这样将导致拒绝零假设的机会比实际情形大。

于是对 $e_t$ 平稳性检验的DF与ADF临界值应该比正常的DF与ADF临界值还要小。

- MacKinnon(1991)通过模拟试验给出了协整检验的临界值，表9.3.1是双变量情形下不同样本容量的临界值。

表 9.3.1 双变量协整 ADF 检验临界值

样本容量	显 著 性 水 平		
	0.01	0.05	0.10
25	-4.37	-3.59	-3.22
50	-4.12	-3.46	-3.13
100	-4.01	-3.39	-3.09
$\infty$	-3.90	-3.33	-3.05

- **例9.3.1** 检验中国居民人均消费水平CPC与人均国内生产总值GDPPC的协整关系。

在前文已知CPC与GDPPC都是I(2)序列，而 § 2.10中已给出了它们的回归式

$$CPC_t = 49.764106 + 0.45831GDPPC_t \quad R^2=0.9981$$

通过对该式计算的残差序列作ADF检验，得适当检验模型

$$\Delta \hat{e}_t = -1.55 \hat{e}_{t-1} + 1.49 \Delta \hat{e}_{t-1} + 2.27 \Delta \hat{e}_{t-3}$$

$$(-4.47) \quad (3.93) \quad (3.05)$$

$$LM(1)=0.00 \quad LM(2)=0.00$$

$t=-4.47 < -3.75 = ADF_{0.05}$ ，拒绝存在单位根的假设，残差项是稳定的，因此中国居民人均消费水平与人均GDP是(2, 2)阶协整的，说明了该两变量间存在长期稳定的“均衡”关系。

## 2、多变量协整关系的检验—扩展的E-G检验

多变量协整关系的检验要比双变量复杂一些，主要在于协整变量间可能存在多种稳定的线性组合。

假设有4个I(1)变量Z、X、Y、W，它们有如下的长期均衡关系：

$$Z_t = \alpha_0 + \alpha_1 W_t + \alpha_2 X_t + \alpha_3 Y_t + \mu_t \quad (*)$$

其中，非均衡误差项 $\mu_t$ 应是I(0)序列：

$$\mu_t = Z_t - \alpha_0 - \alpha_1 W_t - \alpha_2 X_t - \alpha_3 Y_t \quad (**)$$

- 然而，如果Z与W，X与Y间分别存在长期均衡关系：

$$Z_t = \beta_0 + \beta_1 W_t + v_{1t}$$

$$X_t = \gamma_0 + \gamma_1 Y_t + v_{2t}$$

则非均衡误差项 $v_{1t}$ 、 $v_{2t}$ 一定是稳定序列I(0)。于是它们的任意线性组合也是稳定的。例如

$$v_t = v_{1t} + v_{2t} = Z_t - \beta_0 - \gamma_0 - \beta_1 W_t + X_t - \gamma_1 Y_t \quad (***)$$

一定是I(0)序列。

由于 $v_t$ 象(\*\*)式中的 $\mu_t$ 一样，也是Z、X、Y、W四个变量的线性组合，由此(\*\*\*)式也成为该四变量的另一稳定线性组合。

$(1, -\alpha_0, -\alpha_1, -\alpha_2, -\alpha_3)$ 是对应于(\*\*)式的协整向量， $(1, -\beta_0 - \gamma_0, -\beta_1, 1, -\gamma_1)$ 是对应于(\*\*\*)式的协整向量。



## 检验程序：

对于多变量的协整检验过程，基本与双变量情形相同，**即需检验变量是否具有同阶单整性，以及是否存在稳定的线性组合。**

在检验是否存在稳定的线性组合时，需通过设置一个变量为被解释变量，其他变量为解释变量，进行OLS估计并检验残差序列是否平稳。

如果不平稳，则需更换被解释变量，进行同样的OLS估计及相应的残差项检验。

当所有的变量都被作为被解释变量检验之后，仍不能得到平稳的残差项序列，则认为这些变量间不存在  $(d, d)$  阶协整。

同样地，检验残差项是否平稳的DF与ADF检验临界值要比通常的DF与ADF检验临界值小，而且该临界值还受到所检验的变量个数的影响。

表9.3.2给出了MacKinnon(1991)通过模拟试验得到的不同变量协整检验的临界值。

**表 9.3.2      多变量协整检验 ADF 临界值**

样本 容量	变量数=3 显著性水平			变量数=4 显著性水平			变量数=6 显著性水平		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
25	-4.92	-4.1	-3.71	-5.43	-4.56	-4.15	-6.36	-5.41	-4.96
50	-4.59	-3.92	-3.58	-5.02	-4.32	-3.98	-5.78	-5.05	-4.69
100	-4.44	-3.83	-3.51	-4.83	-4.21	-3.89	-5.51	-4.88	-4.56
$\infty$	-4.30	-3.74	-3.45	-4.65	-4.1	-3.81	-5.24	-4.7	-4.42

## \*2、多变量协整关系的检验—JJ检验

- **Johansen于1988年，以及与Juselius于1990年提出了一种用极大或然法进行检验的方法，通常称为JJ检验。**
- **《高等计量经济学》（清华大学出版社，2000年9月）P279-282.**
- **E-views中有JJ检验的功能。**

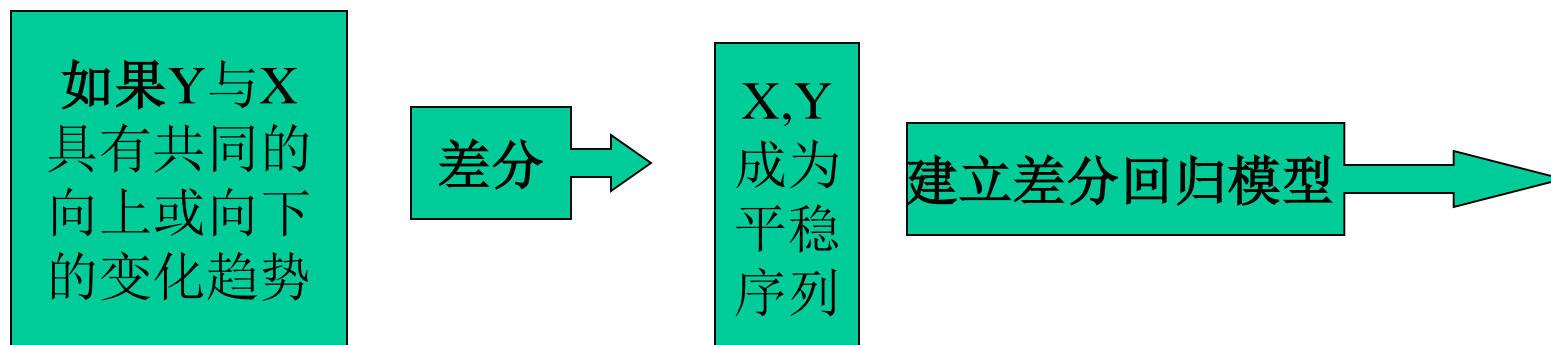
## 四、误差修正模型

# 1、误差修正模型

前文已经提到，对于非稳定时间序列，可通过差分的方法将其化为稳定序列，然后才可建立经典的回归分析模型。

**如：**建立人均消费水平（Y）与人均可支配收入（X）之间的回归模型：

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$$



$$\Delta Y_t = \alpha_1 \Delta X_t + v_t$$

式中，  $v_t = \mu_t - \mu_{t-1}$

然而，这种做法会引起两个问题：

(1) 如果X与Y间存在着长期稳定的均衡关系

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$$

且误差项 $\mu_t$ 不存在序列相关，则差分式

$$\Delta Y_t = \alpha_1 \Delta X_t + v_t$$

中的 $v_t$ 是一个一阶移动平均时间序列，因而是序列相关的；

(2) 如果采用差分形式进行估计，则关于变量水平值的重要信息将被忽略，这时模型只表达了X与Y间的短期关系，而没有揭示它们间的长期关系。

因为，从长期均衡的观点看，Y在第t期的变化不仅取决于X本身的变化，还取决于X与Y在t-1期末的状态，尤其是X与Y在t-1期的不平衡程度。

另外，使用差分变量也往往会得出不能令人满意回归方程。

例如，使用 $\Delta Y_t = \alpha_1 \Delta X_t + v_t$ 回归时，很少出现截距项显著为零的情况，即我们常常会得到如下形式的方程：

$$\Delta Y_t = \hat{\alpha}_0 + \hat{\alpha}_1 \Delta X_t + v_t \quad \hat{\alpha}_0 \neq 0 \quad (*)$$

在X保持不变时，如果模型存在静态均衡（static equilibrium），Y也会保持它的长期均衡值不变。

但如果使用（\*）式，即使X保持不变，Y也会处于长期上升或下降的过程中（Why?），这意味着X与Y间不存在静态均衡。

这与大多数具有静态均衡的经济理论假说不相符。

可见，简单差分不一定能解决非平稳时间序列所遇到的全部问题，因此，误差修正模型便应运而生。

**误差修正模型（Error Correction Model，简记为ECM）**是一种具有特定形式的计量经济学模型，它的主要形式是由Davidson、Hendry、Srba和Yeo于1978年提出的，**称为DHSY模型**。

为了便于理解，我们通过一个具体的模型来介绍它的结构。  
假设两变量X与Y的长期均衡关系为：

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$$

由于现实经济中X与Y很少处在均衡点上，因此实际观测到的只是X与Y间的短期的或非均衡的关系，假设具有如下(1, 1)阶分布滞后形式

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \mu Y_{t-1} + \varepsilon_t$$

该模型显示出第t期的Y值，不仅与X的变化有关，而且与t-1期X与Y的状态值有关。



由于变量可能是非平稳的，因此不能直接运用OLS法。  
对上述**分布滞后模型适当变形**得

$$\begin{aligned}\Delta Y_t &= \beta_0 + \beta_1 \Delta X_t + (\beta_1 + \beta_2) X_{t-1} - (1 - \mu) Y_{t-1} + \varepsilon_t \\ &= \beta_1 \Delta X_t - (1 - \mu) \left( Y_{t-1} - \frac{\beta_0}{1 - \mu} - \frac{\beta_1 + \beta_2}{1 - \mu} X_{t-1} \right) + \varepsilon_t\end{aligned}$$

或 
$$\Delta Y_t = \beta_1 \Delta X_t - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t \quad (**)$$

式中，  $\lambda = 1 - \mu$        $\alpha_0 = \beta_0 / (1 - \mu)$        $\alpha_1 = (\beta_1 + \beta_2) / (1 - \mu)$

如果将 (\*\*) 中的参数，与  $\mathbf{Y_t = \alpha_0 + \alpha_1 X_t + \mu_t}$  中的相应参数视为相等，则 (\*\*) 式中括号内的项就是t-1期的非均衡误差项。

**(\*\*) 式表明：**Y的变化决定于X的变化以及前一时期的**非均衡程度**。同时， (\*\*) 式也弥补了简单差分模型  $\Delta Y_t = \alpha_1 \Delta X_t + v_t$  的不足，因为该式含有用X、Y水平值表示的前期非均衡程度。因此，Y的值已对前期的非均衡程度作出了修正。

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda(Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t \quad (**)$$

称为一阶误差修正模型 (**first-order error correction model**)。

(\*\*) 式可以写成:

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda ecm + \varepsilon_t \quad (***)$$

其中: **ecm**表示**误差修正项**。由**分布滞后模型**

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \mu Y_{t-1} + \varepsilon_t$$

知, 一般情况下  $|\mu| < 1$ , 由关系式  $\lambda = 1 - \mu$  得  $0 < \lambda < 1$ 。可以据此分析 **ecm** 的修正作用:

(1) 若 (t-1) 时刻 **Y** 大于其长期均衡解  $\alpha_0 + \alpha_1 X$ , **ecm** 为正, 则  $(-\lambda ecm)$  为负, 使得  $\Delta Y_t$  减少;

(2) 若 (t-1) 时刻 **Y** 小于其长期均衡解  $\alpha_0 + \alpha_1 X$ , **ecm** 为负, 则  $(-\lambda ecm)$  为正, 使得  $\Delta Y_t$  增大。

(\*\*\*) 体现了长期非均衡误差对的控制。

**需要注意的是：**在实际分析中，变量常以对数的形式出现。

**其主要原因在于**变量对数的差分近似地等于该变量的变化率，而经济变量的变化率常常是稳定序列，因此适合于包含在经典回归方程中。

**于是：**(1) 长期均衡模型

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$$

中的 $\alpha_1$ 可视为Y关于X的**长期弹性 (long-run elasticity)**

(2) 短期非均衡模型

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \mu Y_{t-1} + \varepsilon_t$$

中的 $\beta_1$ 可视为Y关于X的**短期弹性 (short-run elasticity)**。

更复杂的误差修正模型可依照一阶误差修正模型类似地建立。

如具有季度数据的变量，可在短期非均衡模型

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \mu Y_{t-1} + \varepsilon_t$$

中引入更多的滞后项。

引入二阶滞后的模型为

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \mu_1 Y_{t-1} + \mu_2 Y_{t-2} + \varepsilon_t$$

经过适当的衡等变形，可得如下二阶误差修正模型

$$\Delta Y_t = -\mu_2 \Delta Y_{t-1} + \beta_1 \Delta X_t - \beta_3 \Delta X_{t-1} - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t \quad (*)$$

式中， $\lambda = 1 - \mu_1 - \mu_2$ ， $\alpha_0 = \beta_0 / \lambda$ ， $\alpha_1 = (\beta_1 + \beta_2 + \beta_3) / \lambda$

引入三阶滞后项的误差修正模型与 (\*) 式相仿，只不过模型中多出差分滞后项  $\Delta Y_{t-2}$ ， $\Delta X_{t-2}$ 。

多变量的误差修正模型也可类似地建立。

如三个变量如果存在如下长期均衡关系

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Z_t$$

则其一阶非均衡关系可写成

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \gamma_1 Z_t + \gamma_2 Z_{t-2} + \mu Y_{t-1} + \varepsilon_t$$

于是它的一个误差修正模型为

$$\Delta Y_t = \beta_1 \Delta X_t + \gamma_1 \Delta Z_t - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1} - \alpha_2 Z_{t-1}) + \varepsilon_t$$

式中,  $\lambda = 1 - \mu$ ,  $\alpha_0 = \beta_0 / \lambda$ ,  $\alpha_1 = (\beta_1 + \beta_2) / \lambda$ ,  $\alpha_2 = (\gamma_1 + \gamma_2) / \lambda$

## 2、误差修正模型的建立

### (1) Granger 表述定理

误差修正模型有许多明显的**优点**：如

a) 一阶差分项的使用消除了变量可能存在的趋势因素，从而避免了虚假回归问题；

b) 一阶差分项的使用也消除模型可能存在的多重共线性问题；

c) 误差修正项的引入保证了变量水平值的信息没有被忽视；

d) 由于误差修正项本身的平稳性，使得该模型可以用经典的回归方法进行估计，尤其是模型中差分项可以使用通常的t检验与F检验来进行选取；等等。

因此，**一个重要问题就是：是否变量间的关系都可以通过误差修正模型来表述？**

就此问题，Engle 与 Granger 1987年提出了著名的Grange表述定理（Granger representaion theorem）：

如果变量X与Y是协整的，则它们间的短期非均衡关系总能由一个误差修正模型表述：

$$\Delta Y_t = lagged(\Delta Y, \Delta X) - \lambda \mu_{t-1} + \varepsilon_t \quad 0 < \lambda < 1 \quad (*)$$

式中， $\mu_{t-1}$ 是非均衡误差项或者说成是长期均衡偏差项， $\lambda$ 是短期调整参数。

对于(1,1)阶自回归分布滞后模型

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \mu Y_{t-1} + \varepsilon_t$$

如果  $Y_t \sim I(1)$ ,  $X_t \sim I(1)$  ; 那么

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t$$

的左边 $\Delta Y_t \sim I(0)$ ，右边的 $\Delta X_t \sim I(0)$ ，因此，只有Y与X协整，才能保证右边也是 $I(0)$ 。

因此，**建立误差修正模型**，需要

**首先**对变量进行协整分析，以发现变量之间的协整关系，即长期均衡关系，并以这种关系构成误差修正项。

**然后**建立短期模型，将误差修正项看作一个解释变量，连同其它反映短期波动的解释变量一起，建立短期模型，即误差修正模型。

**注意**，由于

$$\Delta Y = \text{lagged}(\Delta Y, \Delta X) + \lambda \mu_{t-1} + \varepsilon_t \quad 0 < \lambda < 1$$

中没有明确指出**Y**与**X**的滞后项数，因此，可以是多个；同时，由于一阶差分项是I(0)变量，因此模型中也允许使用**X**的非滞后差分项 $\Delta X_t$ 。

**Granger表述定理**可类似地推广到多个变量的情形中去。



## (2) Engle-Granger两步法

由协整与误差修正模型的关系，可以得到误差修正模型建立的E-G两步法：

**第一步**，进行协整回归（OLS法），检验变量间的协整关系，估计协整向量（长期均衡关系参数）；

**第二步**，若协整性存在，则以第一步求到的残差作为非均衡误差项加入到误差修正模型中，并用OLS法估计相应参数。

**需要注意的是：**在进行变量间的协整检验时，如有必要可在协整回归式中加入趋势项，这时，对残差项的稳定性检验就无须再设趋势项。

另外，第二步中变量差分滞后项的多少，可以残差项序列是否存在自相关性来判断，如果存在自相关，则应加入变量差分的滞后项。

### (3) 直接估计法

也可以采用打开误差修整模型中非均衡误差项括号的方法直接用OLS法估计模型。

但仍需事先对变量间的协整关系进行检验。

如对双变量误差修正模型

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda(Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t$$

可打开非均衡误差项的括号直接估计下式：

$$\Delta Y_t = \lambda \alpha_0 + \beta_1 \Delta X_t - \lambda Y_{t-1} + \lambda \alpha_1 X_{t-1} + \varepsilon_t$$

这时短期弹性与长期弹性可一并获得。

**需注意的是**，用不同方法建立的误差修正模型结果也往往不一样。

## 例9.3.2 中国居民消费的误差修正模型

经济理论指出，居民消费支出是其实际收入的函数。

以中国国民核算中的居民消费支出经过居民消费价格指数缩减得到中国居民实际消费支出时间序列（C）；

以支出法**GDP**对居民消费价格指数缩减近似地代表国民收入时间序列(**GDP**)

时间段为1978~2000（表9.3.3）

**表 9.3.3 1978~1998 年间中国实际居民消费与实际 GDP 数据（单位：亿元，1990 年价）**

年份	C	GDP	年份	C	GDP	年份	C	GDP
1978	3810	7809	1985	7579	14521	1992	11325	23509
1979	4262	8658	1986	8025	15714	1993	12428	27340
1980	4581	8998	1987	8616	17031	1994	13288	29815
1981	5023	9454	1988	9286	17889	1995	14693	31907
1982	5423	10380	1989	8788	16976	1996	16189	34406
1983	5900	11265	1990	9113	18320	1997	17072	36684
1984	6633	12933	1991	9977	20581	1998	18230	39008

## (1) 对数据lnC与lnGDP进行单整检验

容易验证lnC与lnGDP是一阶单整的，它们适合的检验模型如下：

$$\Delta^2 \ln C_t = 0.056 - 0.744 \Delta \ln C_{t-1}$$

$$(2.76) \quad (-3.23)$$

$$\text{LM}(1)=0.929 \quad \text{LM}(2)=1.121$$

$$\Delta^2 \ln GDP_t = 0.13 - 1.54 \Delta \ln GDP_{t-1} + 0.81 \Delta^2 \ln GDP_{t-1} + 0.59 \Delta^2 \ln GDP_{t-2} + 0.58 \Delta^2 \ln GDP_{t-3}$$

$$(3.81) \quad (-4.01) \quad (2.66) \quad (2.26) \quad (2.54)$$

$$\text{LM}(1)=0.38 \quad \text{LM}(2)=0.67 \quad \text{LM}(3)=2.34 \quad \text{LM}(4)=2.46$$

## (2) 检验lnC与lnGDP的协整性，并建立长期均衡关系

首先，建立lnC与lnGDP的回归模型

$$\ln C_t = 0.047 + 0.923 \ln GDP_t$$

$$(0.30) \quad (57.48)$$

$$R^2=0.994$$

$$DW=0.744$$

发现有残关项有较强的一阶自相关性。考虑加入适当的滞后项，得lnC与lnGDP的分布滞后模型

$$\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1} \quad (*)$$

$$(1.63) \quad (6.62)$$

$$(4.92)$$

$$(-2.17)$$

$$R^2=0.994$$

$$DW=1.92$$

$$LM(1)=0.00$$

$$LM(2)=2.31$$

自相关性消除，因此可初步认为是lnC与lnGDP的长期稳定关系。

## 残差项的稳定性检验：

$$\Delta \hat{e}_t = -0.9975 \hat{e}_{t-1}$$

$$(-4.32)$$

$$R^2=0.994 \quad DW=2.01 \quad LM(1)=0.04 \quad LM(2)=1.34$$

$$t=-4.32 < -3.64 = ADF_{0.05}$$

说明lnC与lnGDP是（1， 1）阶协整的，（\*）式即为它们长期稳定的均衡关系：

$$\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1} \quad (*)$$

### (3) 建立误差修正模型

- 以稳定的时间序列  $\hat{e}_t$  做为误差修正项，可建立如下  
误差修正模型：

$$\Delta \ln C_t = 0.686 \Delta \ln GDP_t + 0.784 \Delta \ln C_{t-1} - 0.484 \Delta \ln GDP_{t-1} - 1.163 \hat{e}_{t-1} \quad (**)$$

(6.96)                      (2.96)                      (-1.91)                      (-3.15)

$$R^2=0.994 \quad DW=2.06 \quad LM(1)=0.70 \quad LM(2)=2.04$$

由(\*)式  $\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1}$   
可得lnC关于lnGDP的长期弹性：

$$(0.698-0.361)/(1-0.622)=0.892;$$

由(\*\*)式可得lnC关于lnGDP的短期弹性： 0.686

下面用打开误差修正项括号的方法直接估计误差修正模型，适当估计式为：

$$\Delta \ln C_t = 0.153 + 0.698 \Delta \ln GDP_t - 0.378 \ln C_{t-1} + 0.337 \ln GDP_{t-1}$$

(1.63) (6.62) (-2.99) (2.88)

$$R^2=0.791 \quad =0.0064 \quad DW=1.93 \quad LM(2)=2.31 \quad LM(3)=2.78$$

写成误差修正模型的形式如下

$$\Delta \ln C_t = 0.698 \Delta \ln GDP_t - 0.378 (\ln C_{t-1} - 0.405 - 0.892 \ln GDP_{t-1}) \quad (***)$$

由 (\*\*\*) 式知， $\ln C$  关于  $\ln GDP$  的短期弹性为 0.698，长期弹性为 0.892。

可见两种方法的结果非常接近。



## (4) 预测

由(\*)式

$$\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1}$$

给出1998年关于长期均衡点的偏差:

$$\begin{aligned}\hat{e}_{98} &= \ln(18230) - 0.152 - 0.698 \ln(39008) - 0.662 \ln(17072) \\ &\quad + 0.361 \ln(36684) = 0.0125\end{aligned}$$

由(\*\*)式

$$\Delta \ln C_t = 0.686 \Delta \ln GDP_t + 0.784 \Delta \ln C_{t-1} - 0.484 \Delta \ln GDP_{t-1} - 1.163 \hat{e}_{t-1}$$

预测1999年的短期波动:

$$\begin{aligned}\Delta \ln C_{99} &= 0.686(\ln(41400) - \ln(39008)) + 0.784(\ln(18230) - \ln(17072)) \\ &\quad - 0.484(\ln(39008) - \ln(36684)) - 1.163 \times 0.0125 = 0.048\end{aligned}$$

于是  $\ln C_{99} = 0.048 + \ln C_{98} = 0.048 + \ln(18230) = 9.859$

$$C_{99} = e^{9.859} = 19125$$

按照 (\*\*\*) 式

$$\Delta \ln C_t = 0.698 \Delta \ln GDP_t - 0.378(\ln C_{t-1} - 0.405 - 0.892 \ln GDP_{t-1})$$

预测的结果为:

$$\begin{aligned} \Delta \ln C_{99} &= 0.698 (\ln(41400) - \ln(39008)) - 0.378(\ln(18230) - 0.405 \\ &\quad - 0.892 \ln(39008)) = 0.051 \end{aligned}$$

于是

$$\ln C_{99} = 0.051 + \ln C_{98} = 0.051 + \ln(18230) = 9.861$$

$$C_{99} = e^{9.861} = 19176$$

以当年价计的1999年实际居民消费支出为39334亿元，用居民消费价格指数（1990=100）紧缩后约为19697亿元，因此：**两个预测结果的相对误差分别为2.9%与2.6%。**

## 五、案例分析

# 1.单位根检验

- 实验基本原理

为了判断时间序列是否平稳，我们通常需要进行单位根检验。↵

对于  $ARMA(p,q)$  模型，其  $MA(q)$  部分显然是平稳的，因为它是有限个白噪声的线性组合。因此， $ARMA(p,q)$  的平稳性取决于其  $AR(p)$  的部分。对于  $p$  阶自回归模型：↵

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

这实质是一个随机差分方程。设  $\lambda_1, \lambda_2, \dots, \lambda_p$  是差分方程  $y_t - \beta_0 - \beta_1 y_{t-1} - \dots - \beta_p y_{t-p} = 0$  的  $p$  个特征根，可以证明，当所有特征根的绝对值都小于 1 时，对应的  $AR(p)$  模型是平稳的。即所有的解都落在复平面的单位圆之内时，模型平稳。如果某个根正好落在单位圆之上，则称为“单位根”（unit root）。对于  $AR(1)$  模型而言，平稳要求  $|\beta_1| < 1$ 。↵

时间序列存在单位根会导致  $t$  检验会失效、大样本理论不适用、“伪回归”（spurious regression）等问题。对于存在单位根的序列，我们可以对其进行差分，然后再拟合  $ARMA$  模型，即使用  $ARIMA$  模型或  $ARIMAX$  模型。对于多个单位根序列，“协整”（Cointegration）也是有效的处理办法。但无论采取哪种方法，必须首先对是否存在单位根进行检验。↵

对单位根进行检验的方法主要有 Dickey-Fuller 检验（DF 检验）、Augmented Dickey-Fuller 检验（ADF 检验）、Phillips-Perron 单位根检验（PP 检验）、DF-GLS 单位根检验等。↵

### (1) Dickey-Fuller 检验

DF 检验使用一阶自回归来检验单位根。即对于序列 $\{y_t\}$ ，检验模型 $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$ 中 $\beta_1$ 是否为 1。其中，扰动项 $\{\varepsilon_t\}$ 为独立白噪声。在方程两边同时减去 $y_{t-1}$ 可得

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

其中， $\delta \equiv \beta_1 - 1$ 。因为我们认为 $\beta_1$ 不会大于 1，这样，检验的原假设与备择假设为：

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

对上述方程做 OLS 回归，可得估计量 $\hat{\delta}$ 及对应的 t 统计量，在 Stata 中记为 Z(t)。DF 检验是左边单侧检验，Z(t)越小，越倾向于拒绝原假设。

### (2) Augmented Dickey-Fuller 检验

DF 检验要求扰动项 $\{\varepsilon_t\}$ 为独立白噪声，即要求扰动项无自相关。如果 $\{\varepsilon_t\}$ 存在自相关，可以引入更高阶的滞后项来控制。假设选择了适当的滞后期 p，使得以下 AR(p)模型的扰动项 $\{\varepsilon_t\}$ 为独立白噪声，

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

为了方便检验，将上式转换为以下形式，

$$\Delta y_t = \beta_0 + \delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \dots + \gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

可以证明，如果 $\delta = 0$ ，则 AR(p)有一个单位根。如果 $\delta < 0$ ，该 AR(p)模型平稳。这样，检验的原假设与备择假设为：

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

OLS 回归可得估计量 $\hat{\delta}$ 及对应的 t 统计量，该统计量在 Stata 中记为 Z(t)。Z(t)越小，则越倾向于拒绝原假设。

### (3) Phillips-Perron 单位根检验↵

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。PP 检验仍然使用一阶自回归，但使用“异方差自相关稳健的标准差”（Newey-West standard error）对 DF 统计量进行修正：↵

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

其中， $\{\varepsilon_t\}$ 可以存在异方差或自相关。PP 检验也是左边单侧检验。由于金融变量常存在异方差与自相关，故 PP 检验在分析金融数据中应用较广。PP 检验的另一个优点是，不必指定滞后期数。↵

### (4) DF-GLS 单位根检验↵

ADF 检验与 PP 检验的共同缺点是，检验的“功效”（power）较低，犯第 II 类错误的概率很大，即如果真实模型接近于单位根但并非单位根时（自回归系数接近于 1），ADF 检验与 PP 检验很可能不能拒绝模型存在单位根。在这种情况下，我们应考虑 DF-GLS 检验。↵

DF-GLS 检验分两步进行。第一步，用 GLS 估计原序列 $\{y_t\}$ 的时间趋势，并计算“去趋势后”（detrended）的序列。第二步，对去趋势后的序列使用 ADF 检验。↵

- 实验内容及数据来源

- 我们要进行时间序列分析，应首先检验序列是否平稳。我们以实验12-5中对国际航线乘客数量的分析为例，即利用本书附带光盘data文件夹下的“airpsn.dta”工作文件。
- 利用这些数据，我们来讲解各种单位根检验的操作及相关选项的含义。

# • 实验操作指导

## • 1 Augmented Dickey-Fuller单位根检验的操作

ADF 单位根检验的基本操作命令为：↵

`dfuller varname [if] [in] [, options]`↵

其中，`dfuller` 代表“Augmented Dickey-Fuller 单位根检验”的基本命令语句，`varname` 代表要检验的变量的名称，`if` 代表条件语句，`in` 代表范围语句，`options` 代表其他选项。表 12.18 显示了各 `options` 选项的具体内容。↵

表 12.18 ADF 单位根检验的 `options` 选项↵

<code>noconstant</code> ↵	设定检验方程中没有截距项或时间趋势项↵	↵
<code>trend</code> ↵	在检验方程中包括时间趋势项（可以同时有漂移项）↵	↵
<code>drift</code> ↵	在检验方程中包括漂移项，但不包括时间趋势项↵	↵
<code>regress</code> ↵	显示检验方程回归结果↵	↵
<code>lags(#)</code> ↵	设定滞后差分的阶数，默认的滞后差分阶数为 0，即为 Dickey-Fuller 单位根检验↵	↵

其中，选项 `noconstant`、`trend`、`drift` 最多只能设定一个。默认情况下，`stata` 的检验方程中不包括时间趋势项，但包含截距项。与选项 `drift` 不同的是，默认情况下假设截距项的总体值为 0，而选项 `drift` 假定截距项的总体值不为 0。↵



- 利用“airpsn.dta”的数据，我们来检验乘客数量的对数是否存在单位根。输入命令：
- `gen lnair=ln(air)`
- `dfuller lnair, lags(3) trend regress`
- 其中，第一步为生成变量lnair，其值为air的对数；第二步对lnair进行单位根检验，设定滞后期为3，在检验方程中包括时间趋势项，并显示回归结果。

## 2 Phillips-Perron单位根检验的操作

Phillips-Perron 单位根检验的基本命令为：

`pperron varname [if] [in] [, options]`

其中，`pperron` 代表“Phillips-Perron 单位根检验”的基本命令语句，`varname` 代表要检验的变量的名称，`if` 代表条件语句，`in` 代表范围语句，`options` 代表其他选项。表 12.19 显示了各 `options` 选项的具体内容。

表 12.19 Phillips-Perron 单位根检验的 `options` 选项

<code>noconstant</code>	设定检验方程中没有截距项
<code>trend</code>	在检验方程中包括时间趋势项（可以同时有漂移项）
<code>regress</code>	显示检验方程回归结果
<code>lags(#)</code>	设定 Newey-West 标准差的计算中的滞后阶数，默认为 $\text{int}\{4(T/100)^{2/9}\}$ 。 其中， <code>int</code> 表示取数值的整数部分。

其中，选项 `noconstant`、`trend` 不能同时设定。默认情况下，`stata` 的检验方程中不包括时间趋势项，但包含截距项。与 ADF 检验不同的是，这里没有 `drift` 选项，这是因为选项 `drift` 的设定实质为选项 `trend` 的特例，所以，这不会对我们的模型设定有什么影响。

对变量 `lnair` 进行 Phillips-Perron 单位根检验，命令为：

`pperron lnair, trend regress`

- 3 DF-GLS单位根检验的操作

DF-GLS 单位根检验的基本命令为: ↵

`dfglst varname [if] [in] [, options]`↵

其中, `dfglst` 代表“DF-GLS 单位根检验”的基本命令语句, `varname` 代表要检验的变量的名称, `if` 代表条件语句, `in` 代表范围语句, `options` 代表其他选项。表 12.20 显示了各 `options` 选项的具体内容。↵

表 12.20 DF-GLS 单位根检验的 `options` 选项↵

<code>maxlag(#)</code> ↵	设定检验方程的最高滞后阶数, 默认为 $\text{int}\{12[(T+1)/100]^{2/9}\}$ ↵↵
<code>notrend</code> ↵	序列平稳无趋势↵↵
<code>ers</code> ↵	利用插值法计算临界值↵↵

- 对变量 `lnair` 进行 DF-GLS 单位根检验, 命令为:
- `dfglst lnair, maxlag(6)`
- 这里, 我们设定最大滞后期为 6。

- 4 时间序列算子
- 在进行时间序列分析时，我们经常要用到某变量的滞后值或差分值等，这可以通过时间序列算子实现。常用的算子及其含义列于表12.4中。

表 12.4 时间序列算子

算子	含义
L.	一阶滞后 $x_{t-1}$
L2.	二阶滞后 $x_{t-2}$
.....	.....
F.	一阶领先 $x_{t+1}$
F2.	二阶领先 $x_{t+2}$
.....	.....
D.	一阶差分 $x_t - x_{t-1}$
D2.	二阶差分 $(x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2}$
.....	.....
S.	季节差分 $x_t - x_{t-1}$
S2.	二阶季节差分 $x_t - x_{t-2}$
.....	.....

## 2. 我国城镇居民的生活费支出与可支配收入关系的研究

下表是我国城镇居民月人均可支配收入（SR）和生活费支出（ZC）的调整序列。现用EG两步法考察它们之间是否存在协整关系

序列	月份	1992	1993	1994	1995	1996	1997	1998
	1	151.83	265.93	273.98	370.00	438.37	521.01	643.40
	2	159.86	196.96	318.81	385.21	561.29	721.01	778.62
	3	124.00	200.19	236.45	308.62	396.82	482.38	537.16
	4	124.88	199.48	248.00	320.33	405.27	492.96	545.79
	5	127.75	200.75	261.16	327.94	410.06	499.90	567.99
	6	134.48	208.50	273.45	338.53	415.38	508.81	555.79
Str	7	145.05	218.82	278.10	361.09	434.70	516.24	570.23
	8	138.31	209.07	277.45	356.30	418.21	509.98	564.38
	9	144.25	223.17	292.71	371.32	442.30	538.46	576.96
	10	143.86	226.51	289.36	378.72	440.81	537.09	599.40
	11	149.12	226.62	296.50	383.58	449.03	534.12	577.40
	12	139.93	210.32	277.60	427.78	449.17	511.22	606.14

...	1.1	139.47	221.74	234.28	307.10	373.58	419.39	585.70
...	2.1	168.07	186.49	272.09	353.55	471.77	528.09	598.82
...	3.1	110.47	185.92	202.88	263.37	350.36	390.04	417.27
...	4.1	113.22	185.26	227.89	281.22	352.15	405.63	455.60
...	5.1	115.82	187.62	235.70	299.73	369.57	426.81	466.20
Zc.1	6.1	118.20	12.11	237.89	308.18	370.41	422.00	455.19
...	7.1	118.03	186.75	239.71	315.87	376.90	428.70	458.57
...	8.1	124.45	187.07	252.52	331.88	387.44	459.29	475.40
...	9.1	147.70	219.23	286.75	385.99	454.93	517.06	591.41
...	10.1	135.14	212.80	270.00	355.92	403.77	463.98	494.57
...	11.1	135.20	205.22	274.37	355.11	410.10	422.96	496.69
...	12.1	128.03	192.64	250.01	386.08	400.48	460.92	516.16