第十章 时间序列计量模型

引言

- 传统的计量经济分析以经济理论为基础，通过建立理论和计量模型，利用样本数据估计模型，然后据此进行结构分析和预测。
- 但是，由于社会经济现象往往受许多因素的影响，并且这些因素又保持着错综复杂的联系，因此，运用结构式的模型进行分析和预测往往比较困难
- 在这种情况下，“让数据自己说话”的建模思想越来越引起人们的重视，时间序列分析就属于这种学派的一个分支。

引言

时间序列分析方法由Box-Jenkins (1976) 提出

- 时间序列分析是一个分析动态数据、揭示数据规律的计量经济学分支
- 依据经济变量自身的历史资料，采用一定的统计方法，建立起能反映变量自身规律性的动态模型，以此对经济变量进行分析和预测
- 将时间序列分析方法与经典计量经济学分析方法相结合研究经济、金融问题，已成为计量经济学新的发展方向

引子：是真回归还是伪回归？

- 经典回归分析的做法是：

首先采用普通最小二乘法（OLS）对回归模型进行估计，然后根据可决系数或F检验统计量值的大小来判定变量之间的相依程度，根据回归系数估计值的t统计量对系数的显著性进行判断，最后在回归系数显著不为零的基础上对回归系数估计值给予经济解释。

- 为了分析美国的个人可支配总收入(I)与个人消费总支出(E)的关系，用OLS法作E关于I的线性回归，得到如下结果：

$$E_t = -174.44 \quad + \quad 0.9672\, I_t$$

$$t = (-7.481) \qquad (119.87)$$

$$R^2 = 0.9941 \qquad \mathrm{DW} = 0.532$$

- 从回归结果来看，可决系数非常高，个人可支配总收入I的回归系数t统计量也非常大，边际消费倾向符合经济假设。凭借经验判断，这个模型的设定是好的，应是非常满意的结果。准备将这个计量结果用于经济结构分析和经济预测。
- 可是有人提出，这个回归结果可能是虚假的！可能只不过是一种“伪回归”！

这里用时间序列数据进行的回归，究竟是真回　　归还是伪回归呢？为什么模型、样本、数据、检验结果都很理想，却可能得到“伪回归”的结果呢？

- 时间序列数据被广泛地运用于计量经济研究。经典时间序列分析和回归分析有许多假定前提，如序列的平稳性、正态性等。直接将经济变量的时间序列数据用于建模分析，实际上隐含了上述假定，在这些假定成立的条件下，据此而进行的t检验、F检验等才具有较高的可靠度。
- 越来越多的经验证据表明，经济分析中所涉及的大多数时间序列是非平稳的。

问题：

- 如果直接将非平稳时间序列当作平稳时间序列来进行分析，会造成什么不良后果；
- 如何判断一个时间序列是否为平稳序列；
- 当我们在计量经济分析中涉及到非平稳时间序列时，应作如何处理？

本章主要讨论

- 时间序列的基本概念
- 时间序列平稳性的单位根检验
- 协整

本节基本内容

- 伪回归问题
- 随机过程的概念
- 时间序列的平稳性

随机过程

- 有些随机现象，要认识它必须研究其发展变化过程，随机现象的动态变化过程就是随机过程。
- 例如，考察一段时间内每一天的电话呼叫次数，需要考察依赖于时间t的随机变量ξ_t，$\{\xi_t\}$就是一个随机过程。
- 又例如，某国某年的GDP总量，是一随机变量，但若考查它随时间变化的情形，则$\{\boldsymbol{GDP_t}\}$就是一个随机过程。

随机过程的简单理解

- 若对于每一特定的$\boldsymbol{t}(\boldsymbol{t} \in \boldsymbol{T})$，$\boldsymbol{Y}_{\mathrm{t}}$为一随机变量，则称这一族随机变量$\{\boldsymbol{Y}_{\mathrm{t}}\}$为一个随机过程。
- 若$\boldsymbol{Y}_{\mathrm{t}}$为一区间，则$\{\boldsymbol{Y}_{\mathrm{t}}\}$为一连续型随机过程。
- 若$\boldsymbol{Y}_{\mathrm{t}}$为离散集合，如$\boldsymbol{T} = (0，1，2，\dots)$或

$\boldsymbol{T} = (\dots，-2，-1，0，1，2，\dots)$，则$\{\boldsymbol{Y}_{\mathrm{t}}\}$为离散型随机过程。

- 离散型时间指标集的随机过程通常称为随机型时间序列，简称为时间序列。

随机过程的严格定义

- 由随机变量组成的一个有序序列称为随机过程，记为{x（s,t），s∈S，t∈T }。其中S表示样本空间，T表示序数集。
- 对于每一个 t，t∈T，x（·，t）是样本空间S中的一个随机变量。对于每一个 s，s∈S，x（s，·）是随机过程在序数集T中的一次实现。

样本空间

$$\{x_1^{\,1},\ x_2^{\,1},\ \ldots,\ x_{T\text{-}1}^{\,1},\ x_T^{\,1}\}$$

一次实现

$$\boxed{\{x_1^{\,2},\ x_2^{\,2},\ \ldots,\ x_{T\text{-}1}^{\,2},\ x_T^{\,2}\}}$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$\{x_1^{\,s},\ x_2^{\,s},\ \ldots,\ x_{T\text{-}1}^{\,s},\ x_T^{\,s}\}$$

随机过程的一次实现称为时间序列，也用{x_t}或x_t表示。

时间序列的平稳性

- 所谓时间序列的平稳性，是指时间序列的统计规律不会随着时间的推移而发生变化。
- 直观上，一个平稳的时间序列可以看作一条围绕其均值上下波动的曲线。
- 从理论上，有两种意义的平稳性，一是严格平稳，另一种是弱平稳。

严格平稳

- 是指随机过程$\{Y_t\}$的联合分布函数与时间的位移无关。设$\{Y_t\}$为一随机过程，$\boldsymbol{n}, \boldsymbol{h}$为任意实数，若联合分布函数满足：

$$F_{Y_{t_1},Y_{t_2},\dots,Y_{t_n}}(y_1,\dots,y_n) = F_{Y_{t_1+h},\dots,Y_{t_n+h}}(y_1,\dots,y_n)$$

则称$\{Y_t\}$为严格平稳过程，它的分布结构不随时间推移而变化。

弱平稳

是指随机过程$\{Y_t\}$的期望、方差和协方差不随时间推移而变化。若$\{Y_t\}$满足：

$$E(Y_t) = \mu$$

$$Cov(Y_t, Y_s) = Cov(Y_{t+h}, Y_{s+h}) = r(t-s, 0) = r_{t-s}$$

$$Var(Y_t) = r_0 = \sigma^2$$

则称$\{Y_t\}$为弱平稳随机过程。在一般的分析讨论中，平稳性通常是指弱平稳。

时间序列的非平稳性

- 是指时间序列的统计规律随着时间的位移而发生变化，即生成变量时间序列数据的随机过程的特征随时间而变化。
- 在实际中遇到的时间序列数据很可能是非平稳序列，而平稳性在计量经济建模中又具有重要地位，因此有必要对观测值的时间序列数据进行平稳性检验。

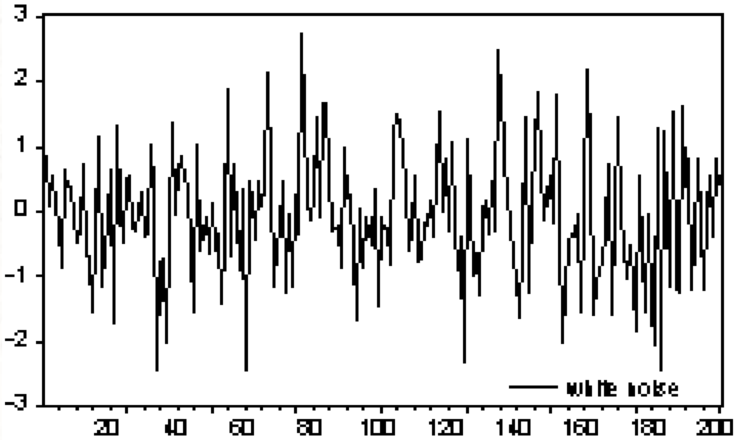
平稳与非平稳的直观图示

- 一种特殊的平稳过程—白噪声（white noise）

$$E(x_t) = 0$$

$$\mathrm{var}(x_t) = \sigma^2$$

$$\mathrm{cov}(x_t, x_{t+k}) = \gamma_k = 0$$



由白噪声过程产生的时间序列

白噪声源于物理学与电学，原指音频和电信号在一定频带中的一种强度不变的干扰声。

平稳与非平稳的直观图示

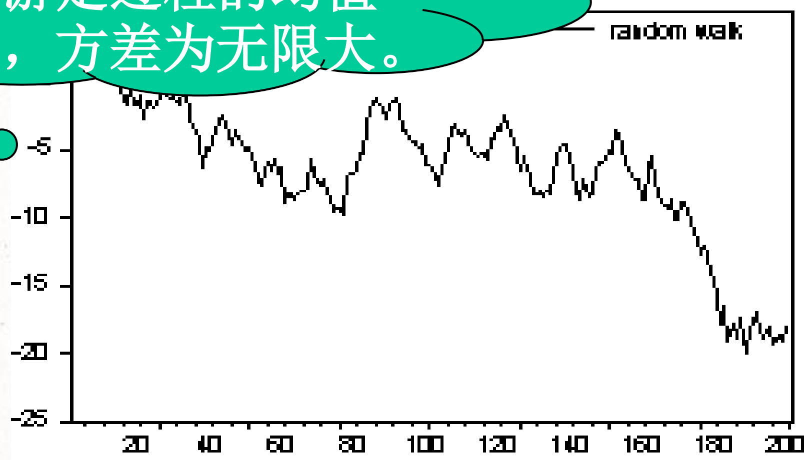
- 一种特殊的非平稳过程——随机游走（random walk）

对于表达式：

$$x_t = x_{t-1} + u_t$$

如果u_t 为白噪声过程，则称x_t 为随机游走过程。

随机游走过程的均值为零，方差为无限大。



由随机游走过程产生时间序列

“随机游走”一词首次出现于1905年自然（Nature）杂志第72卷Pearson K. 和Rayleigh L.的一篇通信中。文中讨论寻找一个被放在野地中央的醉汉的最佳策略是从投放点开始搜索。

伪回归问题

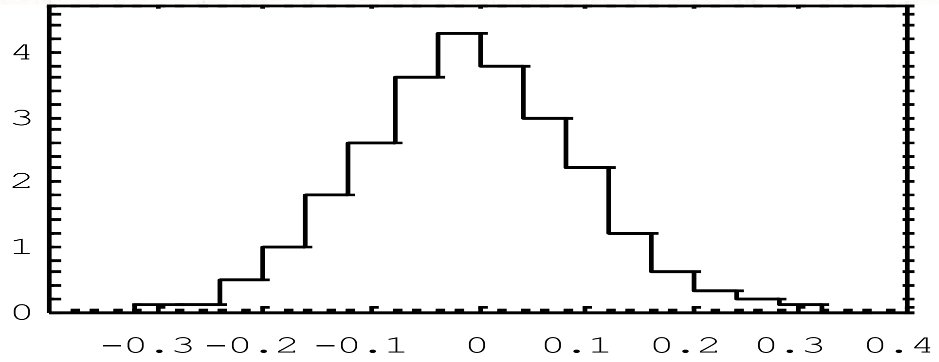
- 传统计量经济学模型的假定条件：序列的平稳性、正态性。
- 所谓“伪回归”，是指变量间本来不存在相依关系，但回归结果却得出存在相依关系的错误结论。
- 20世纪70年代，Grange、Newbold 研究发现，造成“伪回归”的根本原因在于时序序列变量的非平稳性

伪回归的模拟示例

- **用蒙特卡罗模拟方法分析相关系数的分布：**

$$u_t \sim \mathrm{IN}(0, 1),\ \ u_t \sim \mathrm{I}\,(0)$$
$$v_t \sim \mathrm{IN}(0, 1),\ \ v_t \sim \mathrm{I}\,(0)$$

每次生成T=100的相互独立的$\{u_t\}$和$\{v_t\}$，并计算R_{uv}。重复1万次，从而得到R_{uv}的分布为正态分布。

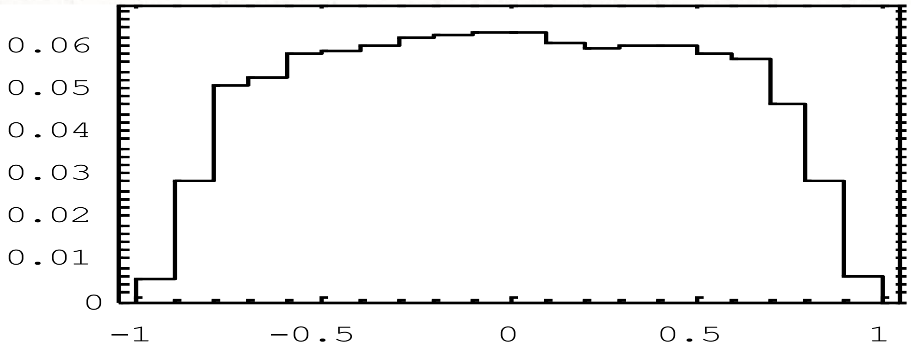


伪回归的模拟示例

$$x_t = x_{t-1} + u_t, \quad x_0 = 0, \quad x_t \sim \mathrm{I}(1)$$

$$y_t = y_{t-1} + v_t, \quad y_0 = 0, \quad y_t \sim \mathrm{I}(1)$$

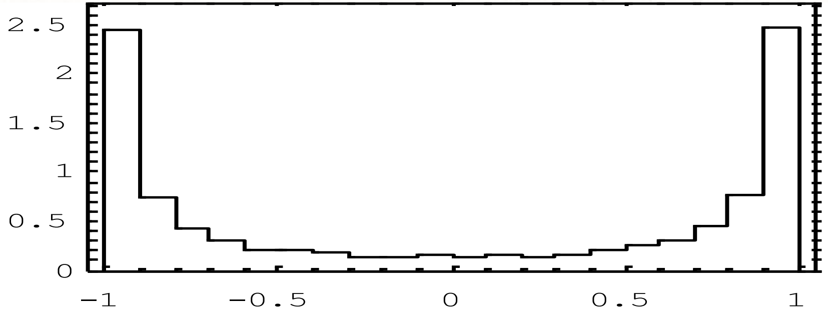
利用$\{u_t\}$和$\{v_t\}$，每次生成T=100的$\{x_t\}$和$\{y_t\}$并计算R_{xy}。重复1万次，从而得到R_{xy}的分布为倒U形 。



伪回归的模拟示例

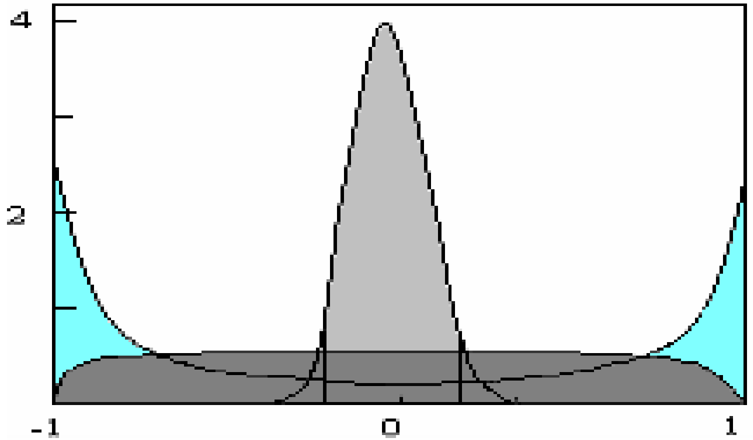
$$p_t = p_{t-1} + x_t, \quad p_0 = 0, \quad p_t \sim \mathrm{I}\,(2)$$
$$q_t = q_{t-1} + y_t, \quad q_0 = 0, \quad q_t \sim \mathrm{I}\,(2)$$

利用 $\{x_t\}$ 和 $\{y_t\}$，每次生成 T=100的 $\{p_t\}$ 和 $\{q_t\}$ 并计算 R_{pq}。重复1万次，从而得到R_{pq}的分布为U形 。



伪回归的模拟示例

- 问题的严重性在于当变量非平稳时，认为R服从的是正态分布，但实际上R服从的却是倒U和U字型分布，因此增加了拒绝概率，本不相关的两个变量结论却是相关！



Granger (1974)的简单介绍

有如下数据生成系统

- $x_t = x_{t-1} + u_t, \quad x_0 = 0, \; u_t \sim \mathrm{IID}(0, 1)$
- $y_t = y_{t-1} + v_t, \quad y_0 = 0, \; v_t \sim \mathrm{IID}\,(0, 1)$

$$\mathrm{E}(u_i v_j) = 0, \qquad \forall i, j$$

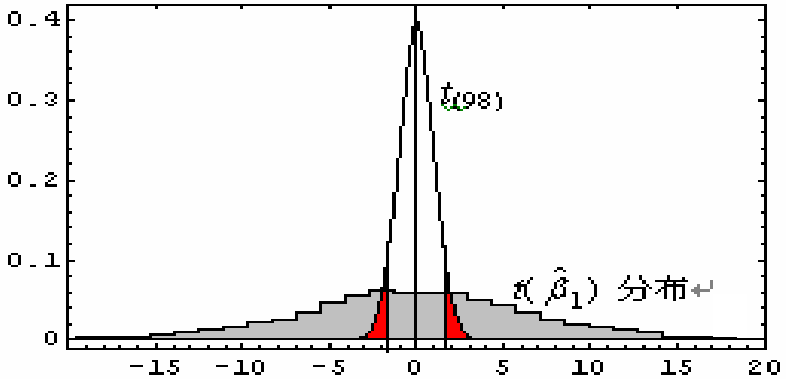
可知x_t和y_t为I(1)变量且相互独立。

作如下回归：

$$y_t = \beta_0 + \beta_1 x_t + w_t$$

Granger (1974)的简单介绍

$t(\hat{\beta}_1)$的分布见下图



拒绝$\beta_1 = 0$的概率大大增加。从而造成虚假回归

第二节 时间序列平稳性的单位根检验

- 单位根检验
- Dickey—Fuller检验
- Augmented Dickey—Fuller检验

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

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

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

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

当 $\boldsymbol{t} \to \infty$ 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：

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

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

$$Y_t \sim I(1)$$

- 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$\{Y_t\}\sim I(0)$。

二 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,ε_t 独立同分布，期望为零，方差为 σ^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: $\hat{\gamma} = \dfrac{\sum Y_{t-1} Y_t}{\sum Y_{t-1}^2}$

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

在 $H_0:\gamma=1$ 成立的条件下，t统计量为：

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

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 τ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：

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

(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：$\hat{\gamma} = \dfrac{\sum Y_{t-1} Y_t}{\sum Y_{t-1}^2}$

(2) 提出假设H_0：$\gamma = 1$　H_1：$\gamma < 1$

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

究竟应该采用哪个模型进行检验？

- 从最一般地模型（模型III）开始，依次对上述三个方程进行检验：
    - 只要有一个拒绝原假设，则可判断序列是平稳的。
    - 若三个模型都接受了原假设，则说明原序列是非平稳的。
- 还可通过画图：
    - 若包含常数项，则检验序列的序列图应该在偏离0的位置随机变动。
    - 若含有线性趋势项，则序列的曲线图应该有随时间变化而变化的波动趋势存在。

三、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： $Y_t = \gamma Y_{t-1} + u_t$

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

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

其中 u_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：$$Y_t = \alpha + \gamma Y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta Y_{t-i} + \varepsilon_t$$

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

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10.1

根据《中国统计年鉴2017》，得到我国1978—2016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—2016年度GDP序列

年度	GDP	年度	GDP	年度	GDP
1978	3678.7	1991	22005.6	2004	161840.2
1979	4100.5	1992	27194.5	2005	187318.9
1980	4587.6	1993	35673.2	2006	219438.5
1981	4935.8	1994	48637.5	2007	270232.3
1982	5373.4	1995	61339.9	2008	319515.5
1983	6020.9	1996	71813.6	2009	349081.4
1984	7278.5	1997	79715.0	2010	413030.3
1985	9098.9	1998	85195.5	2011	489300.6
1986	10376.2	1999	90564.4	2012	540367.4
1987	12174.6	2000	100280.1	2013	595244.4
1988	15180.4	2001	110863.1	2014	643974.0
1989	17179.7	2002	121717.4	2015	689052.1
1990	18872.9	2003	137422.0	2016	744127.2

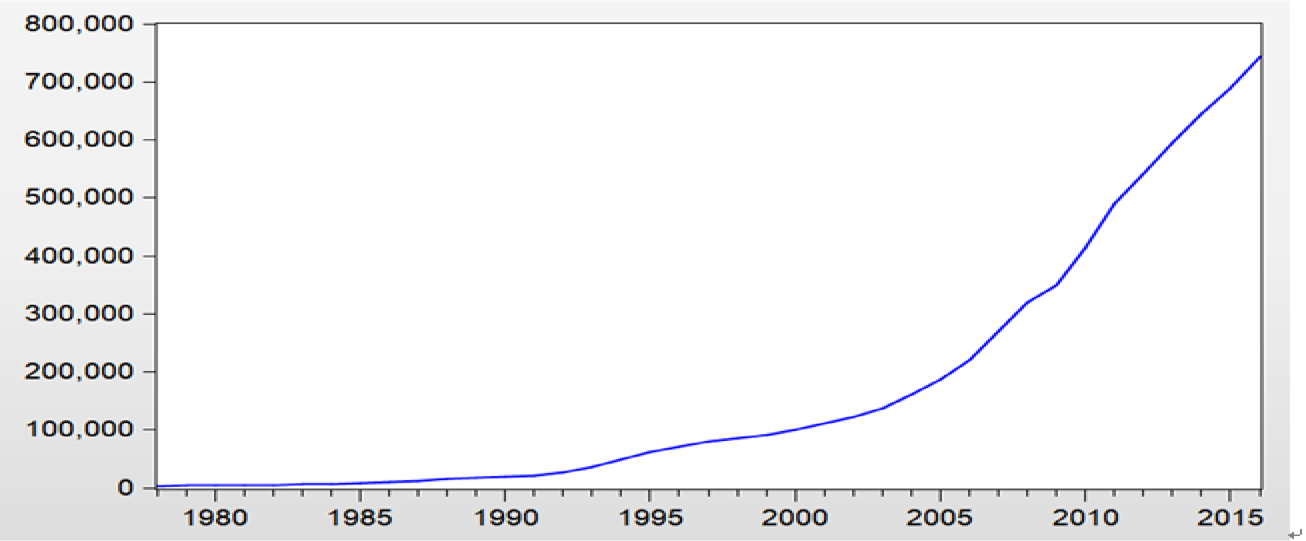


图 10.1　GDP 时间序列图

- 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：

$$\Delta \widehat{GDP}_t = -5636.761 + 638.2892\,t - 0.032007\,GDP_{t-1}$$

$$+0.6592\Delta GDP_{t-1} - 0.2328\Delta GDP_{t-2} + 0.6024\Delta GDP_{t-3}$$

在原假设下，单位根的t检验统计量的值为

$$t = \frac{\hat{\gamma} - \gamma}{\hat{\sigma}_{\hat{\gamma}}} = \frac{-0.032007}{0.026339} = -1.2152$$

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-4.243644、-3.544284、-3.204699，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——2016年度GDP序列存在单位根，是非平稳序列。

第三节 协整

- 协整的概念
- 协整检验
- 误差修正模型

协整的概念

引例：一个货币需求分析的例子。

依照经典理论，一国或一地区的货币需求量主要取决于规模变量和机会成本变量，即实际收入、价格水平以及利率。以对数形式的计量经济模型将货币需求函数描述出来，形式为：

$$\ln M_t = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln Y_t + \beta_3 r_t + u_t$$

其中，M 为货币需求，P 为价格水平，Y 为实际收入总额，r 为利率，u 为扰动项，β 为模型参数。

问题：估计出来的货币需求函数是否揭示了货币需求的长期均衡关系？

（1）如果上述货币需求函数是适当的，那么货币需求对长期均衡关系的偏离将是暂时的，扰动项序列是平稳序列，估计出来的货币需求函数就揭示了货币需求的长期均衡关系。

（2）相反，如果扰动项序列有随机趋势而呈现非平稳现象，那么模型中的误差会逐步积聚，使得货币需求对长期均衡关系的偏离在长时期内不会消失。

- 上述货币需求模型是否具有实际价值，关键在于扰动项序列是否平稳。
- 货币供给量、实际收入、价格水平以及利率可能是I(1)序列。一般情况下，多个非平稳序列的线性组合也是非平稳序列。
- 如果货币供给量、实际收入、价格水平以及利率的任何线性组合都是非平稳的，那么上述货币需求模型的扰动项序列就不可能是平稳的，从而模型并没有揭示出货币需求的长期稳定关系。

反过来说，如果上述货币需求模型描述了货币需求的长期均衡关系，那么扰动项序列必定是平稳序列，也就是说，非平稳的货币供给量、实际收入、价格水平以及利率四变量之间存在平稳的线性组合。

上述例子向我们揭示了这样一个事实：

“包含非平稳变量的均衡系统，必然意味着这些非平稳变量的某种组合是平稳的”

这正是协整理论的思想。

所谓协整，是指多个非平稳变量的某种线性组合是平稳的。

例如，收入与消费，工资与价格，政府支出与税收，出口与进口等，这些经济时间序列一般是非平稳序列，但它们之间却往往存在长期均衡关系。

下面给出协整的严格定义：

对于两个序列$\{\boldsymbol{X}\}$和$\{\boldsymbol{Y}\}$如果$\boldsymbol{y_t \sim I(1)}$，$\boldsymbol{x_t \sim I(1)}$，而且存在一组非零常数$\boldsymbol{\alpha_1}$、$\boldsymbol{\alpha_2}$，使得$\boldsymbol{\alpha_1 x_t + \alpha_2 y_t \sim I(0)}$则称$\{\boldsymbol{X}\}$和$\{\boldsymbol{Y}\}$之间是协整的。

一般的，设有$\boldsymbol{k}(\geq 2)$个序列$\{\boldsymbol{y}_{1t}\},\{\boldsymbol{y}_{2t}\},\ldots,\{\boldsymbol{y}_{kt}\}$，用$\boldsymbol{Y}_t=(\boldsymbol{y}_{1t},\boldsymbol{y}_{2t},\ldots,\boldsymbol{y}_{kt})'$表示由此$\boldsymbol{k}$个序列构成的$\boldsymbol{k}$维向量序列，

如果：

(1)每一个序列$\{\boldsymbol{y}_{1t}\},\{\boldsymbol{y}_{2t}\},\ldots,\{\boldsymbol{y}_{kt}\}$都是$\boldsymbol{d}$阶单整序列，即$\boldsymbol{y}_{it}\sim \boldsymbol{I}(\boldsymbol{d})$；

(2) 存在非零向量 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_k)'$ 使得

$Y_t=a_1y_{1t}+a_2y_{2t}+\cdots+a_ky_{kt}$ 为（$d-b$）阶单整序列，

即 $\alpha' Y_t \sim \mathrm{I}(d-b)$，$0<b\le d$。

则称向量序列 $Y_t=(y_{1t},y_{2t},\cdots,y_{kt})'$ 的分量间是 d、b 阶协整的，记为 $Y_t\sim \mathrm{CI}(d,b)$，

向量 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_k)'$ 称为协整向量。

特别地，若$d=b=1$，则 $Y_t \sim CI(1,1)$，说明尽管各个分量序列是非平稳的一阶单整序列，但它们的某种线性组合却是平稳的。这种（1，1）阶协整关系在经济计量分析中较为常见。例如，假设变量 y_{1t} 与变量 y_{it}（$i=2,\ldots,m$）之间为（1，1）阶协整关系，协整向量为 $\boldsymbol{\alpha}=(1,-\beta_2,\cdots,-\beta_m)'$，

则这种协整关系可表示为：

$$y_{1t}=\alpha+\beta_2 y_{2t}+\cdots+\beta_m y_{mt}+u_t$$

组合变量u_t 就为I(0)过程。

例如，居民收入时间序列 Y_t 为 1 阶单整序列，居民消费时间序列 C_t 也为 1 阶单整序列，如果二者的线性组合 $\alpha_1 Y_t + \alpha_2 C_t$ 构成的新序列为 0 阶单整序列，于是认为序列 Y_t 与 C_t 是(1,1)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相同时，才可能协整。例如上面的居民收入 Y_t 和居民消费 C_t；如果它们的单整阶不相同时，就不可能协整。例如居民消费 C_t 和居民储蓄余额 S_t（一般讲作为存量的居民储蓄余额 S_t 为 2 阶单整）。

从协整的定义可以看出协整的经济意义在于：两个变量，虽然它们具有各自的长期波动规律，但是如果它们是协整的，则它们之间存在着一个长期稳定的比例关系。[redacted]收入 Y_t 和居民消费 C_t，如果它们各自都是 1 阶单整，并且它们是(1.1)阶协整，则说明它们之间存在着一个长期稳定的比例关系，而这个比例关系就是消费倾向，也就是说消费倾向是不变的。

从计量经济学模型的意义上讲，建立如下消费函数模型

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

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。

反过来，如果两个变量，具有各自的长期波动规律，但是它们不是协整的，则它们之间就不存在着一个长期稳定的比例关系。例如居民消费 C_t 和居民储蓄余额 S_t，由于它们单整阶数不同，所以它们不是协整的，则说明它们之间不存在着一个长期稳定的比例关系。从计量经济学模型的意义上讲，建立如下消费函数模型

$$C_t = \alpha_0 + \alpha_1 S_t + \mu_t$$

或者

$$C_t = \alpha_0 + \alpha_1 Y_t - \alpha_2 S_t + \mu_t$$

变量选择是不合理的，随机误差项一定不是“白噪声”，模型参数没有合理的经济解释。

1. 在建立计量经济模型时，检验变量之间的协整关系是非常重要的；
2. 从协整理论出发，在建立消费函数模型时，就不会选择居民储蓄余额作为居民消费的解释变量。

协整概念的提出对于用非平稳变量建立经济计量模型，以检验这些变量之间的长期均衡关系非常重要。

（1）如果多个非平稳变量具有协整性，则这些变量可以合成一个平稳序列。这个平稳序列就可以用来描述原变量之间的均衡关系。

（2）当且仅当多个非平稳变量之间具有协整性时，由这些变量建立的回归模型才有意义。所以协整性检验也是区别真实回归与伪回归的有效方法。

（3）具有协整关系的非平稳变量可以用来建立误差修正模型。由于误差修正模型把长期关系和短期动态特征结合在一个模型中，因此既可以克服传统计量经济模型忽视伪回归的问题，又可以克服建立差分模型忽视水平变量信息的弱点。

协整检验

协整性的检验有两种方法

- 基于回归残差的协整检验，这种检验也称为单一方程的协整检验；
- 基于回归系数的完全信息协整检验。

这里我们仅介绍单一方程的EG两步法协整检验。

EG两步检验法

前提：对所有的解释变量和被解释变量进行单位根检验，若不平稳，在其单整阶数满足要求的条件下，进行如下两步操作。

第一步：用OLS法对回归方程：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t$$

进行估计，得到残差序列：

$$e_t = Y_t - \left(\hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \cdots + \hat{\beta}_k X_{kt}\right)$$

- 第二步，检验e_t的平稳性。若e_t为平稳的，则$\{Y_t\},\{X_{2t}\},\ldots,\{X_{kt}\}$ 是协整的，反之则不是协整的。因为若$\{Y_t\},\{X_{2t}\},\ldots,\{X_{kt}\}$不是协整的，则它们的任一线性组合都是非平稳的．因此残差e_t将是非平稳。换言之，对残差序列e_t是否具有平稳性的检验，也就是对 $\{Y_t\},\{X_{2t}\},\ldots,\{X_{kt}\}$ 是否存在协整的检验。

检验 e_t 为非平稳的假设可用两种方法：

一种方法是对残差序列进行DF检验，即对进行单位根检验，其检验方法在前面已介绍，但要注意的是，DF检验和ADF检验使用的临界值应该用Engle-Granger编制的专用临界值表：

$$C(\alpha) = \phi_\infty + \phi_1 / T + \phi_2 / T^2$$

其中，*T*表示样本容量，其它值见附录。

协整回归DW检验

具体做法：用协整回归所得的残差构造DW统计量：

$$CRDW=\frac{\sum(e_t-e_{t-1})^2}{\sum e_t^2}$$

若$\{e_t\}$是随机游动的，可以证明DW也应接近于0。因此，只需检验

$$H_0\text{：}DW=0$$

是否成立，若成立，$\{e_t\}$为随机游走，X_t与Y_t间不存在协整，反之则存在协整。

Sargan和Bhargava最早编制了用于检验协整的DW临界值表。表10.2是观察数为100时，该检验的临界值。例如，当DW＝0.71时，在1％的显著性水平上我们能拒绝，即拒绝非协整假设。

表10.2　检验DW=0的临界值

显著性水平%	DW临界值
1	0.511
5	0.386
10	0.322

均衡与误差修正机制

- 均衡指一种状态，达到均衡时将不存在破坏均衡的内在机制。即当系统受到干扰后会偏离均衡点，而内在均衡机制将努力使系统重新回到均衡状态

- 若两个变量x_t，y_t永远处于均衡状态，则偏差为零。然而由于各种因素的影响，x_t，y_t并不是永远处于均衡位置上，从而使$u_t \neq 0$，称u_t为非均衡误差

误差修正模型

- 当系统偏离均衡点时，平均来说，系统将在下一期移向均衡点。这是一个动态均衡过程。
- 本期非均衡误差u_t是y_t下一期取值的重要解释变量。当$u_t > 0$时，说明y_t相对于x_t取值高出均衡位置。平均来说，变量y_t在T+1期的取值y_{t+1}将有所回落。
- 因此，$u_t = f（y_t，x_t）$具有一种误差修正机制。

三、误差修正模型
（Error Correction Model ,ECM）

误差修正模型(ECM，也称误差修正模型)是一种具有特定形式的计量经济模型。

建立误差修正模型一般采用两步，分别建立区分数据长期特征和短期待征的计量经济学模型。

第一步，建立长期关系模型。即通过水平变量和OLS法估计出时间序列变量间的关系。若估计结果形成平稳的残差序列时，那么这些变量间就存在相互协整的关系．长期关系模型的变量选择是合理的，回归系数具有经济意义。

第二步，建立误差修正模型。将长期关系模型 各个变量以一阶差分形式重新构造，并将第一步中的残差引入。在一个从一般到特殊的检验过程中，对短期动态关系进行逐项检验，剔除不显著项，直到得到最适当的模型形式。

注意，解释变量引入的长期关系模型的残差，代表着在取得长期均衡的过程中各时点上出现“偏误”的程度，使得第二步可以对这种偏误的短期调整或误差修正机制加以估计。

举例:货币需求函数

以建立我国货币需求函数为例，说明误差修正模型的建模过程。

货币需求函数通常在局部调整的结构下加以设定。在这种模型中，当前实际货币需求余额是关于实际货币需求余额滞后值、实际国民收入(通常用GDP表示)和机会成本等变量的回归。那么这种依据交易方程设定的模型可作为长期关系模型。

其一般形式为：

$$(\frac{M}{P})_t = \beta_0 + \beta_1 Y_t + \beta_2 \pi_t + \beta_3 (\frac{M}{P})_{t-1} + \upsilon_t$$

其中：M 为相应的名义货币余额，P 为物价指数(通常用GDP的平减指数表示)，Y 为实际的国民收入(GDP)，π 为季度通货膨胀率(根据综合物价指数衡量)。这里关于实际收入(产业规模)和机会成本变量的长期弹性分别由 $\beta_1/(1-\beta_3)$ 和 $\beta_2/(1-\beta_3)$ 给出。

第二阶段误差修正方程的一般形式是：

$$\Delta(\frac{M}{P})_t = \alpha_0 + \sum_{i=0}^{l} \beta_i \Delta Y_{t-i} + \sum_{i=0}^{l} \gamma_i \Delta \pi_{t-i} + \sum_{i=0}^{l} \sigma_i \Delta(\frac{M}{P})_{t-i-1} + \lambda \mathrm{EC}_{t-1} + \upsilon_t$$

其中，EC ＝长期关系模型中的残差。

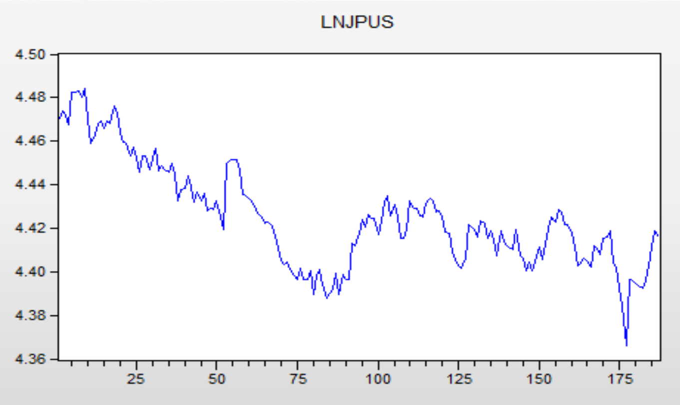
在具体建模中，首先要对长期关系模型的设定是否合理进行单位根检验，以保证 EC为平稳序列。其次，对短期动态关系中各变量的滞后项，进行从一般到特殊的检验，将不显著的滞后项逐渐剔除，直到找出了最佳形式为止。通常滞后期在 ＝ 0, 1, 2, 3 中进行试验。

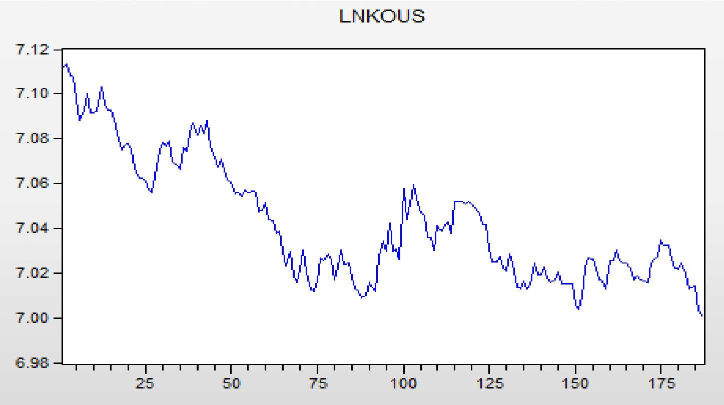
案例分析

在许多关于国际汇率问题的研究中，研究者都发现：（1）汇率一般都存在单位根；（2）不同货币的汇率之间往往存在联系。在本章的案例分析中，使用日元兑美元与韩元兑美元的汇率作为例子，来验证以上的两个结论，分析两者之间是否存在长期均衡关系。

本案例使用的汇率数据来自FRED Economic Data，选取了从2010年7月1日至2011年3月31日的日元兑美元（jpus）与韩元兑美元（kous）的当日午间买入汇率，两个时间序列各187个观测值（具体数据见表10.3）。

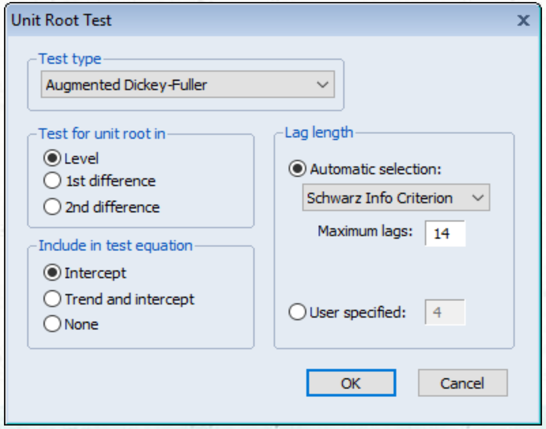
- 在EViews中建立文档，录入日元兑美元（jpus）与韩元兑美元（kous）序列的数据。依照惯例，我们将汇率数据取自然对数，分别生成lnjpus和lnkous序列。双击打开序列，点击View / Graph…，可以直观地看到两序列关于时间的变化。





考察变量的单整阶数

- ADF检验的操作如下：双击打开Injpus序列，并点击View打开下拉菜单，选择Unit Root Test…，随后出现对话窗口。其他选项保持不变，滞后阶数Lag length选择Automatic selection，并在下拉菜单中选择Schwarz Info Criterion，即用SIC这种方法自动选择ADF中的修项的阶数，如图10.6。选择OK即得到ADF检验的结果。



注：最好就用ADF检验，而不要用DF检验（强制令阶数为0）。DF检验是ADF检验的特例，如果真是DF检验就可以，那ADF检验的最优阶数也会选到0。

考察变量的单整阶数

- 从检验结果看，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-3.4656、-2.8769、-2.5751， t检验统计量值-2.392625大于相应临界值，从而不能拒绝 H_0 ，表明lnjpus序列存在单位根，是非平稳序列。

Null Hypothesis: LNJPUS has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.392625	0.1452
Test critical values:	1% level	-3.465585	
	5% level	-2.876927	
	10% level	-2.575051	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LNJPUS)
Method: Least Squares
Date: 11/06/19 Time: 17:10
Sample (adjusted): 2 187
Included observations: 186 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNJPUS(-1)	-0.048380	0.020221	-2.392625	0.0177
C	0.213790	0.089478	2.389299	0.0179

R-squared	0.030173	Mean dependent var	-0.000295
Adjusted R-squared	0.024903	S.D. dependent var	0.006649
S.E. of regression	0.006566	Akaike info criterion	-7.203136
Sum squared resid	0.007933	Schwarz criterion	-7.168450
Log likelihood	671.8916	Hannan-Quinn criter.	-7.189080
F-statistic	5.724655	Durbin-Watson stat	2.177509
Prob(F-statistic)	0.017735		

考察变量的单整阶数

由此知道lnjpus序列是非平稳的。为了确定序列的单整阶数，对lnjpus的一阶差分序列进行单位根检验，即在单位根检验窗口的“Test for unit root in”中选择“1st difference”，其他选项保持不变。从检验结果看，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-3.4658、-2.8770、-2.5751，**t检验统计量值为-15.10486，小于相应临界值，从而拒绝H_0，表明lnjpus的差分序列不存在单位根，是平稳序列。即lnjpus序列是一阶单整的，lnjpus～I（1）。**

Null Hypothesis: D(LNJPUS) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-15.10486	0.0000
Test critical values:	1% level	-3.465780	
	5% level	-2.877012	
	10% level	-2.575097	

*MacKinnon (1996) one-sided p-values.

考察变量的单整阶数

对lnkous进行同样的操作，得到lnkous序列也是一阶单整的。

这与文献中大部分关于汇率的研究结果相同。

Null Hypothesis: LNKOUS has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.256176	0.1875
Test critical values:	1% level	-3.465585	
	5% level	-2.876927	
	10% level	-2.575051	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(LNKOUS) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-15.09621	0.0000
Test critical values:	1% level	-3.465780	
	5% level	-2.877012	
	10% level	-2.575097	

*MacKinnon (1996) one-sided p-values.

协整检验-EG两步法

为了分析lnjpus和lnkous序列之间是否存在协整关系，使用EG两步法，先作两变量之间的回归，然后检验回归残差的平稳性。

第一步，以lnkous为被解释变量，以lnjpus为解释变量，用OLS方法估计回归模型。

Dependent Variable: LNKOUS
Method: Least Squares
Date: 11/06/19 Time: 17:34
Sample: 1 187
Included observations: 187

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.710116	0.193068	14.03708	0.0000
LNJPUS	0.979048	0.043631	22.43931	0.0000

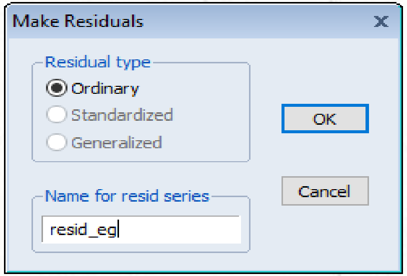
R-squared	0.731309	Mean dependent var	7.042372
Adjusted R-squared	0.729856	S.D. dependent var	0.027269
S.E. of regression	0.014173	Akaike info criterion	-5.664295
Sum squared resid	0.037163	Schwarz criterion	-5.629737
Log likelihood	531.6116	Hannan-Quinn criter.	-5.650292
F-statistic	503.5226	Durbin-Watson stat	0.407049
Prob(F-statistic)	0.000000		

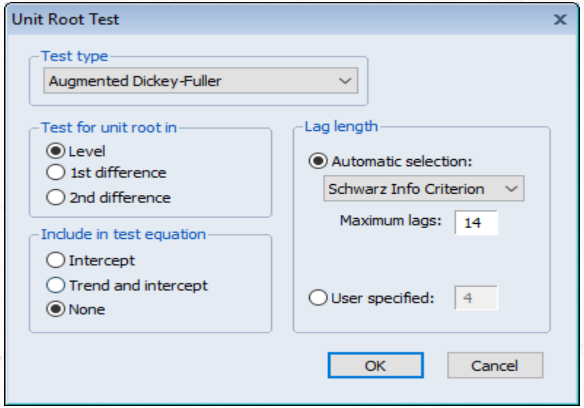
估计的回归模型为：

$$\ln kous_t = 2.710116 + 0.979048 \ln jpus_t + e_t$$

协整检验-EG两步法

- 第二步，对上述回归的残差进行平稳性检验。在回归结果窗口中，点击Proc -> Make Residual Series，将回归的残差命名为resid_eg，并对此序列进行无截距项、无趋势项的ADF检验（即在单位根检验的窗口中，“Include in estimate equation”选择“None”），得到的单位根检验的结果





协整检验-EG两步法

结果图中的t统计量为-4.499077，根据公式：

$$C(\alpha)=\phi_{\infty}+\phi_1/T+\phi_2/T^2$$

与本书附录表6，在5%的显著性水平下，EG两步法检验的临界值为

$$-3.3377-5.967/187-8.98/187^2=-3.3699$$

大于t统计量，从而拒绝H_0，表明残差序列不存在单位根，是平稳序列，说明日元lnjpus和韩元lnkous之间存在协整关系。

Null Hypothesis: RESID_EG has a unit root
Exogenous: None
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.499077	0.0000
Test critical values:	1% level	-2.577454	
	5% level	-1.942545	
	10% level	-1.615565	

*MacKinnon (1996) one-sided p-values.

注意：这里不能直接看p值！

误差修正模型

两序列之间存在协整，表明两者之间有长期均衡关系，但从短期来看，可能会出现失衡。为了增强模型的精度，可以把协整回归式

$$\boldsymbol{ln\ kous_t = 2.710116 + 0.979048ln\ jpus_t + e_t}$$

中的误差项$\boldsymbol{e_t}$看作均衡误差，通过建立误差修正模型把韩币兑美元汇率的短期行为与长期变化联系起来。误差修正模型的结构如下：

$$\boldsymbol{\Delta ln\ kous_t = \alpha + \beta \Delta ln\ jpus_t + \gamma e_{t-1} + \varepsilon_t}$$

其中$\boldsymbol{e_{t-1}}$即是生成的残差序列resid_eg。

误差修正模型

要估计ECM，我们首先要生成lnkous与lnjpus的一阶差分序列。点击Genr，并输入“dlnkous = lnkous - lnkous(-1) ”和“dlnjpus = lnjpus - lnjpus(-1) ”。

以dlnkous为被解释变量，dlnjpus和resid_eg(-1)为解释变量，估计回归式。在EViews命令栏中输入

ls dlnkous c dlnjpus resid_eg(-1)

回车后可得方程输出结果

也可以直接输入：**ls d(lnkous) c d(lnjpus) resid_eg(-1)**

误差修正模型

Dependent Variable: DLNKOUS
Method: Least Squares
Date: 11/06/19 Time: 18:07
Sample (adjusted): 2 187
Included observations: 186 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000572	0.000435	-1.314333	0.1904
DLNJPUS	0.025496	0.066706	0.382216	0.7027
RESID_EG(-1)	-0.119601	0.031765	-3.765138	0.0002

R-squared	0.072365	Mean dependent var	-0.000601
Adjusted R-squared	0.062227	S.D. dependent var	0.006124
S.E. of regression	0.005930	Akaike info criterion	-7.401593
Sum squared resid	0.006435	Schwarz criterion	-7.349565
Log likelihood	691.3482	Hannan-Quinn criter.	-7.380510
F-statistic	7.137922	Durbin-Watson stat	2.122393
Prob(F-statistic)	0.001035		

最终得到误差修正模型的估计结果：

$$\Delta \widehat{ln\ kous_t} = -0.000572 + 0.025496\Delta ln\ jpus_t - 0.119601e_{t-1}$$
$$t = (-1.314333) \qquad (0.382216) \qquad (-3.765138)$$

从回归结果可以看到，误差修正项前的系数的估计值为-0.1196，且显著。

误差修正模型

- 上述结果表明，日币与韩币兑美元的汇率之间存在协整关系，韩币兑美元汇率的变化一方面取决于日币兑美元汇率的变化，另一方面还取决于上一期韩币兑美元汇率对均衡水平的偏离。当上一期的汇率偏离两者的长期均衡关系时，误差修正机制使汇率会更倾向于回到它们的长期均衡，误差项 $\boldsymbol{e_{t-1}}$ 估计的系数-0.1196体现了对偏离的修正，上一期偏离越远，本期修正的量就越大，即系统存在误差修正机制。