

第五章

时间序列 计量经济学模型

第五章 时间序列计量经济学模型

5.1 时间序列的平稳性及其检验

5.2 协整与误差修正模型

5.3 格兰杰因果关系检验

5.1 时间序列的平稳性及其检验

经典回归分析暗含的重要假设：数据是平稳的。

非平稳数据的后果——伪回归（两个本来没有因果关系的变量，却有很高的相关性，表现出较高的 R^2 ）

时间序列分析方法：以通过揭示时间序列自身的变化规律为主线，发展起来的计量经济学方法。

一、进行平稳性检验的原因

（一）平稳序列才满足基本假设

时间序列的平稳性可以替代随机抽样假定，采用平稳时间序列作为样本，建立经典计量经济学模型，在模型设定正确的前提下，模型随机干扰项仍满足极限法则和经典模型的基本假设（序列无关假设除外）。

（二）避免伪回归问题产生

采用平稳时间序列建立经典计量经济学结构模型，可以有效减少虚假回归。

问题：是否只有非平稳序列之间才会出现伪回归？

从逻辑上说，一个统计关系式，不管多强或多么有启发性，本身不可能意味着任何因果关系。要谈因果关系，必须来自统计学之外，诉诸先验的或者理论上的思考。

——古扎拉蒂

所以，伪回归不仅可以出现在非平稳序列之间，也可以出现在平稳时间序列和截面数据之间。但是，非平稳时间序列之间出现伪回归的可能性更大。

【注意】 杜绝伪回归的根本方法，是正确的模型设定。

二、时间序列平稳性的相关概念

随机过程：有些随机现象，要认识它必须研究它的发展

变化过程，随机现象的动态变化过程就是随机过程。

【例】考察一段时期内每天电话被呼叫次数，需要考察依赖于时间 t 的随机变量 v_t ， $\{v_t\}$ 就是一个随机过程。

【例】某国某年的GDP总量，是一个随机变量，考察它随时间变化的情形， $\{GDP_t\}$ 就是一个随机过程。

时间序列平稳性的概念：

假定某个时间序列是由某一随机过程生成，即 $\{X_t\}$ 的每一个数值都是从一个概率分布中随机得到的， $\{X_t\}$ 满足

- (1) 均值 $E(X_t) = \mu$ 是与时间 t 无关的常数
- (2) 方差 $\text{Var}(X_t) = \sigma^2$ 是与时间 t 无关的常数
- (3) 协方差 $\text{Cov}(X_t, X_{t+k}) = \gamma_k$ 只与 k 有关，与时间 t 无关

则称该序列是平稳的，该随机过程是一个平稳随机过程。

白噪声：白噪声过程（white noise）是最简单的随机时间序列，即具有零均值、同方差的独立分布序列。这种序列被称为是一个白噪声。

$$X_t = \mu_t, \quad \mu_t \sim N(0, \sigma^2)$$

白噪声序列满足：

$$\left\{ \begin{array}{l} E(X_t) = 0 \\ \text{Var}(X_t) = \sigma^2 \\ \text{Cov}(X_t, X_{t+k}) = 0 \end{array} \right.$$

【注意】白噪声序列是平稳的。

随机游走：随机游走 (random walk) 由随机过程

$X_t = X_{t-1} + \mu_t$ 生成， μ_t 是一个白噪声。

设初始值为 X_0 (X_0 为常数， μ_t 为白噪声)，则

$$\begin{cases} X_1 = X_0 + \mu_1 \\ X_2 = X_1 + \mu_2 = X_0 + \mu_1 + \mu_2 \\ \dots \\ X_t = X_{t-1} + \mu_t = X_0 + \mu_1 + \dots + \mu_t \end{cases}$$

均值： $E(X_t) = E(X_{t-1}) + E(\mu_t) = E(X_{t-1})$

方差： $Var(X_t) = Var(X_0) + Var(\mu_1) + \dots + Var(\mu_t) = t\sigma^2$

因此，随机游走序列为非平稳序列。

对 $\{X_t\}$ 取一阶差分(first difference)

$$\Delta X_t = X_t - X_{t-1} = \mu_t$$

由于 μ_t 是一个白噪声，因此序列 $\{\Delta X_t\}$ 是平稳的。

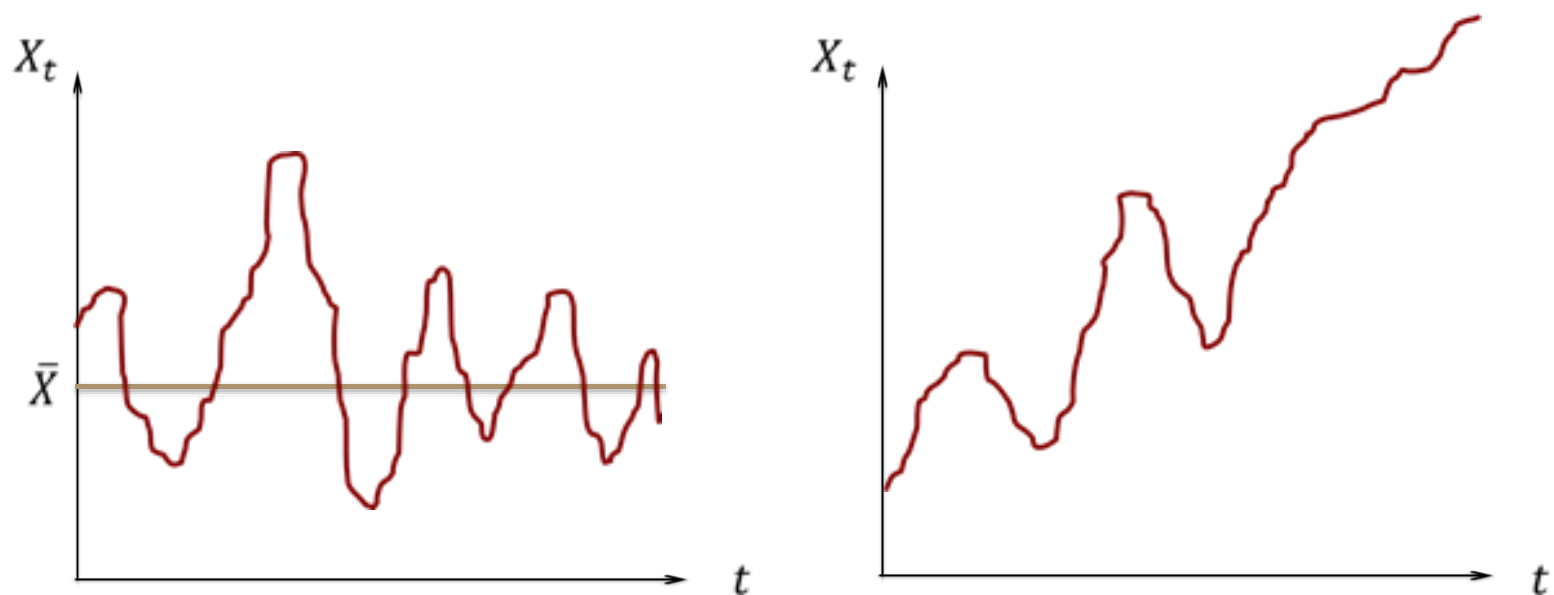
事实上，对于非平稳时间序列，往往都可以通过取差分方式使其形成平稳序列。

三、平稳性检验的图示判断

1. 序列图判断平稳性

平稳过程：围绕均值不断波动。

非平稳过程：不同时间段有不同的均值（如持续上升或下降）。



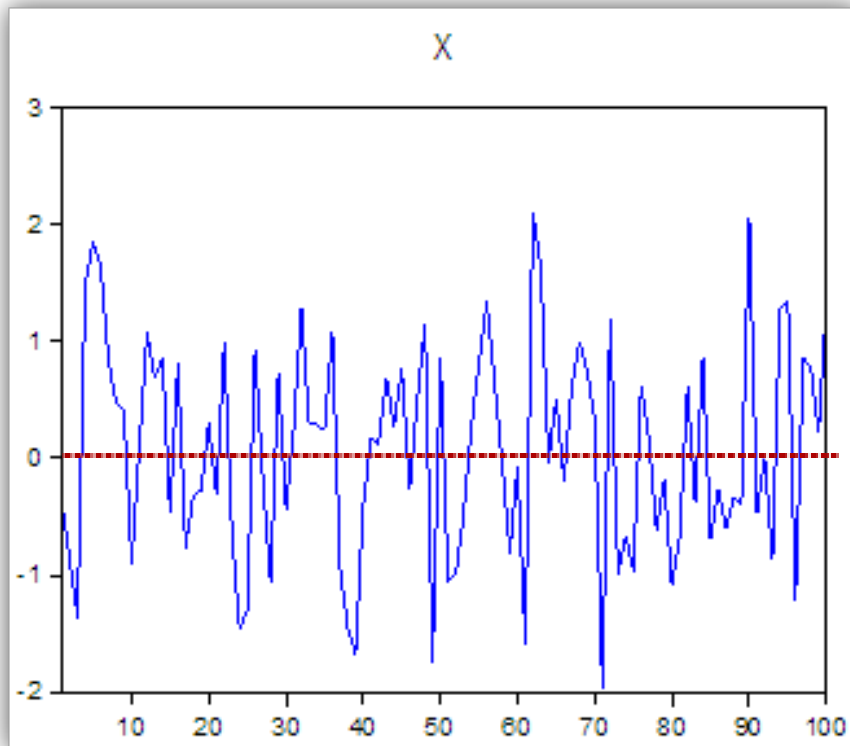


图1 白噪声序列

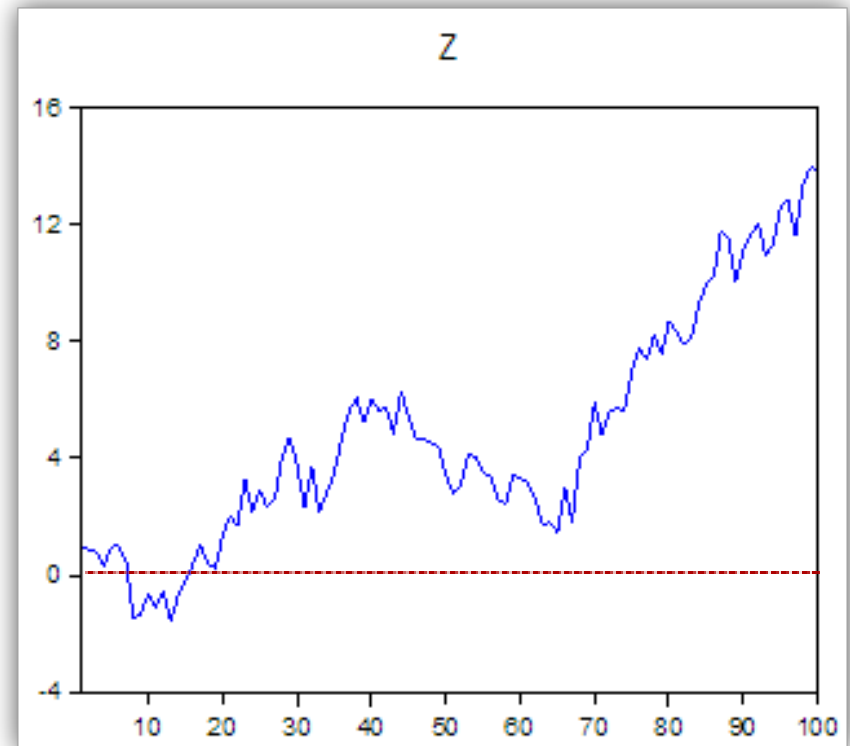
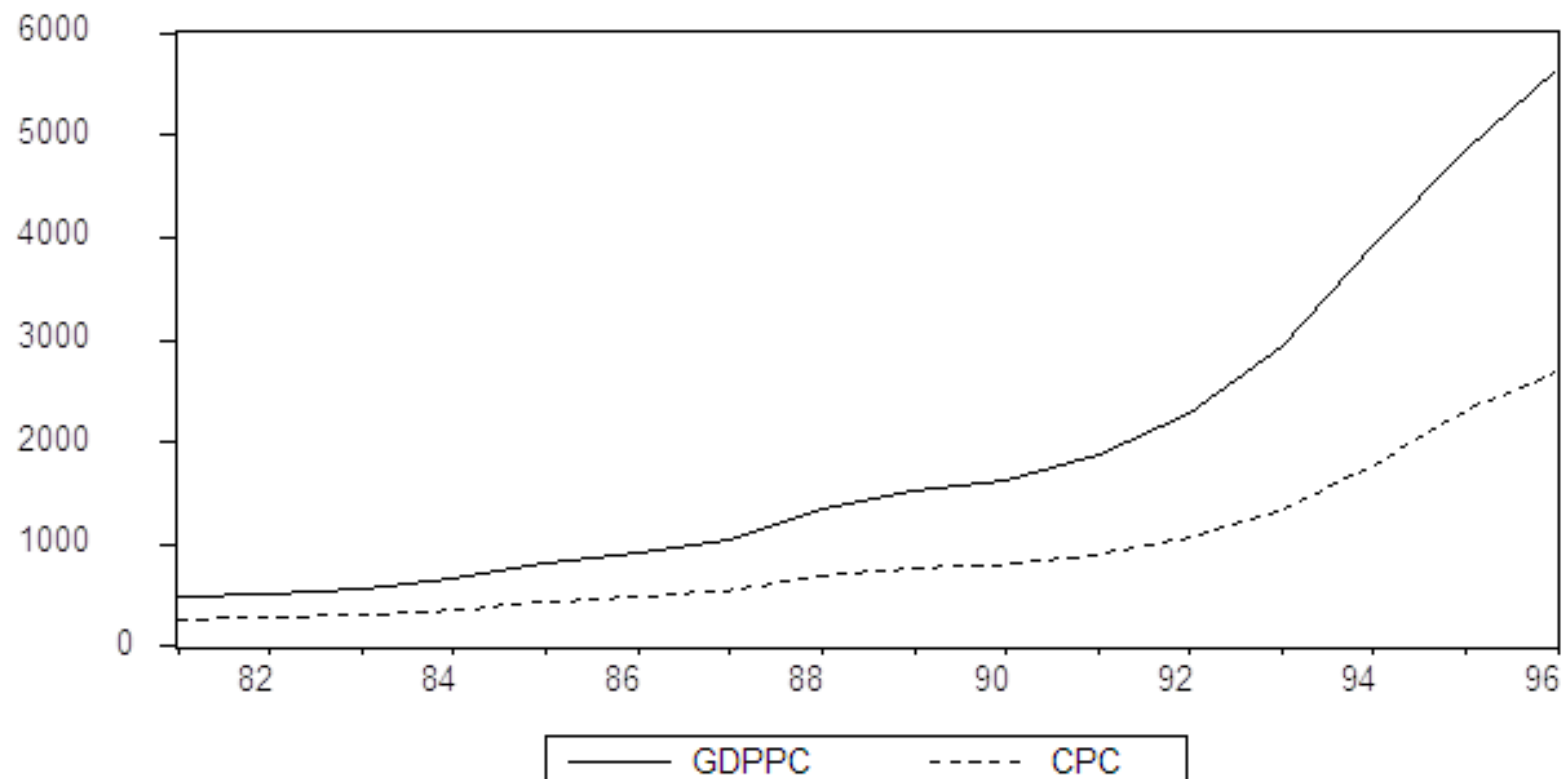


图2 随机游走序列



【例】从图形上来看，人均居民消费CPC与人均国内生产总值GDPPC是非平稳的，就此来说，运用传统方法对其进行回归是没有意义的。

2. 自相关函数图 (ACF) 判断平稳性

直观的图示往往会产生误导，因此需要进行进一步判别。通常采用检验样本自相关函数及图形，一个随机过程只有一个样本，因此，只能计算样本自相关函数，定义为

$$\gamma_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}$$
$$k = 1, 2, 3 \dots$$

- 随着 k 的增加，样本自相关函数逐渐下降且趋于零
- 但是从下降速度来看，平稳序列要比非平稳序列快得多。

四、平稳性的单位根检验

随机游走序列

$$X_t = X_{t-1} + \mu_t$$

可以看做是随机游走模型

$$X_t = \rho X_{t-1} + \mu_t$$

中 $\rho = 1$ 时的特例，即发现 $\rho = 1$ ，则称随机变量 X_t 有一个**单位根**（一个有单位根的时间序列就是随机游走序列，是非平稳的）。

因此，要判断某时间序列是否平稳，可通过

$$X_t = \rho X_{t-1} + \mu_t$$

判断它是否有单位根。这就是时间序列的**单位根检验**。

(一) DF 检验

$$X_t = \alpha + \rho X_{t-1} + \mu_t$$
$$\Delta X_t = \alpha + (\rho - 1)X_{t-1} + \mu_t$$

DF 检验原理：判断某序列是否平稳，即判断它是否有单位根。

DF 检验方法：对带有截距项的一阶自回归模型

$$\Delta X_t = \alpha + \delta X_{t-1} + \mu_t$$

进行 OLS 下的 t 检验（ DF 分布下）。

事实上：

$$X_t = \rho X_{t-1} + \mu_t \text{ — } \rho \geq 1 \text{ 时, } X_t \text{ 非平稳}$$

$$\Delta X_t = \alpha + \delta X_{t-1} + \mu_t \text{ — } \delta \geq 0 \text{ 时, } X_t \text{ 非平稳}$$

非平稳

DF检验步骤：

<1> 假设 $H_0 : \delta = 0$, $H_1 : \delta < 0$

<2> 计算统计量 $t = \frac{\hat{\delta}}{s_{\hat{\delta}}}$

这里的 t 统计量也称 τ 统计量，呈现围绕小于0值的偏态分布

<3> 查DF分布临界值表，得到临界值（左侧单尾检验）

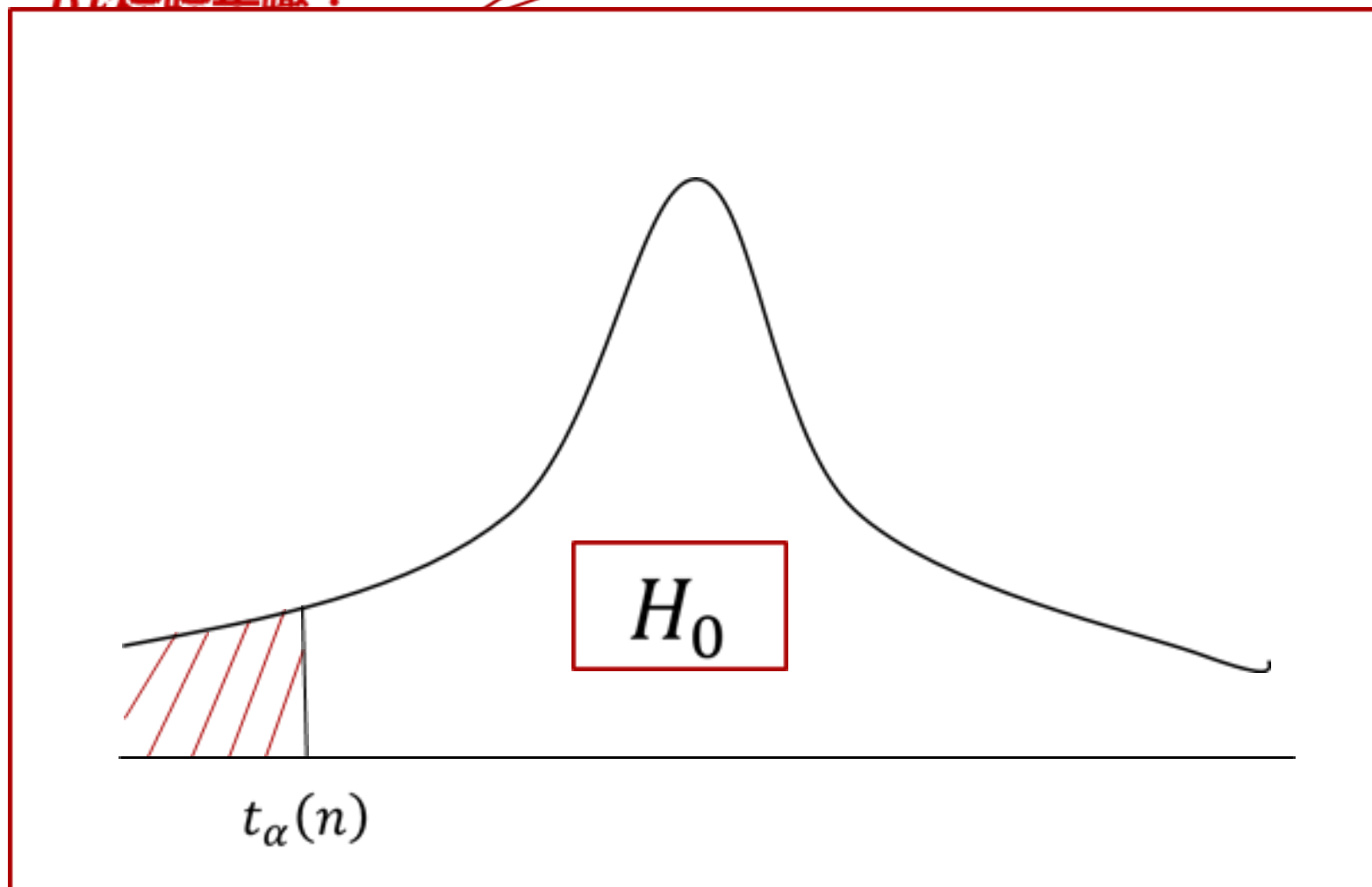
<4> 比较，判断

若 $t > t_{\delta}$ ，则接受 H_0 ，原时间序列存在单位根，是非平稳序列。

若 $t < t_{\delta}$ ，则接受 H_1 ，原时间序列不存在单位根，是平稳序列。

非平稳

拒绝域。



15.75%

表5-1 DF 分布临界值表

显著性 水平	样本容量					t 分布临界值 ($n = +\infty$)
	25	50	100	500	∞	
0.01	-3.75	-3.58	-3.51	-3.44	-3.43	-2.33
0.05	-3.00	-2.93	-2.89	-2.86	-2.86	-1.65
0.10	-2.63	-2.60	-2.58	-2.57	-2.57	-1.28

在实际检验中

$$\Delta X_t = \alpha + \delta X_{t-1} + \mu_t$$

中的 μ_t 可能出现自相关，导致 DF 检验无效，为了保证 μ_t 的白噪声特性，Dickey和Fuller对 DF 检验进行了扩展，形成了 ADF 检验（augment Dickey-Fuller test）。

引入 ADF 检验的目标：保证 DF 检验中随机干扰项的白噪声特性。

(二) ADF检验

ADF检验通过以下三个模型完

增加 ΔX_t 的滞后项，是为了消除时间序列由更高阶的自回归过程生成时，模型随机干扰项的序列相关，保证随机干扰项是白噪声。滞后阶数一般依据LM检验或AIC准则和SC准则确定。

模型1 $\Delta X_t = \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$

常数项

模型2 $\Delta X_t = \alpha + \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$

趋势项：代表序列随时间变化的某种趋势

模型3 $\Delta X_t = \alpha + \beta T + \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$

检验步骤：实际检验时从模型3开始，然后模型2、模型1。

（只要有一个模型拒绝存在单位根，就认为序列平稳；三个模型都不拒绝，就是非平稳序列。）

检验原理：与 DF 检验相同，只是对模型1、2、3进行检验时，有各自相应的临界值表。

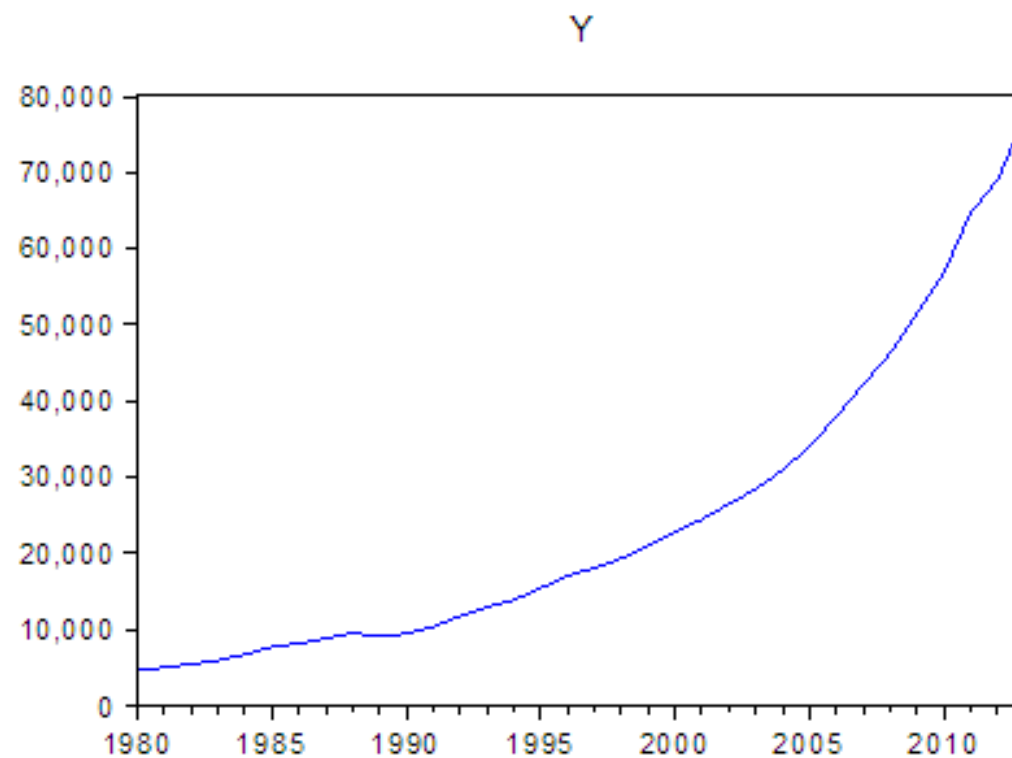
模型滞后项阶数的确定：以随机项不存在序列相关为准则。
（软件利用AIC、SC准则自动确定）

【例】 1980-2013年，中国名义居民消费总量（CONS）数据，消除价格因素后，得到以1990年价格计算的实际居民消费总量（ Y_t ）的时间序列数据

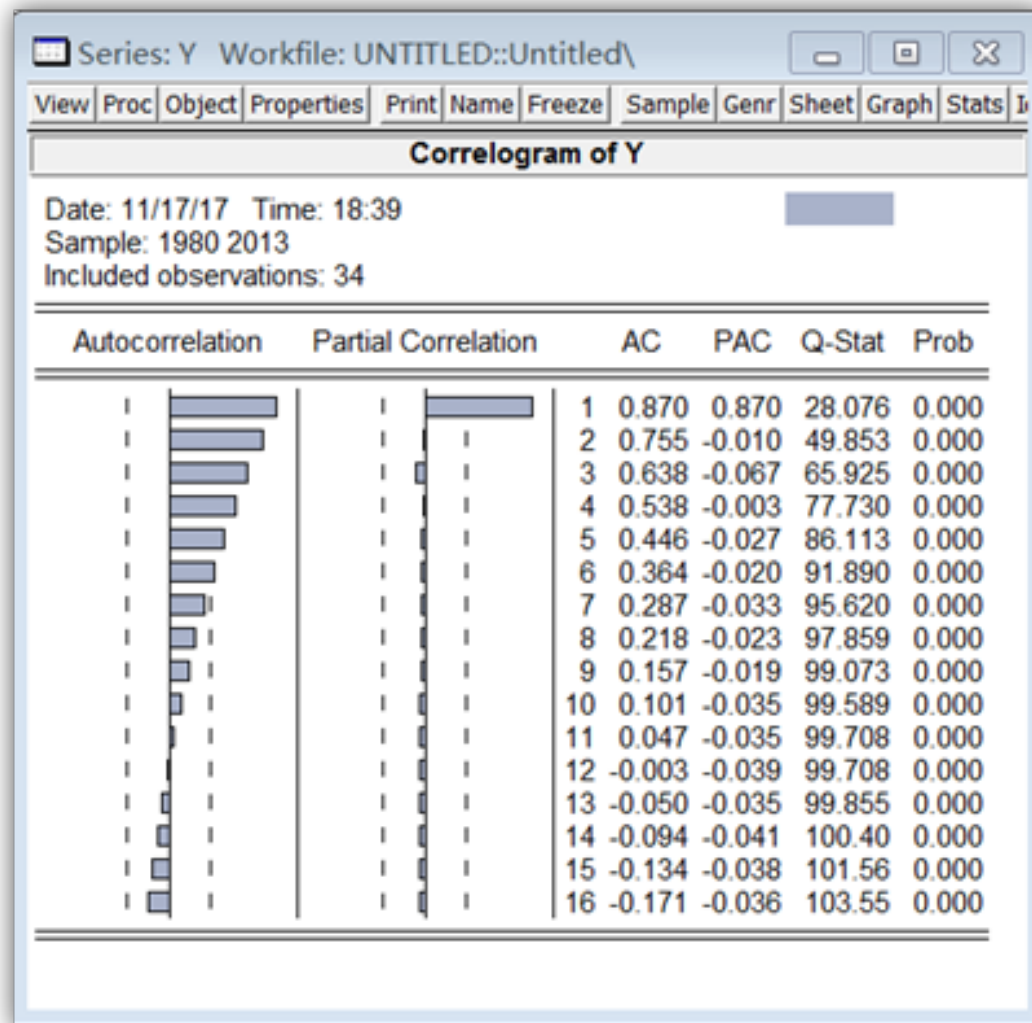
Obs	Y	Obs	Y
1980	4605.29	1997	18080.16
1981	5063.90	1998	19363.89
1982	5482.34	1999	20989.59
1983	5983.52	2000	22864.42
1984	6745.99	2001	24480.49
1985	7728.61	2002	26485.92
1986	8211.40	2003	28436.74
1987	8839.97	2004	30963.54
1988	9560.27	2005	34026.07
1989	9085.15	2006	37939.58
1990	9450.90	2007	42232.57
1991	10375.75	2008	46232.67
1992	11815.05	2009	51530.08
1993	13004.83	2010	56817.07
1994	13944.59	2011	64712.02
1995	15467.91	2012	69002.39
1996	17092.47	2013	77198.39

1.对 Y_t 水平序列的检验

(1) 图示法判断平稳性



(2) 自相关系数判断平稳性



(3) ADF检验

$$\text{检验模型3 } \Delta X_t = \alpha + \beta T + \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$$

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	6.049722	1.0000
Test critical values:		
1% level	-4.273277	
5% level	-3.557759	
10% level	-3.212361	

时间趋势项 T 的 t_T 检验，就是一般的变量显著性检验，

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.169788	0.028065	6.049722	0.0000
D(Y(-1))	-0.448348	0.199390	-2.248597	0.0326
C	-232.0864	298.5248	-0.777444	0.4434
@TREND(1980)	-34.37253	35.34982	-0.972354	0.3392
R-squared	0.903684	Mean dependent var	2254.203	
Adjusted R-squared	0.893264	S.D. dependent var	2122.066	

结论：通过对模型3的检验，发现序列 Y 接受 $\delta = 0$ 的原假设，即序列存在单位根，为非平稳序列。

接受 $\beta = 0$ 的原假设，认为序列不存在时间趋势项。

需要继续检验模型2。

【注意】

如果检验在此时发生矛盾，即检验结果是接受零假设

$\delta = 0$ ，但是拒绝零假设 $\beta = 0$ ，即序列为非平稳，并且存在时间趋势项。此时回到模型2就不再合理。

一种经验处理方法是，采用正态分布临界值检验是否存在单位根，即将临界值适当放松，如果仍然存在单位根，即停止检验，得到该序列为非平稳的结论。

模型3的检验结果及应对策略：

- $\delta = 0$, $\beta = 0$ 继续检验模型2
- $\delta = 0$, $\beta \neq 0$ 放宽临界值，仍不通过，得到序列非平稳结论
- $\delta \neq 0$, $\beta = 0$ 直接得出结论，序列为平稳序列

检验模型2 $\Delta X_t = \alpha + \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	7.487675	1.0000
Test critical values: 1% level	-3.653730	
5% level	-2.957110	
10% level	-2.617434	

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

Test Equation				
Dependent Variable: D(Y)				
Method: Least Squares				
Date: 11/17/17 Time: 19:55				
Sample (adjusted): 1982 2013				
Included observations: 32 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.150801	0.020140	7.487675	0.0000
D(Y(-1))	0.422244	0.197389	2.139150	0.0410
C	-439.9432	208.1808	-2.113274	0.0433

结论：通过对模型2的检验，发现序列Y接受 $\delta = 0$ 的原假设，即序列存在单位根，为非平稳序列。

接受 $\alpha = 0$ 的原假设，认为序列不存在常数项。

需要继续检验模型1。

检验模型1 $\Delta X_t = \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	22.87095	1.0000
Test critical values:		
1% level	-2.636901	
5% level	-1.951332	
10% level	-1.610747	

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.101171	0.004424	22.87095	0.0000
R-squared	0.877902	Mean dependent var		2199.791
Adjusted R-squared	0.877902	S.D. dependent var		2112.781
S.E. of regression	738.2579	Akaike info criterion		16.07630
Sum squared resid	17440789	Schwarz criterion		16.12165
Log likelihood	-264.2589	Hannan-Quinn criter.		16.09156
Durbin-Watson stat	2.451442			

结论：通过对模型1的检验，发现序列Y接受 $\delta = 0$ 的原假设，即中国实际居民消费总量序列存在单位根，为非平稳时间序列。

五、单整时间序列

随机游走序列 $X_t = X_{t-1} + \mu_t$ 经差分后变形为

$$\Delta X_t = \mu_t$$

由于 μ_t 为白噪声序列，因此差分序列 $\{\Delta X_t\}$ 是平稳的。

- 如一个时间序列经过一次差分变成平稳的，就称原序列是**一阶单整序列**，记为 $I(1)$ 。
- 如一个时间序列经过 d 次差分后变成平稳序列，则称原序列是 **d 阶单整序列**，记为 $I(d)$ 。
- 平稳时间序列记做 $I(0)$ 。
- 如无论如何进行差分，都不能平稳的序列，称为非单整序列。

六、趋势平稳与差分平稳随机过程

- 为了避免因趋势造成的伪回归，通常引入作为趋势变量的时间，用这样包含有时间趋势变量的回归，可以消除这种趋势性的影响。
- 但这种做法，只有当趋势性变量是确定性的而非随机性的时，才会有效。
- 换言之，一个包含有某种确定性趋势的非平稳时间序列，可以通过引入表示这一确定性趋势的变量，而将确定性趋势分离出来。

对于含有一阶自回归的随机过程

$$X_t = \alpha + \beta t + \rho X_{t-1} + \mu_t$$

μ_t 为白噪声过程， t 为时间趋势

- 若 $\rho = 1$ ， $\beta = 0$ ，则 $X_t = \alpha + X_{t-1} + \mu_t$ ， X_t 成为一个带位移（截距项）的随机游走过程，根据 α 的正负， X_t 表现出明显的上升或者下降趋势，这种趋势称为随机性趋势，是典型的随机游走过程。
- 若 $\rho = 0$ ， $\beta \neq 0$ ，则 $X_t = \alpha + \beta t + \mu_t$ ， X_t 成为一个带时间趋势的随机变化过程，根据 β 的正负， X_t 表现出明显的上升或者下降趋势，这种趋势称为确定性趋势或趋势平稳过程。
- 若 $\rho = 1$ ， $\beta \neq 0$ ，则 $X_t = \alpha + \beta t + X_{t-1} + \mu_t$ 包含随机性和确定性两种趋势。

● 确定性趋势和随机趋势的辨别方法

辨别方法：通过ADF检验中所用的第3个模型判断是确定性趋势还是随机性趋势。

(该模型中已引入了表示确定性趋势的时间变量，即分离出了确定性趋势的影响。)

$$\text{模型3 } \Delta X_t = \alpha + \beta T + \delta X_{t-1} + \sum_{i=1}^m \beta_i \Delta X_{t-i} + \varepsilon_t$$

若检验结果表明所给时间序列有单位根， $\delta = 0$ ，且 $\beta = 0$ ，则该序列显示出**随机性趋势**；

若检验结果表明所给时间序列无单位根， $\delta < 0$ ，且 $\beta \neq 0$ ，则该序列显示出**确定性趋势**。

● 克服办法

1. 随机性趋势的消除方法——差分法

$X_t = \alpha + X_{t-1} + \mu_t$ 可以通过差分变换为 $\Delta X_t = \alpha + \mu_t$,
则该时间序列 X_t 称为差分平稳过程。

2. 确定性趋势的消除方法——除去趋势项

$X_t = \alpha + \beta t + \mu_t$ 通过去除 βt 变换为 $X_t - \beta t = \alpha + \mu_t$ 。

【注意】 趋势平稳过程代表了一个时间序列长期稳定的变化过程，因而用于进行长期预测则是更为可靠的。