第 13 章　平稳时间序列

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

13.1　时间序列的自相关

时间序列指同一个体在不同时点上的观测数据。

比如，在 1978-2022 年期间，中国每年的国内生产总值。

对于离散时间$\{1, 2, \cdots, T\}$，可将时间序列写为$\{y_1, y_2, \cdots, y_T\}$，其中每个$y_t$都是随机变量。

时间序列的最大特点是通常存在自相关，即不同期的观测值之间存在相关性。

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

$$\gamma_k \equiv \mathrm{Cov}(y_t, y_{t+k}) = \mathrm{E}\left[(y_t - \mu)(y_{t+k} - \mu)\right] \qquad (13.1)$$

其中，$\mu \equiv \mathrm{E}(y)$为总体均值。

γ_k反映同一变量(y)相隔k期之间的自相关程度。

当$k=0$时，$\gamma_0 \equiv \mathrm{Var}(y)$。

对γ_k的估计值为**样本自协方差**：

$$\hat{\gamma}_k \equiv \frac{1}{T-k}\sum_{t=1}^{T-k}(y_t-\overline{y})(y_{t+k}-\overline{y}) \qquad (13.2)$$

其中，$\overline{y} \equiv \dfrac{1}{T}\sum_{t=1}^{T} y_t$ 为样本均值。

自协方差受变量单位的影响，可将其标准化。

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

$$\rho_k \equiv \mathrm{Corr}(y_t, y_{t+k}) \equiv \frac{\mathrm{Cov}(y_t, y_{t+k})}{\mathrm{Var}(y_t)} \qquad (13.3)$$

自相关系数ρ_k将自协方差γ_k标准化为介于$[-1, 1]$之间的量。

对于严格平稳过程，ρ_k不依赖于具体时间，而仅仅是滞后阶数k的函数，故称为“自相关函数”(Autocorrelation Function，简记 ACF)。

将(k, ρ_k)画成图，即为“自相关图”(correlogram)。

对 ρ_k 的估计值为**样本自相关系数**：

$$\hat{\rho}_k \equiv \hat{\gamma}_k / \hat{\gamma}_0 \tag{13.4}$$

其中，$\hat{\gamma}_0 \equiv \dfrac{1}{T-1}\sum_{t=1}^{T}(y_t - \overline{y})^2$ 为样本方差。

这些时间序列的数字特征，都是时间序列固有的特征，不依赖于模型的设定。

在设定模型时，则希望尽可能地与这些数字特征相一致。

例 使用数据集 gdp_china.dta 考察在 1978-2013 年期间，中国的年度国内生产总值(1978 年不变价格，亿元)，记为 *y*。

定义时间变量后，首先看 GDP 的时间趋势(结果参见图 13.1)。

```
. use gdp_china.dta,clear

. tsset year

. tsline y,xlabel(1980(10)2010)
```

其中，“`tsline`”表示画时间趋势图，在此等价于命令“`line gdp year`”(`year` 为时间变量)。选择项“`xlabel(1980(10)2010)`”表示在横轴 1980-2010 期间，每隔 10 年做个标注(label)。

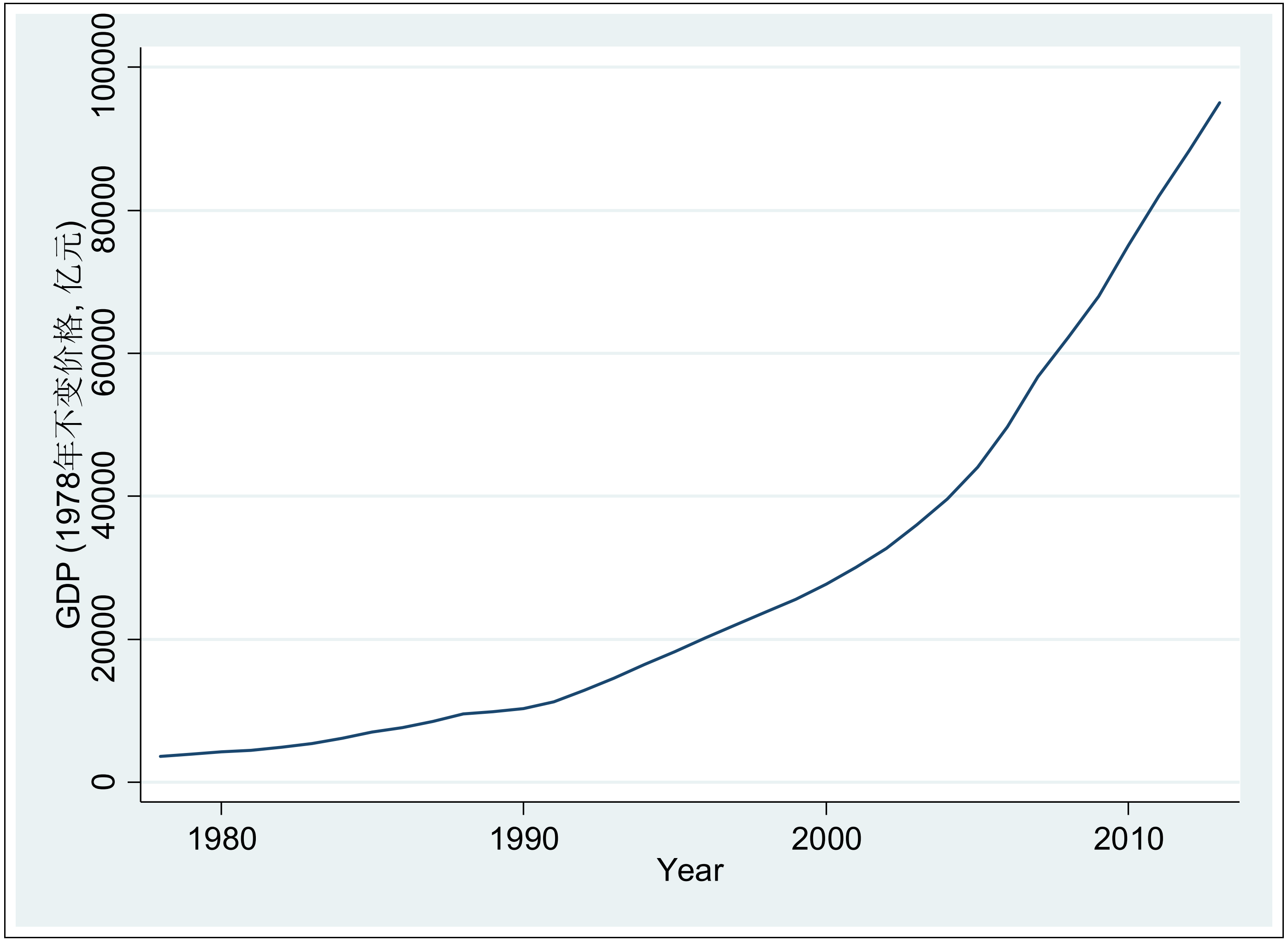

图 13.1　GDP 的指数增长趋势(1978-2013)

GDP 存在指数增长(exponential growth)的趋势，类似于“利滚利”的情形，即 $y_t \approx (1+r)^t$，其中 r 为 GDP 的平均增长率。

通常的处理方法是，将 GDP 取对数，把指数增长趋势变为线性增长趋势。

计算 GDP 对数，并再次画时间趋势图(结果参见图 13.2)。

```
. gen lny=log(y)

. tsline lny,xlabel(1980(10)2010)
```

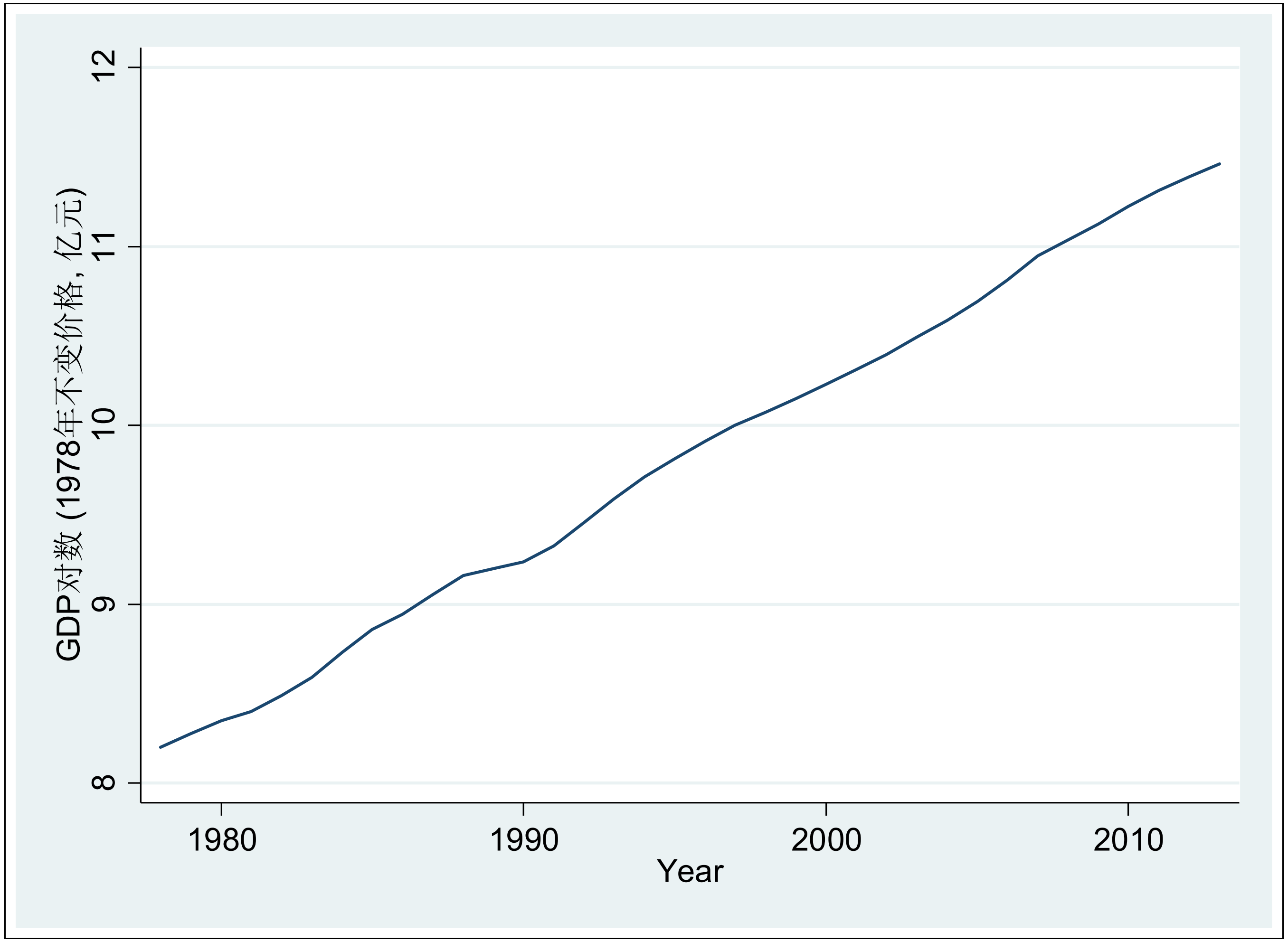

图 13.2　GDP 对数的线性增长趋势(1978-2013)

GDP 对数(lny)存在线性增长趋势，但依然不是平稳序列(其期望值不断增长)。

为此，将 GDP 对数进行一阶差分，即$\Delta \ln y_t = \ln y_t - \ln y_{t-1}$(可使用 Stata 的差分算子 `d.`)，然后再画时间趋势图(结果参见图 13.3)。

```
. gen dlny=d.lny

. tsline dlny,xlabel(1980(10)2010)
```

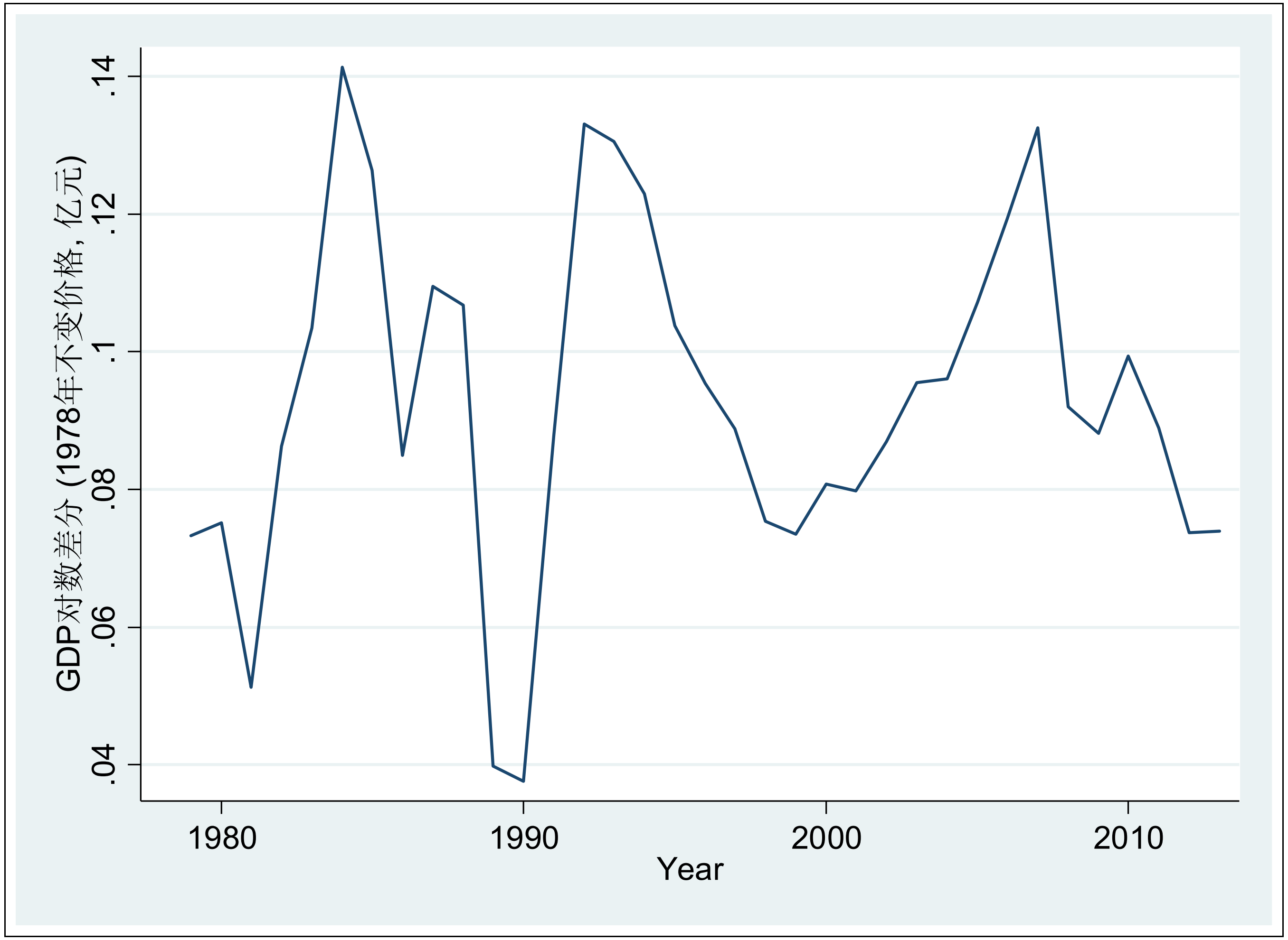

图 13.3　GDP 对数差分的时间趋势(1978-2013)

$\Delta \ln y_t$ 已不存在明显的时间趋势，可大致视为平稳序列。

之所以考察 GDP 对数差分，是因为它约等于 GDP 的增长率，

$$
\begin{aligned}
\Delta \ln y_t \equiv \ln y_t - \ln y_{t-1} &= \ln\left(\frac{y_t}{y_{t-1}}\right) \\
&= \ln\left(\frac{y_{t-1} + \Delta y_t}{y_{t-1}}\right) = \ln\left(1 + \frac{\Delta y_t}{y_{t-1}}\right) \approx \frac{\Delta y_t}{y_{t-1}}
\end{aligned}
\tag{13.5}
$$

根据泰勒展开的一阶近似，当 $x \approx 0$ 时，$\ln(1+x) \approx x$。

在有些研究中，直接将$\Delta \ln y_t$视为y_t的增长率(当然，如果此增长率较高，则误差较大)。

为了比较二者的差别，直接计算 GDP 的增长率(记为 g)，并与 GDP 对数差分进行画图比较(结果参见图 13.4)。

```
. gen g=(y-l.y)/l.y
(1 missing value generated)

. tsline dlny g,xlabel(1980(10)2010)
lpattern(dash)
```

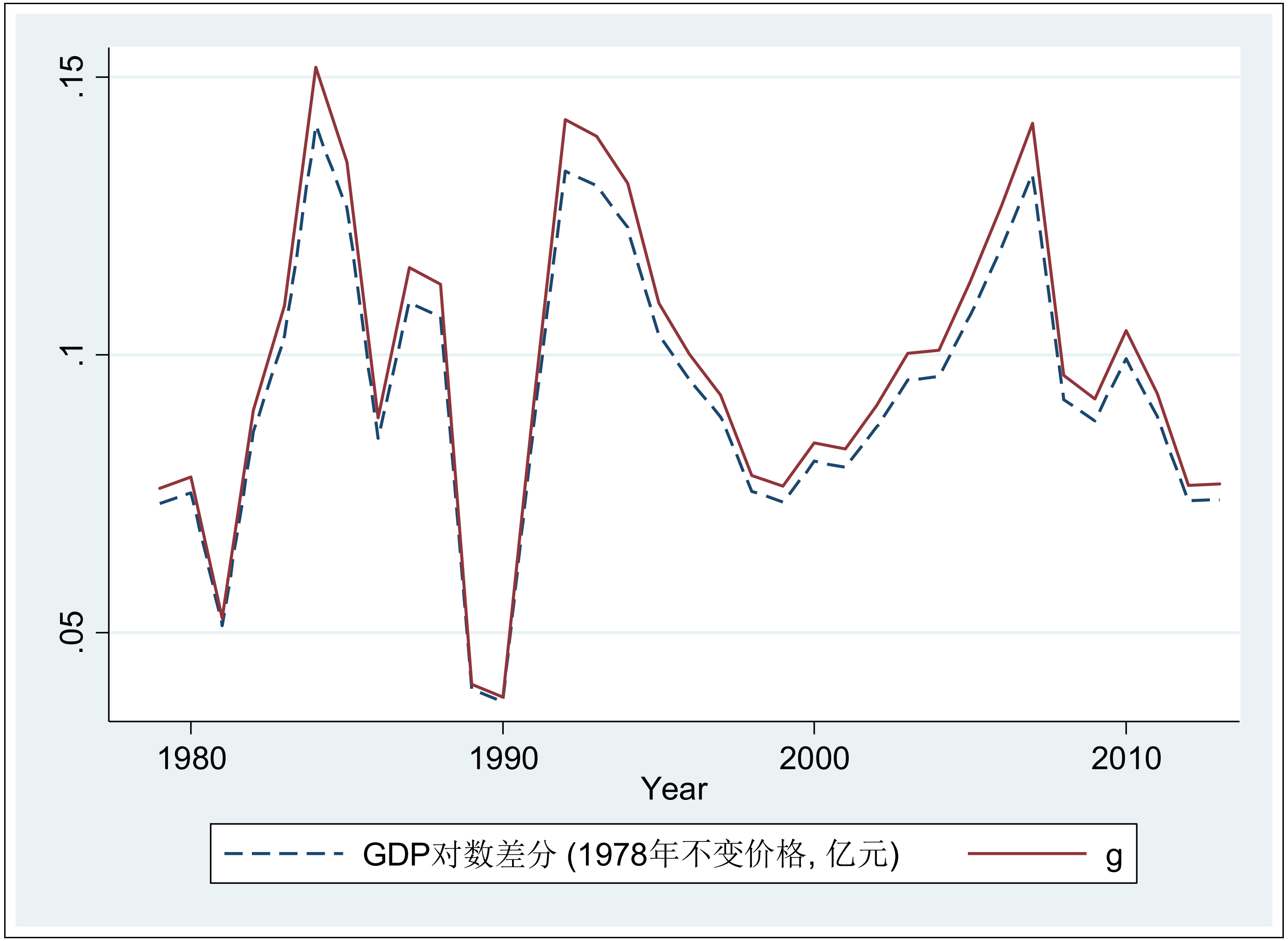

图 13.4　GDP 增长率的两种计算方法(1978-2013)

GDP 增长率(以实线表示)与 GDP 对数差分(以虚线表示)十分接近。

通过画自相关图，来考察 GDP 对数差分的各阶自相关系数。

```
. corrgram dlny
```

其中，“`corrgram`”表示 correlogram，即画自相关图。

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial autocor]
1	0.5360	0.5454	10.943	0.0009		
2	-0.0298	-0.4515	10.978	0.0041		
3	-0.2579	0.0205	13.669	0.0034		
4	-0.3405	-0.3311	18.514	0.0010		
5	-0.4687	-0.3671	27.998	0.0000		
6	-0.4371	-0.3092	36.531	0.0000		
7	-0.1425	-0.0790	37.47	0.0000		
8	0.1774	-0.0418	38.98	0.0000		
9	0.3220	-0.1600	44.143	0.0000		
10	0.2768	-0.0670	48.113	0.0000		
11	0.1179	-0.1722	48.863	0.0000		
12	0.0341	-0.0569	48.928	0.0000		
13	-0.0123	-0.2000	48.937	0.0000		
14	-0.0322	-0.0378	49.001	0.0000		
15	-0.0743	-0.0851	49.359	0.0000		

第 1 列表示滞后阶数(LAG)，第 2 列表示自相关系数(AC)。

右边倒数第 2 列为自相关图(以线条长度表示各阶自相关系数)。

第 3 列 PAC 为偏自相关系数(Partial Autocorrelation)，即在给定$(y_{t+1},\cdots,y_{t+k-1})$的情况下，$y_t$与 y_{t+k}的相关系数；右边倒数第 1 列为 PAC 的图示。

第 4-5 列为 Q 统计量(Q)及其相应的p值(Prob>Q)，其原假设为“各阶自相关系数均为 0”。

下面使用画自相关图的另一命令。

```
. ac dlny,lags(20)
```

其中，“ac”表示 autocorrelation ；选择项“lags(20)”表示画 1-20 阶的自相关图，默认所画的最高阶数为$\min\{\text{floor}(n/2)-2,\ 40\}$，其中$\text{floor}(n/2)$为不超过$n/2$的最大整数。结果参见图 13.5。

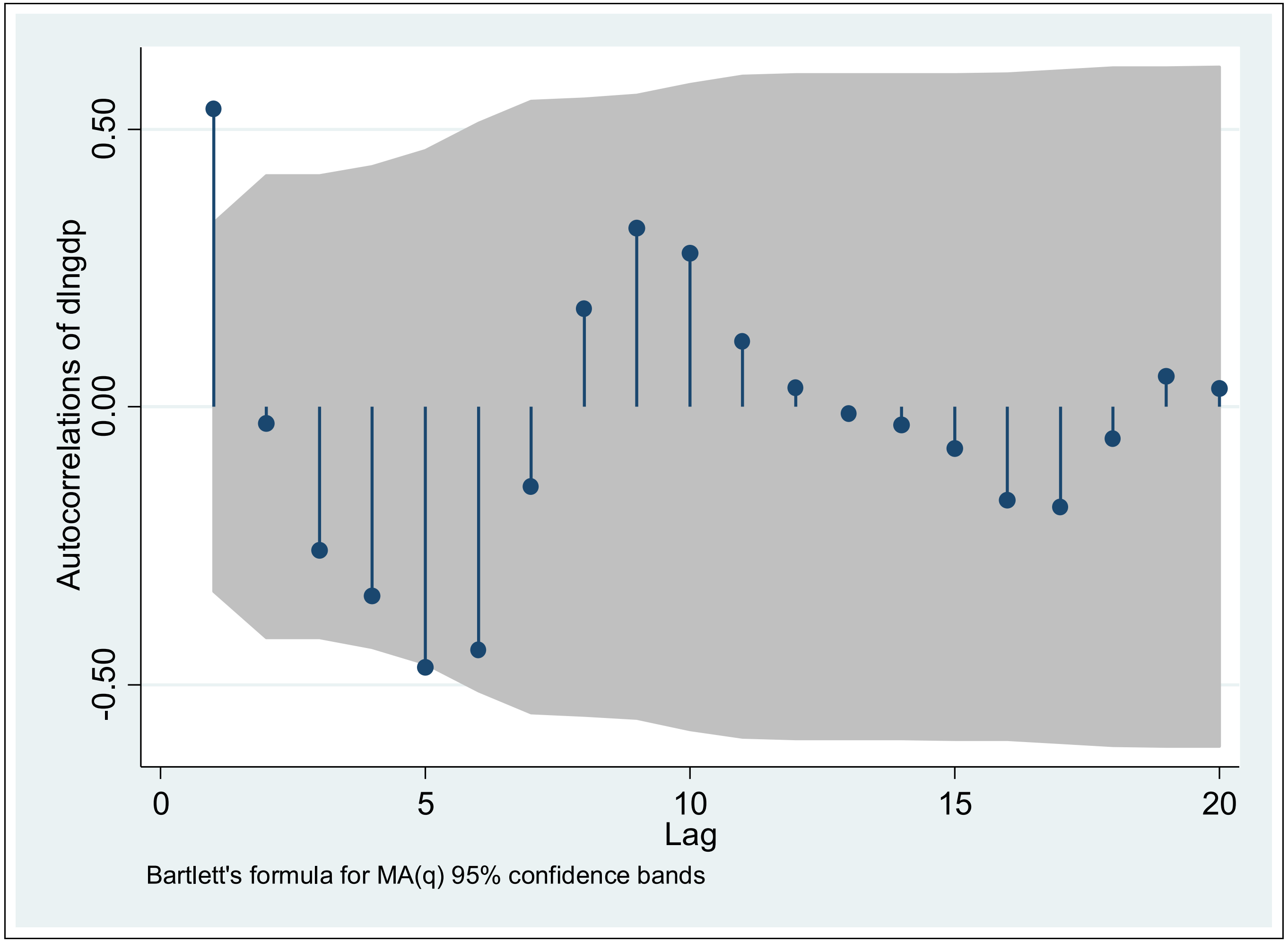

图 13.5　GDP 对数差分的自相关图

阴影部分表示置信区域(confidence band)。

如果某阶自相关系数落在置信区域之外，则说明该阶自相关系数显著地不为 0；反之，则不显著。

一阶与五阶自相关系数显著不为 0，而其他阶的自相关系数不显著。

13.2 一阶自回归

如果从一个客户的角度仅关心某变量(比如股价)的未来值，则可以用该变量的过去值来预测其未来值。

这种模型被称为**单变量时间序列**(univariate time series)。

由于时间序列一般存在自相关，故最简单的预测方法为，使用过去值 y_{t-1} 来预测当前值 y_t，即**一阶自回归**(First Order Autoregression，简记 AR(1))：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \quad (t = 2, \cdots, T) \qquad (13.6)$$

其中，扰动项 ε_t 为白噪声，故无自相关，即 $\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0, \quad \forall\, t \neq s$。

假设自回归系数$\left|\beta_1\right| < 1$，则$\left\{y_t\right\}$为渐近独立的平稳过程。

由于解释变量y_{t-1}依赖于$\left\{\varepsilon_{t-1}, \cdots, \varepsilon_1\right\}$，而扰动项$\varepsilon_t$与$\left\{\varepsilon_{t-1}, \cdots, \varepsilon_1\right\}$不相关，故$y_{t-1}$为前定变量(predetermined)，与$\varepsilon_t$不相关，因此 OLS 一致。

但使用 OLS 只能用观测值$t = 2,\cdots,T$进行回归，将损失一个样本容量。

为了提高估计效率，可使用最大似然估计(须假设扰动项服从正态分布)。

继续上节的例子，以 OLS 估计$\Delta \ln y_t$的一阶自回归模型。

为了演示目的，仅使用 2013 年以前的数据进行回归，然后预测 2013 年的 GDP。

```
. reg dlny l.dlny if year<2013,r
```

由于假设扰动项ε_t无自相关，故使用异方差稳健的标准误即可，而不必使用异方差自相关稳健的 HAC 标准误。

```
Linear regression                               Number of obs     =         33
                                                F(1, 31)          =      12.99
                                                Prob > F          =     0.0011
                                                R-squared         =     0.2879
                                                Root MSE          =     .02147

------------------------------------------------------------------------------
             |               Robust
        dlny | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        dlny |
         L1. |   .5362727   .1487888     3.60   0.001     .2328159    .8397295
             |
       _cons |   .0437698   .0144049     3.04   0.005     .0143908    .0731488
------------------------------------------------------------------------------
```

可得如下回归方程(常数项与斜率均在 1%水平上显著):

$$\widehat{\Delta \ln y_t} = 0.0438698 + 0.5362727\,\Delta \ln y_{t-1} \qquad (13.7)$$

计算此回归方程的拟合值，即$\widehat{\Delta \ln y_t}$，并记为 dlny1。

```
. predict dlny1
(option xb assumed; fitted values)
(2 missing values generated)

. list dlny1 if year==2013
```

	dlny1
36.	.083309

因此，$\widehat{\Delta \ln y_{2013}} = 0.083309$。

由于$\widehat{\ln y_{2013}} = \ln y_{2012} + \widehat{\Delta \ln y_{2013}}$，故 2013 年 GDP 的预测值为 $\widehat{y_{2013}} = \exp\left(\ln y_{2012} + \widehat{\Delta \ln y_{2013}}\right)$。

在 Stata 中，可使用“`x[n]`”表示变量x的第n个观测值，故可计算如下：

```
. dis exp(lny[35]+dlny1[36])
95985.114
```

其中,“`lny[35]`”表示变量 *lny* 的第 35 个观测值(即 2012 年)，

“`dlny1[36]`”表示变量 *dlny1* 的第 36 个观测值(即 2013 年)，因为样本容量为 36。

根据 AR(1)模型,2013 年 GDP 的预测值为 95,985.114 亿元(1978 年不变价格)。

对比 2013 年的实际 GDP，并计算预测误差，即$(y_{2013} - \widehat{y_{2013}})$：

```
. dis y[36]
95089.211

. dis y[36]-exp(lny[35]+dlny1[36])
-895.90347
```

预测误差为-895.90347 亿元，高估了 895.90347 亿元。

13.3 高阶自回归

在 AR(1)模型中，我们假设扰动项为无自相关，故可使用 OLS 进行一致的估计。

如果模型为 AR(2)，但却被误设为 AR(1)，则意味着二阶滞后项 $\beta_2 y_{t-2}$ 被纳入扰动项：

$$y_t = \beta_0 + \beta_1 y_{t-1} + (\beta_2 y_{t-2} + \varepsilon_t) \qquad (13.8)$$

由于扰动项为$(\beta_2 y_{t-2} + \varepsilon_t)$，故扰动项与解释变量$y_{t-1}$相关，因为$\mathrm{Cov}(y_{t-1}, y_{t-2}) \neq 0$。

OLS 不再一致，须引入$\beta_2 y_{t-2}$才能得到一致估计。

从预测的角度，更高阶的滞后项也可能包含有用的信息。

考虑p阶自回归模型，记为 AR(p)：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t \tag{13.9}$$

其中，扰动项ε_t为白噪声(无自相关)，故 OLS 为一致估计。

通常我们不知道滞后期p。如何估计$\hat{p}$在实践中有着重要意义。

方法之一：设一个最大滞后期$p_{\max}$，然后令$\hat{p} = p_{\max}$进行估计，并对最后一个滞后期系数的显著性进行t检验。如果接受该系数为 0，则令$\hat{p} = p_{\max} - 1$，重新进行估计，再对(新的)最后一个滞后期的系数进行t检验，如果显著，则停止；否则，令$\hat{p} = p_{\max} - 2$；以此类推。

此准则被称为**由大到小的序贯t规则**(general-to-specific sequential t rule)。

方法之二：使用信息准则，选择 $\hat{p}$ 使得 AIC 或 BIC 最小化，分别记为 $\hat{p}_{\text{AIC}}$ 与 $\hat{p}_{\text{BIC}}$。比如，

$$\min_{p} \text{AIC} \equiv \ln\left(\frac{\text{SSR}}{T}\right) + \frac{2}{T}(p+1) \qquad (13.10)$$

其中，SSR 为残差平方和。

$\hat{p}_{\text{BIC}}$ 是真实滞后阶数 p 的一致估计，但 $\hat{p}_{\text{AIC}}$ 在大样本中可能高估 p。

这两种信息准则在小样本中难分优劣，都很常用。

在实践中，可以结合以上两种方法来确定 $\hat{p}$。

如果二者不一致，为了保守起见(即尽量避免遗漏变量偏差)，可取二者滞后阶数的大者。

还可检验模型的残差是否存在自相关(比如，使用 Q 检验)；如果残差存在自相关，则须扩大滞后阶数。

回到上节 GDP 对数差分的例子。

首先，使用信息准则来确定滞后阶数 p，即从 AR(1)开始，分别计算 AIC 与 BIC，然后逐次增加滞后阶数，直至 AIC 与 BIC 二者出现最小值并开始上升。

```
. quietly reg dlny l.dlny if year<2013,r

. estat ic
```

```
Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	33	75.35938	80.96115	2	-157.9223	-154.9293

Note: BIC uses N = number of observations. See **[R] IC note**.

AR(1)的 AIC 为-157.9223，而 BIC 为-154.9293。

下面估计 AR(2)模型，并计算其信息准则。

```
. reg dlny l(1/2).dlny if year<2013,r
```

其中，“`l(1/2).dlny`” 表示变量 *dlny* 的 1-2 阶滞后。

```
Linear regression                                   Number of obs     =         32
                                                    F(2, 29)          =      17.51
                                                    Prob > F          =     0.0000
                                                    R-squared         =     0.4234
                                                    Root MSE          =     .01979

------------------------------------------------------------------------------
             |               Robust
        dlny | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        dlny |
         L1. |   .7711595   .1304462     5.91   0.000     .5043671    1.037952
         L2. |  -.4487175   .1530057    -2.93   0.007    -.7616494   -.1357857
             |
       _cons |   .0641134   .0128498     4.99   0.000     .0378326    .0903943
------------------------------------------------------------------------------
```

dlny 的二阶滞后 `L2.dlny` 依然在 1%水平上显著，故根据序贯 *t* 规则，滞后阶数 *p* 应至少大于或等于 2。

```
. estat ic
```

```
Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	32	72.88943	81.69936	3	-157.3987	-153.0015

Note: BIC uses N = number of observations. See **[R] IC note**.

AR(2)的 AIC 为-157.3987，而 BIC 为-153.0015；均比 AR(1)略有上升。

根据信息准则，应选择 $p=1$，即 AR(1)模型(尽管 AR(1)仅比 AR(2)有微弱优势)。

进一步估计 AR(3)模型。

```
. reg dlny l(1/3).dlny if year<2013,r
```

Linear regression						
				Number of obs = 31		
				F(3, 27) = 12.82		
				Prob > F = 0.0000		
				R-squared = 0.4459		
				Root MSE = .01907		

dlny	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
dlny						
L1.	.7557034	.1261359	5.99	0.000	.4968939	1.014513
L2.	-.4943798	.1785691	-2.77	0.010	-.8607733	-.1279864
L3.	.0204783	.1711994	0.12	0.906	-.3307938	.3717504
_cons	.0692315	.0154137	4.49	0.000	.0376052	.1008579

dlny 的三阶滞后 `L3.dlny` 很不显著(p值高达 0.906)，故根据序贯t规则，应选择$p=2$，即 AR(2)模型。

综合以上结果，为避免遗漏变量偏差，应按照序贯t规则选择AR(2)模型。

进一步，使用命令 `corrgram` 对残差进行 Q 检验也表明，AR(1)的残差存在自相关，而 AR(2)的残差无自相关(参见习题)。

下面使用 AR(2)模型预测 GDP，并与 AR(1)模型的预测效果进行对比。

```
. quietly reg dlny l(1/2).dlny if year<2013,r

. predict dlny2
(option xb assumed; fitted values)
(3 missing values generated)
```

```
. dis exp(lny[35]+dlny2[36])
95769.998

. dis y[36]-exp(lny[35]+dlny2[36])
-680.78688
```

对于 2013 年的 GDP，AR(2)模型的预测误差为-680.78688 亿元，高估了 680.78688 亿元

AR(1)模型则高估了 895.90347 亿元。

AR(2)的预测效果明显优于 AR(1)，因为二阶滞后仍包含不少有用信息。

13.4 自回归分布滞后模型

在自回归 AR(p)模型中，为了提高预测力或解释力，也可引入其他解释变量，构成**自回归分布滞后模型**(Autoregressive Distributed Lag Model，简记 ADL(p, q)或 ARDL(p, q))：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \gamma_1 x_{t-1} + \cdots + \gamma_q x_{t-q} + \varepsilon_t \quad (13.11)$$

其中，p为被解释变量y的自回归阶数，而q为解释变量x的滞后阶数。

假定扰动项ε_t为白噪声，则 OLS 为一致估计。

对于滞后阶数(p, q)的选择，可使用信息准则(AIC 或 BIC)，或进行序贯检验，即使用t或F检验来检验最后一阶系数的显著性。

在 ADL 模型(13.11)中，也可引入更多的解释变量；比如，变量z的r阶滞后$(z_{t-1}, \cdots, z_{t-r})$。

对于 ADL 模型(13.11)，解释变量x_{t-1}对于y_t的边际效应为γ_1，但这并非长期效应。

由于$\{y_t\}$与$\{x_t\}$为平稳序列，故均值不随时间而变，记其均值分别为y^*与x^*。将方程(13.11)两边同时求期望可得

$$y^* = \beta_0 + \beta_1 y^* + \cdots + \beta_p y^* + \gamma_1 x^* + \cdots + \gamma_q x^* \qquad (13.12)$$

将上式移项整理可得

$$(1-\beta_1-\cdots-\beta_p)y^* = \beta_0+(\gamma_1+\cdots+\gamma_q)x^* \tag{13.13}$$

故x^*增加一单位对y^*的边际效应为

$$\frac{dy^*}{dx^*}=\frac{\gamma_1+\cdots+\gamma_q}{1-\beta_1-\cdots-\beta_p} \tag{13.14}$$

这就是解释变量x永久性增加一单位对y的长期效应，也称为**长期乘数**(long-run multiplier)。

例 作为 ADL 的实例，Chen (2015)研究了中原王朝与北方游牧民族边界维度(*border*)的决定因素。

时间序列数据集 border.dta 以每十年作为观测单位(时间变量为 *decade*)，从公元前 221 年秦朝建立至 1911 年清朝灭亡，共有 213 个观测值。

主要解释变量：中原王朝早于游牧政权建立的年数(*diff*)，以及中国北方在十年中发生旱灾的年数比例(*drought*)。其他控制变量包括：中原王朝的绝对年龄(*age*)，游牧对手的数目(*rival*)，中原是否在长城的有效保护之下(*wall*)，以及中国是否统一(*unified*)。

信息准则与序贯检验均支持估计以下 ADL(2, 1)模型：

$$
\begin{aligned}
border_t = {} & \beta_0 + \beta_1 border_{t-1} + \beta_2 border_{t-2} + \gamma_1 drought_{t-1} \\
& + \gamma_2 diff_t + \gamma_3 age_t + \gamma_4 rival_t + \gamma_5 wall_t + \gamma_6 unified_t + \varepsilon_t
\end{aligned}
\tag{13.15}
$$

其中，变量 *diff*, *age*, *rival*, *wall* 与 *unified* 被认为只有当期作用，而气候变量 *drought* 则存在滞后效应。

```
. use border.dta, clear

. tsset decade

. reg border l(1/2).border l.drought diff age
rival wall unified,r
```

Linear regression

Number of obs	= 211
F(8, 202)	= 2040.71
Prob > F	= 0.0000
R-squared	= 0.9854
Root MSE	= 1.0889

border	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
border						
L1.	1.518284	.133108	11.41	0.000	1.255825	1.780744
L2.	-.5586965	.1278243	-4.37	0.000	-.8107376	-.3066555
drought						
L1.	-.6333046	.3009855	-2.10	0.037	-1.226781	-.0398281
diff	-.0069699	.0028159	-2.48	0.014	-.0125222	-.0014175
age	-.0264399	.0123658	-2.14	0.034	-.0508224	-.0020573
rival	.34148	.1455019	2.35	0.020	.0545827	.6283772
wall	.7339998	.2203202	3.33	0.001	.2995774	1.168422
unified	.4078538	.2230656	1.83	0.069	-.0319819	.8476894
_cons	.8189595	.472603	1.73	0.085	-.1129085	1.750827

被解释变量的两阶滞后(*L1.border* 与 *L2.border*)均在 1%水平上显著。

变量 *L1.drought* 在 5%水平上显著为负，说明气候越干旱，则游牧民族越会为了生存而进攻中原王朝，从而将游牧边界推向南方。

变量 *diff* 也在 5%水平上显著为负，说明中原王朝越早于游牧政权建立(根据王朝周期假说，中原相对而言更弱)，则北方边界纬度越低。

其他控制变量也至少在 10%水平上显著，而且系数符号均与理论预期一致。

从上表还可计算气候冲击对游牧边界的长期效应。

假如 *drought* 永久性增加一单位，即从 0 增加到 1(从年年无灾到年年旱灾)，则中国北方边界纬度将变化$[\gamma_1/(1-\beta_1-\beta_2)]$度。

代入上表中相应的系数估计值可得：

```
. dis -.6333046/(1-1.518284+.5586965)
-15.671008
```

根据此模型，气候冲击对游牧边界的长期效应为 15.67 度，这是一个很大的效应(从北京到海口的纬度差距约为 20 度)。

13.5　误差修正模型

ADL 是一种动态模型。

从经济理论而言，相关的变量之间可能存在长期的均衡关系，而变量的短期变动则是向着这个长期均衡关系的部分调整。

误差修正模型(Error Correction Model，简记 ECM)正是这一思想在计量经济学的体现。

考虑最简单的 ADL(1, 1)模型：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_1 x_{t-1} + \varepsilon_t \quad (13.16)$$

其中，$|\beta_1| < 1$，故为平稳过程。

假设经济理论认为(y, x)之间存在以下长期均衡关系：

$$y = \phi + \theta x \qquad (13.17)$$

其中，ϕ与θ为待定参数。

对方程(13.16)两边求期望，并令$y^* = \mathrm{E}(y_t) = \mathrm{E}(y_{t-1})$，$x^* = \mathrm{E}(x_t) = \mathrm{E}(x_{t-1})$，可得

$$y^* = \beta_0 + \beta_1 y^* + \gamma_1 x^* \qquad (13.18)$$

移项整理可得

$$y^* = \frac{\beta_0}{(1-\beta_1)} + \frac{\gamma_1}{(1-\beta_1)} x^* \qquad (13.19)$$

由此可知，$\phi = \dfrac{\beta_0}{1-\beta_1}$，$\theta = \dfrac{\gamma_1}{1-\beta_1}$。

$\theta = \dfrac{\gamma_1}{1-\beta_1}$为长期乘数，衡量当$x$永久性变化一单位时，将导致$y$的永久性变化幅度。

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

在方程(13.16)两边同时减去y_{t-1}可得

$$\Delta y_t = \beta_0 - (1-\beta_1)y_{t-1} + \gamma_1 x_{t-1} + \varepsilon_t \qquad (13.20)$$

代入$\beta_0 = (1-\beta_1)\phi$以及$\gamma_1 = (1-\beta_1)\theta$，可得

$$\Delta y_t = (1-\beta_1)\phi - (1-\beta_1)y_{t-1} + (1-\beta_1)\theta x_{t-1} + \varepsilon_t \quad (13.21)$$

经整理则有

$$\Delta y_t = \underbrace{(\beta_1 - 1)(y_{t-1} - \phi - \theta x_{t-1})}_{\text{error correction}} + \varepsilon_t \qquad (13.22)$$

这就是误差修正的形式。

$(y_{t-1} - \phi - \theta x_{t-1})$衡量上一期对均衡条件“$y = \phi + \theta x$”的偏离(误差)

$(\beta_1 - 1)(y_{t-1} - \phi - \theta x_{t-1})$为根据上期的误差所作的反向修正，称为**误差修正项**(error correction term)。

比如，$(y_{t-1}-\phi-\theta x_{t-1})>0$，即 y_{t-1} 高于其均衡值，则 $(\beta_1-1)(y_{t-1}-\phi-\theta x_{t-1})<0$(因为$(\beta_1-1)<0$)，故平均而言(忽略扰动项)，$\Delta y_t<0$，使得下一期能更靠近均衡条件。

一般来说，ADL 模型都可以转换成 ECM 模型。

误差修正模型的优点是，经济含义十分明确，而且可以分别考察长期效应(长期均衡关系)与短期效应(误差修正效应)。

13.6 移动平均与 ARMA 模型

另一类时间序列模型为**移动平均过程**(Moving Average Process，简记 MA)。记一阶移动平均过程为 MA(1)：

$$y_t=\mu+\varepsilon_t+\theta\varepsilon_{t-1} \tag{13.23}$$

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

由于y_t可被看成是白噪声的移动平均，故得名。

考虑q阶移动平均过程，记为 MA(q)：

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \qquad (13.24)$$

对于 MA(q)模型，假设$\{\varepsilon_t\}$为 iid 且服从正态分布，则可进行 MLE 估计。

将 AR(p)与 MA(q)结合起来，可得到 ARMA(p, q)模型：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \qquad (13.25)$$

其中，$\{\varepsilon_t\}$为白噪声。

对于 ARMA(p, q)模型，也可进行 MLE 估计。

对于 MA(q)，如果$q \to \infty$，则得到无穷阶的移动平均过程，记为 MA(∞)：

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \quad (13.26)$$

其中，$\theta_0 = 1$(标准化为 1)。

从哲学的角度看，MA(∞)相当于将y_t的决定因素追溯到无穷远的过去。

无穷多个随机变量之和，能收敛到某个随机变量吗？

一个常用的充分条件是，序列$\left\{\theta_j\right\}_{j=0}^{\infty}$为**绝对值可加总**(Absolutely Summable，简记 AS)，即$\sum_{j=0}^{\infty}\left|\theta_j\right|<\infty$(有限)。在 AS 的条件下，则 MA(∞)有定义。

虽然样本容量T通常有限，且一般无法追溯到无穷远的过去，但 MA(∞)在理论上仍然有着重要意义，因为 AR(p)与 ARMA(p, q)都可写为 MA(∞)的形式。

13.7　脉冲响应函数

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

证明：反复使用迭代法可得

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \\
&= \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= (\beta_0 + \beta_0\beta_1) + \beta_1^2 y_{t-2} + \beta_1\varepsilon_{t-1} + \varepsilon_t \\
&= (\beta_0 + \beta_0\beta_1) + \beta_1^2(\beta_0 + \beta_1 y_{t-3} + \varepsilon_{t-2}) + \beta_1\varepsilon_{t-1} + \varepsilon_t \\
&= \beta_0(1 + \beta_1 + \beta_1^2) + \beta_1^3 y_{t-3} + \beta_1^2\varepsilon_{t-2} + \beta_1\varepsilon_{t-1} + \varepsilon_t \\
&= \cdots \\
&= \beta_0(1 + \beta_1 + \beta_1^2 + \cdots) + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_1^2\varepsilon_{t-2} + \beta_1^3\varepsilon_{t-3} + \cdots \\
&= \frac{\beta_0}{1-\beta_1} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_1^2\varepsilon_{t-2} + \beta_1^3\varepsilon_{t-3} + \cdots \qquad (13.27)
\end{aligned}
$$

其中，无穷等比级数之和$(1 + \beta_1 + \beta_1^2 + \cdots)$等于$1/(1-\beta_1)$。

上式即为 MA(∞)的形式。

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

从 AR(1)的MA(∞)表达式可以看出

$$IRF(j) \equiv \frac{\partial y_{t+j}}{\partial \varepsilon_t} = \beta_1^j \qquad (13.28)$$

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

动态乘子与绝对时间 t 无关，是时间间隔 j 的函数。

将 $\left(\partial y_{t+j} / \partial \varepsilon_t\right)$ 视为 j 的函数，则称为**脉冲响应函数**(Impulse Response Function，简记 IRF)。

它刻画的是 y_{t+j} 对 ε_t 的 1 单位脉冲(impulse)的响应(response)。

将 $\left(j,\ \partial y_{t+j} / \partial \varepsilon_t\right)$ 画图，即可得到对 IRF 的直观认识，称为**脉冲响应图**。

类似地，AR(p)也是MA(∞)。

更一般地，ARMA(p, q)也是 MA(∞)。

例 继续以数据集 gdp_china.dta 为例，考察 GDP 对数差分(dlny)的自回归模型。

为了计算脉冲响应函数(IRF)，可将 AR(p)视为一维的向量自回归(Vector Autogression，简记 VAR)，并使用以下命令：

```
. varbasic x y z,lags(numlist) irf
```

其中，“varbasic”为估计 VAR 模型的便捷命令，而“x y z”为 VAR 模型所包含的变量(在此例中只有一个变量)。

选择项“lags(numlist)”表示滞后阶数，默认为“lags(1 2)”或“lags(1/2)”，即滞后二阶。

选择项“irf”表示画脉冲响应图。

首先，估计 *dlny* 的 AR(1)模型(为保持与上文的一致，不包括 2013 年的观测值)，并将其脉冲响应函数画图(结果参见图 13.6)。

```
. use gdp_china.dta,clear

. gen lny=log(y)

. gen dlny=d.lny

. varbasic dlny if year<2013,lags(1) irf
```

```
Vector autoregression

Sample: 1980 thru 2012                          Number of obs     =         33
Log likelihood =   80.96115                     AIC               =  -4.785524
FPE            =   .0004889                     HQIC              =  -4.755008
Det(Sigma_ml)  =   .0004331                     SBIC              =  -4.694827

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
dlny                  2     .021471   0.2879   13.34017   0.0003
----------------------------------------------------------------


------------------------------------------------------------------------------
        dlny | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
dlny         |
        dlny |
         L1. |   .5362727   .1468267     3.65   0.000     .2484976    .8240478
             |
       _cons |   .0437698   .0143197     3.06   0.002     .0157037     .071836
------------------------------------------------------------------------------
```

使用命令 varbasic 的估计系数与命令“reg dlngdp l.dlngdp”完全相同，但不提供异方差稳健标准误的选择项。

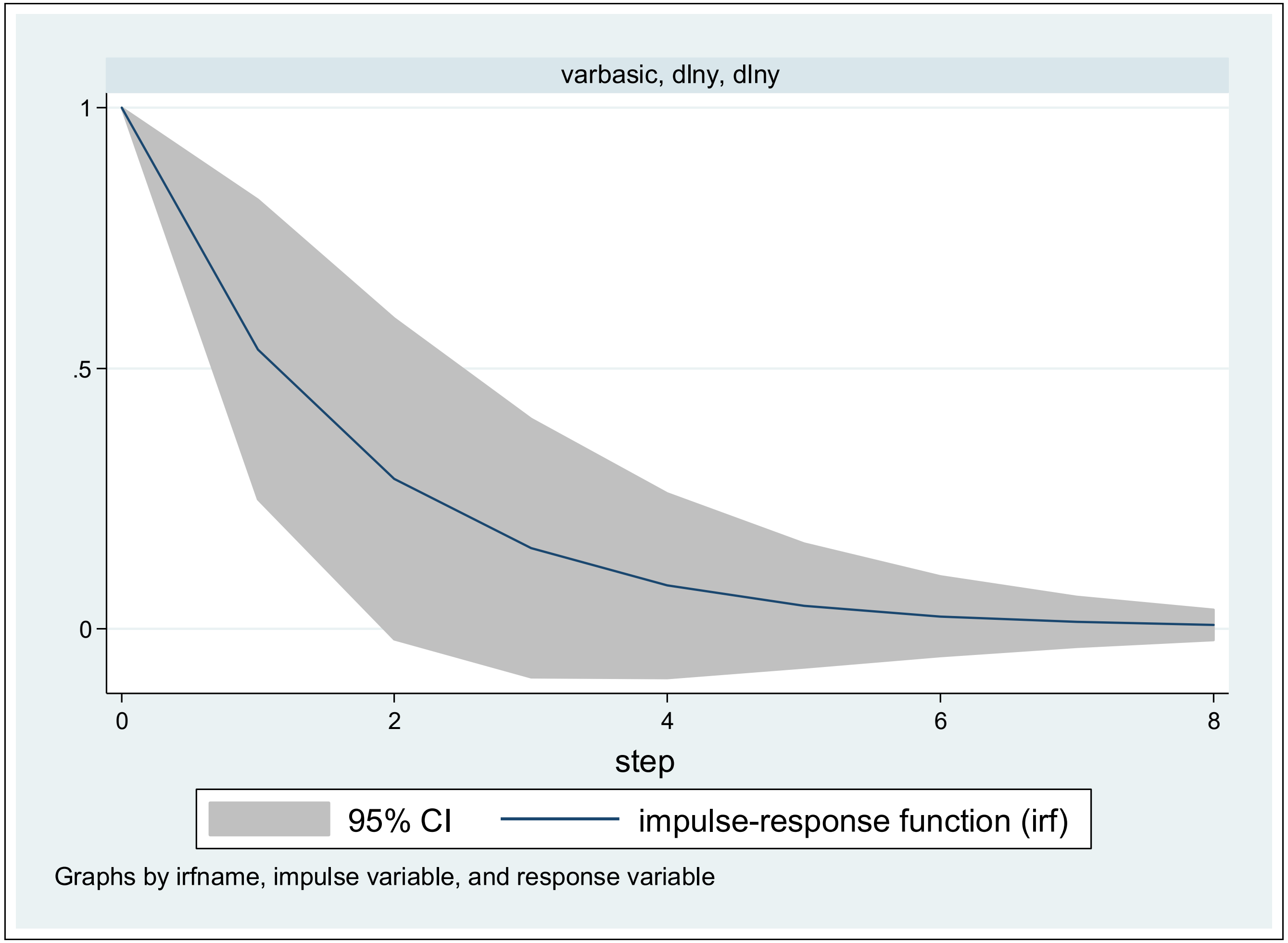

图 13.6　AR(1)模型的脉冲响应函数

AR(1)模型的脉冲响应函数呈指数衰减，从当期的一单位冲击逐渐衰减为 0，与方程(13.28)的 IRF 表达式一致。

灰色区域为 IRF 的 95%置信区间。

其次，估计 *dlny* 的 AR(2)模型，并将其 IRF 画图(结果参见图 13.7)。

```
. varbasic dlny if year<2013,irf
```

```
Vector autoregression

Sample: 1981 thru 2012                          Number of obs     =         32
Log likelihood =   81.69936                     AIC               =   -4.91871
FPE            =   .0004282                     HQIC              =  -4.873162
Det(Sigma_ml)  =   .0003548                     SBIC              =  -4.781298

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
dlny                  3     .019785   0.4234   23.49854   0.0000
----------------------------------------------------------------


------------------------------------------------------------------------------
        dlny | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
dlny         |
        dlny |
         L1. |   .7711595   .1593746     4.84   0.000      .458791    1.083528
         L2. |  -.4487175    .157661    -2.85   0.004    -.7577275   -.1397076
             |
       _cons |   .0641134   .0148348     4.32   0.000     .0350377    .0931892
------------------------------------------------------------------------------
```

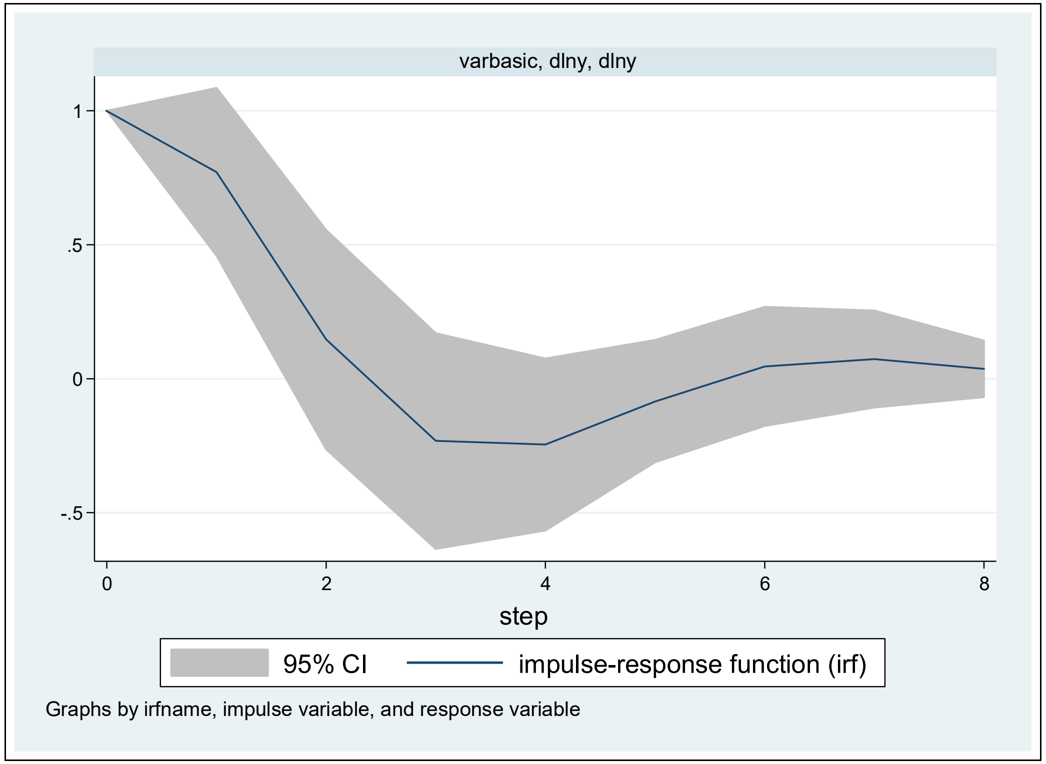

图 13.7　AR(2)模型的脉冲响应函数

AR(2)模型的脉冲响应函数不再单调递减，更具动态特征，即先下降，变为负数后再反弹上升，然后又下降并趋于 0。

13.8　向量自回归过程

我们常常同时关心几个经济变量的预测，比如 GDP 增长率与失业率。

一种方法是用单变量时间序列的方法对每个变量分别作预测。

另一种方法则是将这些变量放在一起，作为一个系统来预测，以使得预测相互自洽(mutually consistent)，称为**多变量时间序列**(multivariate time series)。

由 Sims (1980)所提倡的**向量自回归**(Vector Autoregression，简记 VAR)正是这样一种方法。

假设有两个时间序列$\{y_{1t}, y_{2t}\}$，分别作为两个回归方程的被解释变量；而解释变量为这两个变量的p阶滞后值，构成一个二元的 VAR(p)系统：

$$\begin{cases} y_{1t} = \beta_{10} + \beta_{11} y_{1,t-1} + \cdots + \beta_{1p} y_{1,t-p} + \gamma_{11} y_{2,t-1} + \cdots + \gamma_{1p} y_{2,t-p} + \varepsilon_{1t} \\ y_{2t} = \beta_{20} + \beta_{21} y_{1,t-1} + \cdots + \beta_{2p} y_{1,t-p} + \gamma_{21} y_{2,t-1} + \cdots + \gamma_{2p} y_{2,t-p} + \varepsilon_{2t} \end{cases} \tag{13.29}$$

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

$$\mathrm{Cov}(\varepsilon_{1t}, \varepsilon_{2s}) = \begin{cases} \sigma_{12} & \text{若} t = s \\ 0 & \text{其他} \end{cases} \tag{13.30}$$

表达式(13.29)的两个方程，解释变量完全相同。可将这两个方程更简洁地写在一起：

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix} y_{1,t-1} + \cdots + \begin{pmatrix} \beta_{1p} \\ \beta_{2p} \end{pmatrix} y_{1,t-p} \\ + \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \end{pmatrix} y_{2,t-1} + \cdots + \begin{pmatrix} \gamma_{1p} \\ \gamma_{2p} \end{pmatrix} y_{2,t-p} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{13.31}$$

将同期变量合成列向量，并把相应的系数合并为矩阵可得

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

(13.32)

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

$$\boldsymbol{y}_t = \underbrace{\begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix}}_{\boldsymbol{\Gamma}_0} + \underbrace{\begin{pmatrix} \beta_{11} & \gamma_{11} \\ \beta_{21} & \gamma_{21} \end{pmatrix}}_{\boldsymbol{\Gamma}_1} \boldsymbol{y}_{t-1} + \cdots + \underbrace{\begin{pmatrix} \beta_{1p} & \gamma_{1p} \\ \beta_{2p} & \gamma_{2p} \end{pmatrix}}_{\boldsymbol{\Gamma}_p} \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t \quad (13.33)$$

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

$$\boldsymbol{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t \qquad (13.34)$$

这个形式与 AR(p)很相似，故名“VAR(p)”。

其中，$\{\boldsymbol{\varepsilon}_t\}$是一维白噪声过程的推广，称为**向量白噪声过程**(vector white noise process)或**新息过程**(innovation process)。

由于 VAR(p) 系统中的解释变量$\{\boldsymbol{y}_{t-1}, \cdots, \boldsymbol{y}_{t-p}\}$依赖于$\{\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \cdots\}$，而$\boldsymbol{\varepsilon}_t$与$\{\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \cdots\}$不相关，故可视所有解释变量为前定变量(predetermined)，与当期扰动项$\boldsymbol{\varepsilon}_t$不相关，故可用 OLS 对每个方程分别进行一致估计。

在进行 VAR 建模时，需要确定变量的滞后阶数，以及 VAR 系统中包含几个变量。

滞后阶数的选择

方法之一是使用信息准则，比如 AIC 或 BIC。

方法之二是检验最后一阶系数的显著性(类似于由大到小的序贯规则)。

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

方法之三是检验 VAR 模型的残差是否为白噪声，即是否存在自相关。

如果真实模型为 VAR(p)，但被错误地设置为 VAR(p–1)，则解释变量的最后一阶滞后$\boldsymbol{y}_{t-p}$被纳入扰动项$\boldsymbol{\varepsilon}_t$，导致扰动项出现自相关。由于$\{\boldsymbol{y}_t\}$的相关性，包含$\boldsymbol{y}_{t-p}$的扰动项$\boldsymbol{\varepsilon}_t$将与解释变量$\{\boldsymbol{y}_{t-1}, \cdots, \boldsymbol{y}_{t-p+1}\}$相关，导致 OLS 估计不一致。

为此，需要检验 VAR 模型的残差是否存在自相关。

如果存在自相关，则意味着应加入更高阶的滞后变量。

VAR 变量个数的选择

VAR 系统中包含的变量个数越多，则需要估计的系数越多。

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

待估系数过多将使得有效样本容量过小(损失自由度)，增大估计误差，降低预测精度。

VAR 模型通常仅包含为数不多的几个变量。

在设定 VAR 模型时，主要应根据经济理论来确定哪些变量应在 VAR 模型中。

比如，经济理论告诉我们，通货膨胀率、失业率、短期利息率互相关联，可以构成一个三变量的 VAR 模型。

另外，也可以在 VAR 系统中引入其他外生解释变量，比如 $\left\{w_{1t}, w_{2t}, \cdots, w_{Kt}\right\}$，与扰动项不相关。

13.9 VAR 的脉冲响应函数

由于VAR模型包含许多参数，而这些参数的经济意义很难解释，故通常将注意力集中于脉冲响应函数。

考虑如下n元 VAR(p)系统：

$$\boldsymbol{y}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t \qquad (13.35)$$

其中，$\boldsymbol{y}_t$包含n个变量。

正如 AR(p)可写为 MA(∞)，此 VAR(p)系统也可写成**向量移动平均过程**(Vector Moving Average Process) VMA(∞)的形式：

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t + \boldsymbol{\psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots = \boldsymbol{\alpha} + \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \boldsymbol{\varepsilon}_{t-i} \tag{13.36}$$

其中，$\boldsymbol{\psi}_0 \equiv \mathbf{I}_n$，而$\boldsymbol{\psi}_j$为$n$维方阵。

$\boldsymbol{\varepsilon}_{t-1}$对$\boldsymbol{y}_t$的“边际效应”为$\boldsymbol{\psi}_1$。更严格地，可以证明，

$$\frac{\partial \boldsymbol{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'} = \boldsymbol{\psi}_s \tag{13.37}$$

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

假设$n=2$，则

$$\frac{\partial \boldsymbol{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_t'}=\boldsymbol{\psi}_s=\begin{pmatrix} \dfrac{\partial y_{1,\,t+s}}{\partial \varepsilon_{1t}} & \dfrac{\partial y_{1,\,t+s}}{\partial \varepsilon_{2t}} \\ \dfrac{\partial y_{2,\,t+s}}{\partial \varepsilon_{1t}} & \dfrac{\partial y_{2,\,t+s}}{\partial \varepsilon_{2t}} \end{pmatrix} \qquad (13.38)$$

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

它表示的是，当第j个变量在第t期的扰动项ε_{jt}增加 1 单位时(而其他变量与其他期的扰动项均不变)，对第i个变量在第$(t+s)$期的取值$y_{i,\,t+s}$的影响。

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

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

此假定只有当扰动项不存在“同期相关”(contemporaneous correlation)时才成立。

但现实中，同期相关是普遍存在的。

可从扰动项$\boldsymbol{\varepsilon}_t$中分离出相互正交的部分，记为$\boldsymbol{v}_t$。

新扰动项$\mathbf{v}_t$的各分量正交(不相关)，且方差均被标准化为 1(故变化一单位，就是变化一个标准差)。

然后计算当$\boldsymbol{v}_t$中的某分量变动时，对各变量在不同时期的影响，称为**正交化脉冲响应函数**(Orthogonalized Impulse Response Function，简记 OIRF)。

但 OIRF 依然有缺点。

首先，正交化冲击(orthogonalized shocks)$\boldsymbol{v}_t$的经济含义不易解释($\boldsymbol{v}_t$为$\boldsymbol{\varepsilon}_t$中各分量的线性组合)。

其次，OIRF 依赖于变量的次序(order of variables)；如果改变变量次序，可能得到很不相同的结果。

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

在实践中，可借助借助格兰杰因果检验来确定两个变量之间的排序。

在难以确定变量次序的情况下，可进行稳健性检验，即对于不同的变量排序，分别画其正交化脉冲响应图，然后进行比较。

13.10 格兰杰因果检验

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

格兰杰 [Granger (1969)] 提出了**格兰杰因果检验** (Granger causality test)。

首先，原因必然发生于结果之前。

其次，原因包含有关结果的独特信息，对结果具有解释力或预测力。

如果x是y的因，但y不是x的因，则x的过去值可以帮助预测y的未来值，而y的过去值却不能帮助预测x的未来值。

考虑以下 ADL(p, p)模型：

$$y_t = \gamma + \sum_{m=1}^{p} \alpha_m y_{t-m} + \sum_{m=1}^{p} \beta_m x_{t-m} + \varepsilon_t \qquad (13.39)$$

其中，滞后阶数p可根据“信息准则”或“由大到小的序贯规则”来确定。

估计此模型后，可检验原假设“$H_0: \beta_1 = \cdots = \beta_p = 0$”，即$x$的过去值对预测$y$的未来值有无帮助。

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

将以上回归模型中x与y的位置互换，则可以检验y是否为x的格兰杰因。

在实际操作中，常将(x, y)构成一个二元 VAR 系统，然后在 VAR 的框架使用 Stata 命令 `vargranger` 进行格兰杰因果检验。

格兰杰因果关系并非真正意义上的因果关系。它充其量只是一种动态相关关系，表明一个变量是否对另一变量有“预测能力”(predictability)。

在某种意义上，它顶多是因果关系的必要条件，而且格兰杰因果关系也可能由第三个变量所引起。

另外，格兰杰因果检验仅适用于平稳序列，或者有协整关系的单位根过程。

对于不存在协整关系的单位根变量，则只能先差分，得到平稳序列后再进行格兰杰因果检验。

13.11 VAR 的 Stata 命令及实例

与 VAR 相关的 Stata 命令包括(假设变量为x, y, z)

```
. varsoc x y z, maxlag(#)
```

此命令用来计算不同滞后期的信息准则，其中“soc”表示 selection-order criteria

选择项“maxlag(#)”表示最大滞后期，默认值为 4。

在未使用选择项“maxlag(#)”的情况下，如果恰好选择最优滞后 4 期，则应通过选择项“maxlag(#)增加最大滞后期数。

```
. varbasic x y z,lags(numlist) irf
```

这是估计 VAR 模型的便捷命令。

选择项“lags(numlist)”表示滞后阶数，默认为“lags(1 2)”或“lags(1/2)”，即滞后二阶。

选择项“irf”表示画(未正交化)脉冲响应图，默认为“oirf”(画正交化脉冲响应图)。

估计 VAR 的正式命令为

```
. var x y z, lags(numlist) exog(w1 w2)
```

其中，选择项“lags(numlist)”表示滞后阶数，默认为“lags(1/2)”，即滞后二阶。

如果要滞后三阶，可使用选择项“lags(1/3)”。

选择项“exog(w1 w2)”表示在 VAR 模型中引入外生变量 *w1* 与 *w2*。

```
. varlmar,mlag(#)
```

估计 VAR 后，对残差是否存在自相关进行 *LM* 检验。

选择项“`mlag(#)`”指定所检验自相关的最大滞后阶数(maximum order of autocorrelation)，默认为“`mlag(2)`”。

```
. varstable,graph
```

估计 VAR 后，通过特征值检验该 VAR 系统是否为平稳过程。

如果所有特征值都在单位圆内部，则为平稳过程(参见第 14 章)。选择项“`graph`”表示画出特征值的几何分布图。

```
. varwle
```

估计 VAR 后，对每个方程以及所有方程的各阶系数的联合显著性进行沃尔德检验，其中“`wle`”表示 Wald lag-exclusion statistics。

```
. vargranger
```

估计 VAR 后，进行格兰杰因果检验。

```
. irf create irfname, set(filename) step(#) replace order(varlist)
```

估计 VAR 后，将有关脉冲响应的结果存为“irfname”(可自行命名)。

选择项“set(filename)”表示建立脉冲文件“filename”，使之成为当前的脉冲文件(make filename active)，并将脉冲结果“irfname”存入此脉冲文件“filename”(若未使用选择项“set(filename)”指定脉冲文件，则将脉冲响应结果存入当前的脉冲文件)。

选择项“step(#)”表示考察截止#期的脉冲响应函数，默认为“step(8)”。

选择项“replace”表示替代已有的同名脉冲响应结果 irfname (如果有)

一个脉冲文件“filename”可存储多个脉冲响应结果“irfname”。

选择项“order(varlist)”指定变量排序，默认使用估计 VAR 时的变量排序计算正交化 IRF。

. irf graph irf,impulse(varname) response(varname)

画脉冲响应图(未正交化)。

选择项“impulse(varname)”用于指定脉冲变量

选择项“response(varname)”用来指定反应变量；默认画出所有变量的脉冲响应图。

. irf graph oirf,impulse(varname) response(varname)

画正交化的脉冲响应图，选择项的含义同上。

如果将以上命令中的“irf graph”改为“irf table”，则将相应信息列表而非画图。

```
. fcast compute prefix,step(#)
```

估计 VAR 后，计算被解释变量未来#期的预测值，并把预测值赋予被解释变量加上前缀“prefix”(自行确定)的变量名。

```
. fcast graph varlist,observed
```

运行命令“fcast compute”后，将变量“varlist”的预测值画图，其中选择项“observed”表示与实际观测值相比较。

以数据集 macro_swatson.dta 为例，进行 VAR 估计。

该数据集包含美国 1960 年第 2 季至 2002 年第 1 季的以下宏观经济季度变量：*inf* 为通货膨胀率，*dinf* 为通货膨胀率的一阶差分，*unem* 为失业率，*quarter* 为季度(时间变量)。

由于通胀率 *inf* 可能不平稳，故考虑其一阶差分 *dinf* 与失业率 *unem* 所构成的二元 VAR 系统。

首先，打开数据集后，看一下二者的时间趋势(结果参见图 13.8)。

```
. use macro_swatson.dta,clear
. tsline dinf unem,lp(solid dash)
```

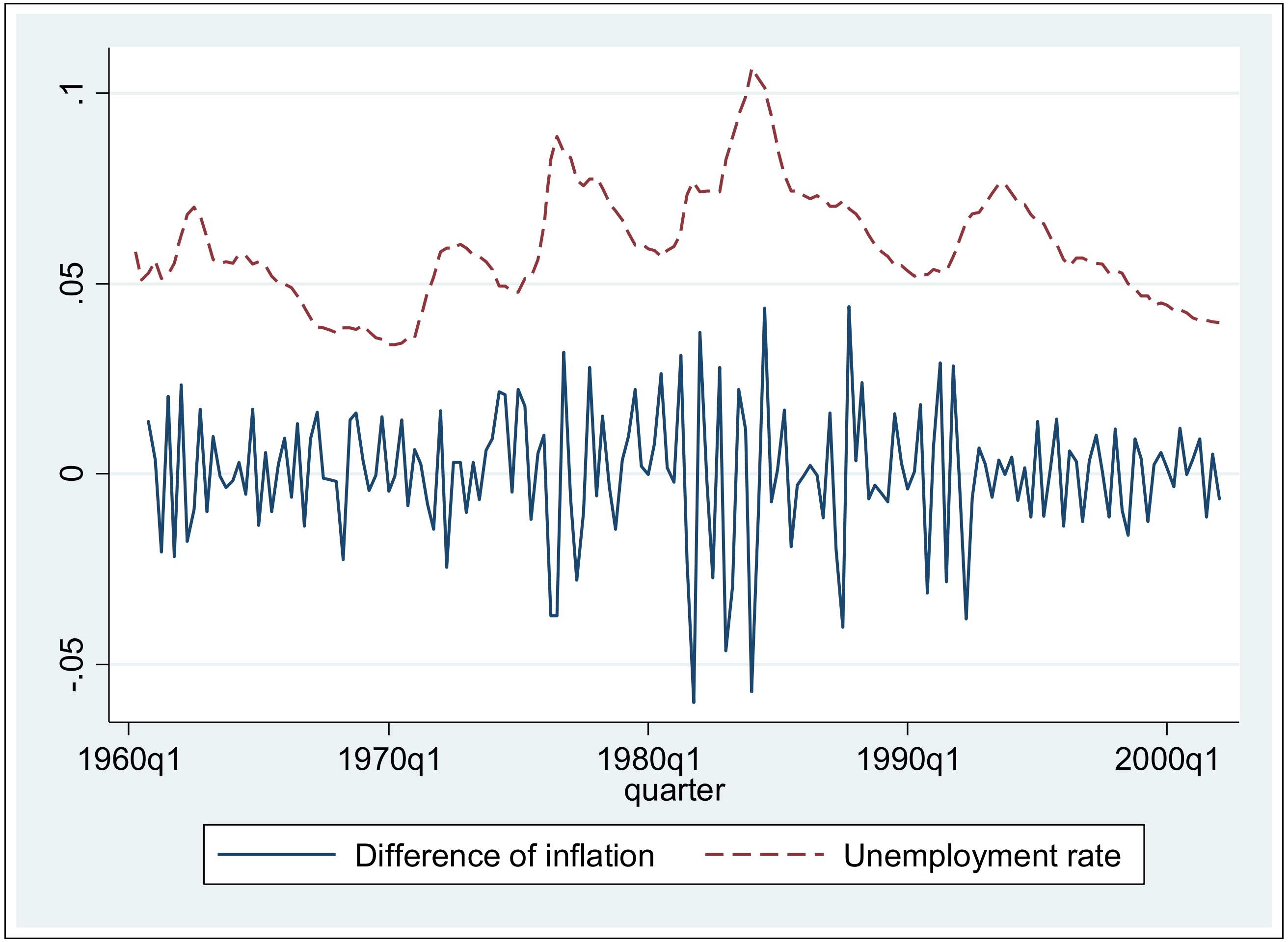

图 13.8　通胀差分与失业率的时间趋势图

其次，根据信息准则，估计此 VAR 系统的阶数。

```
. varsoc dinf unem
```

```
Lag-order selection criteria

   Sample: 1961q4 thru 2002q1                              Number of obs = 162
```

Lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	881.581				6.6e-08	-10.859	-10.8435	-10.8209
1	1141.4	519.63	4	0.000	2.8e-09	-14.0172	-13.9708	-13.9029
2	1213.86	144.93*	4	0.000	1.2e-09*	-14.8625*	-14.7851*	-14.6719*
3	1217.32	6.9224	4	0.140	1.2e-09	-14.8558	-14.7475	-14.589
4	1219.45	4.2537	4	0.373	1.2e-09	-14.8327	-14.6934	-14.4896

```
   * optimal lag
   Endogenous: dinf unem
    Exogenous: _cons
```

当 $p=2$ 时(打星号“*”者)，AIC 与 BIC 信息准则最小化。

下面，估计二阶向量自回归模型：

```
. var dinf unem,lags(1/2)
```

Vector autoregression

Sample: 1961q2 - 2002q1		No. of obs	= 164
Log likelihood =	1224.456	AIC	= -14.81044
FPE =	1.27e-09	HQIC	= -14.73371
Det(Sigma_ml) =	1.12e-09	SBIC	= -14.62143

Equation	Parms	RMSE	R-sq	chi2	P>chi2
dinf	5	.013815	0.3621	93.11047	0.0000
unem	5	.002508	0.9731	5926.554	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dinf						
dinf						
L1.	-.4709436	.0676118	-6.97	0.000	-.6034603	-.338427
L2.	-.401031	.0654944	-6.12	0.000	-.5293976	-.2726644
unem						
L1.	-2.24205	.3371897	-6.65	0.000	-2.90293	-1.58117
L2.	2.03417	.3356694	6.06	0.000	1.37627	2.69207
_cons	.0123764	.0045265	2.73	0.006	.0035047	.0212482
unem						
dinf						
L1.	.0305223	.0122746	2.49	0.013	.0064645	.05458
L2.	-.0113172	.0118902	-0.95	0.341	-.0346216	.0119871
unem						
L1.	1.638031	.0612152	26.76	0.000	1.518051	1.75801
L2.	-.6725807	.0609392	-11.04	0.000	-.7920194	-.5531421
_cons	.0020139	.0008218	2.45	0.014	.0004032	.0036245

大多数的系数均很显著。下面检验各阶系数的联合显著性。

```
. varwle
```

Equation: dinf

lag	chi2	df	Prob > chi2
1	72.72164	2	0.000
2	72.9294	2	0.000

Equation: unem

lag	chi2	df	Prob > chi2
1	741.8916	2	0.000
2	123.1287	2	0.000

Equation: All

lag	chi2	df	Prob > chi2
1	796.2418	4	0.000
2	187.9991	4	0.000

无论是单一方程(`dinf` 与 `unem`),还是两个方程作为整体(`All`),各阶系数均高度显著。

下面检验残差是否为白噪声，即残差是否存在自相关。

```
. varlmar
```

Lagrange-multiplier test

lag	chi2	df	Prob > chi2
1	7.3130	4	0.12024
2	5.0645	4	0.28074

H0: no autocorrelation at lag order

可以接受残差“无自相关”的原假设。

进一步检验估计出来的 VAR 系统是否为平稳过程，并画图(结果参见图 13.9)。

```
. varstable,graph
```

```
Eigenvalue stability condition
```

Eigenvalue	Modulus
.82182	.82182
.7970223	.797022
-.2258777 + .6292178*i*	.668533
-.2258777 - .6292178*i*	.668533

```
All the eigenvalues lie inside the unit circle.
VAR satisfies stability condition.
```

所有特征值均在单位圆之内，故此 VAR 模型满足平稳性条件。

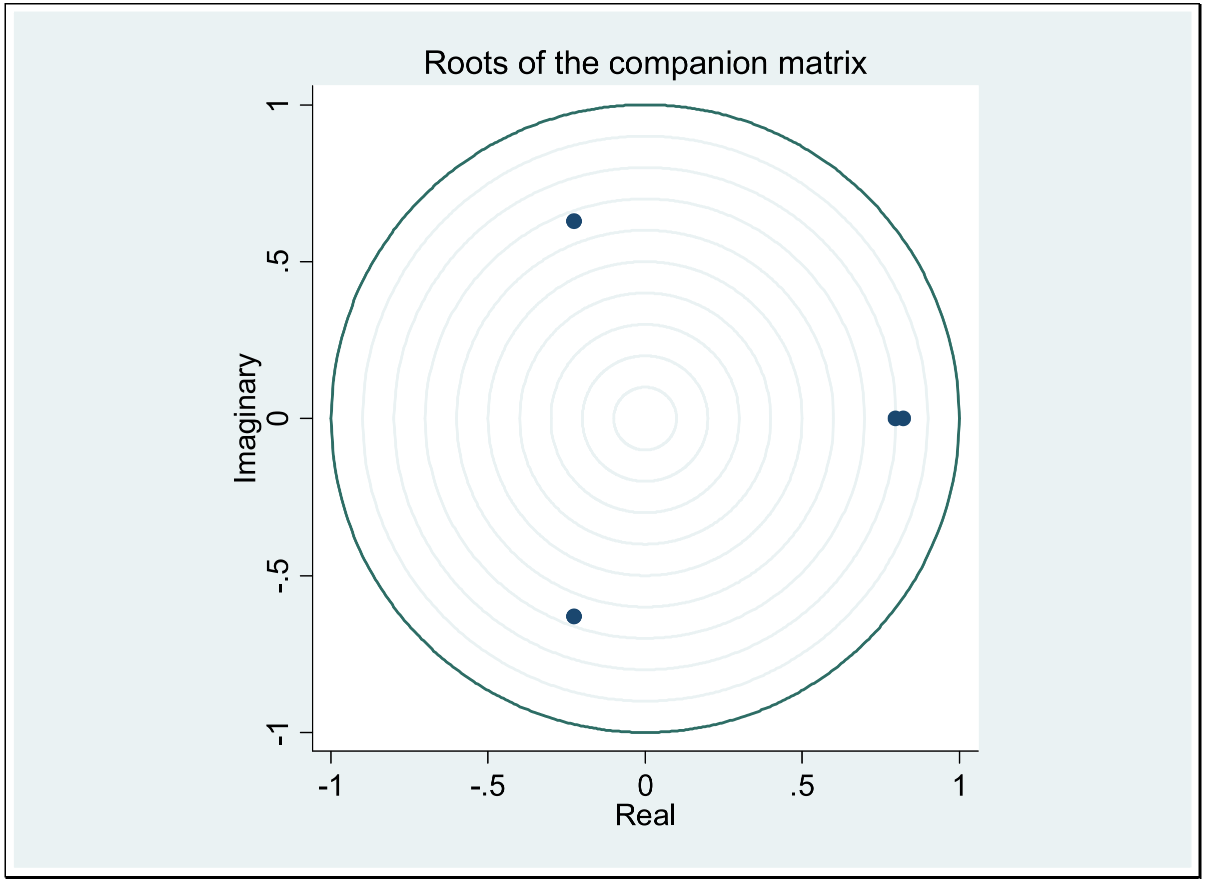

图 13.9　VAR 系统稳定性的判别图

所有特征值均在单位圆之内，故此 VAR 系统是稳定的。

下面考察变量 *dinf* 与 *unem* 之间的格兰杰因果关系。

```
. vargranger
```

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
dinf	unem	48.123	2	0.000
dinf	ALL	48.123	2	0.000
unem	dinf	9.009	2	0.011
unem	ALL	9.009	2	0.011

无论以 *dinf* 还是 *unem* 为被解释变量，其 p 值均远小于 0.05。

二者互为格兰杰原因，故格兰杰因果检验无法提供变量排序的信息。

下面考察正交化脉冲响应函数，将脉冲文件命名为“macro”，并将脉冲结果命令为“iu”(表示变量排序为 *dinf*，*unem*)。

```
. irf create iu, set(macro) replace
(file macro.irf created)
(file macro.irf now active)
(file macro.irf updated)
```

此命令建立了脉冲文件“macro.irf”，并将脉冲结果 iu 存入此脉冲文件。选择项“replace”表示，若已有同名脉冲结果，可覆盖之。

根据此脉冲结果，画正交化的脉冲响应图(结果参见图 13.10)。

```
. irf graph oirf,yline(0)
```

选择项“yline(0)”表示在纵轴 y=0 处画一条水平线。

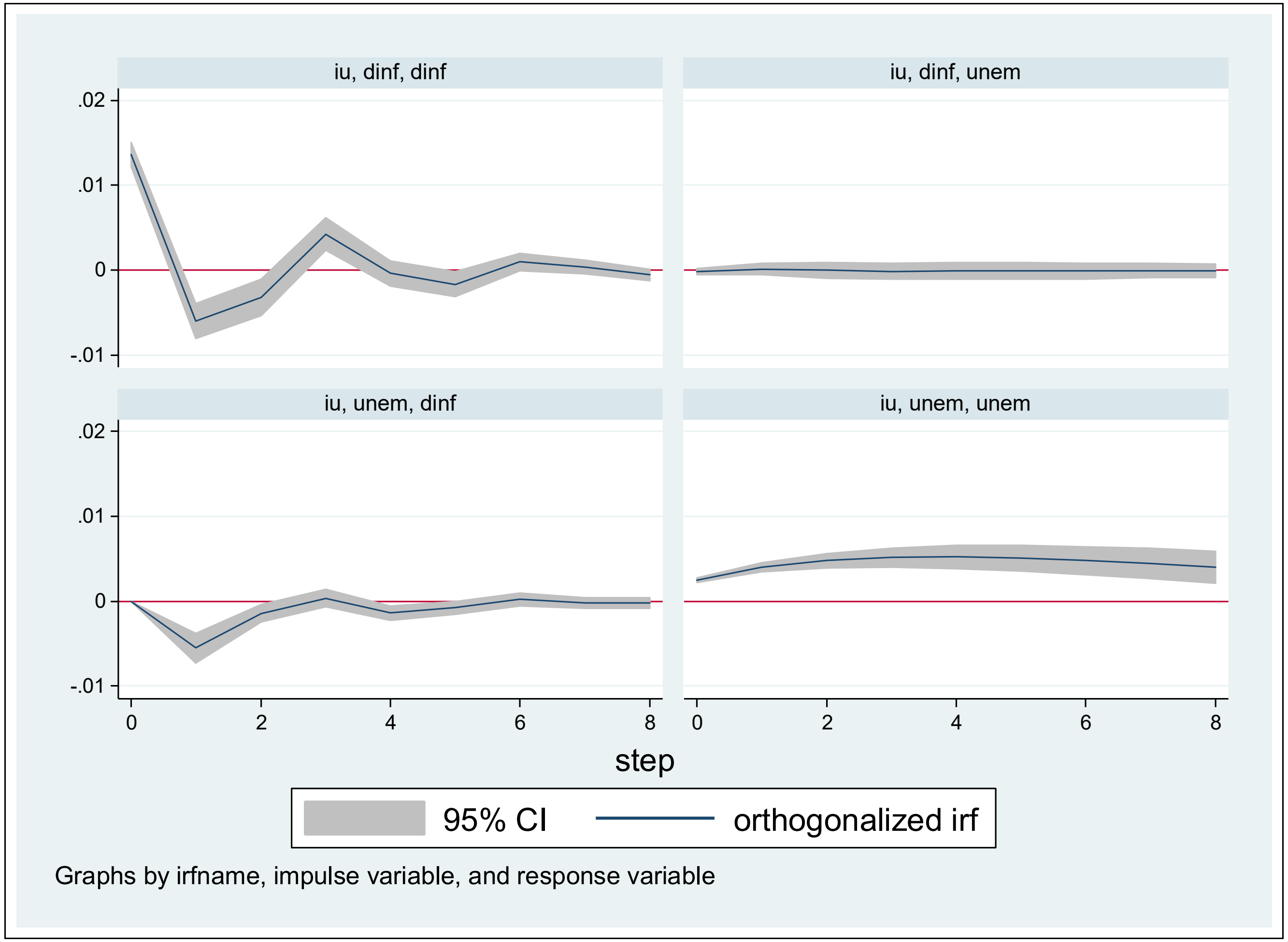

图 13.10　正交化的脉冲响应图

图 13.10 的最后一行为“`Graphs by irfname,impluse variable, and response variable`”，这表明四个小图的标题命名顺序为“脉冲名称、冲击变量、响应变量”。

比如，左下小图的标题为“`iu,unem,dinf`”，表明此图为根据脉冲结果 `iu`，冲击变量 *unem*，响应变量 *dinf* 所画的脉冲响应图。

它表明，失业率 *unem* 的一个标准差的正向冲击，将使未来一期的 *dinf* 下降(根据菲利普斯曲线，失业率上升可缓解通胀压力)，但未来二期的 *dinf* 即反弹，然后此影响逐渐消失归于 0。

下面变换变量次序，考察正交化脉冲响应函数的稳健性。

```
. irf create ui, order(unem dinf) replace
 (file macro.irf updated)
```

其中，选择项“`order(unem dinf)`”表示变量 *unem* 排在 *dinf* 之前。

此命令将脉冲结果记为 `ui`。

以上两个脉冲结果 `iu` 与 `ui`，都已存储在当前脉冲文件 `macro.irf` 中；故可直接画图，比较在两种变量排序下，*dinf* 对于 *unem* 冲击的脉冲响应(结果参见图 13.11)。

```
. irf graph oirf,i(unem) r(dinf) yline(0)
```

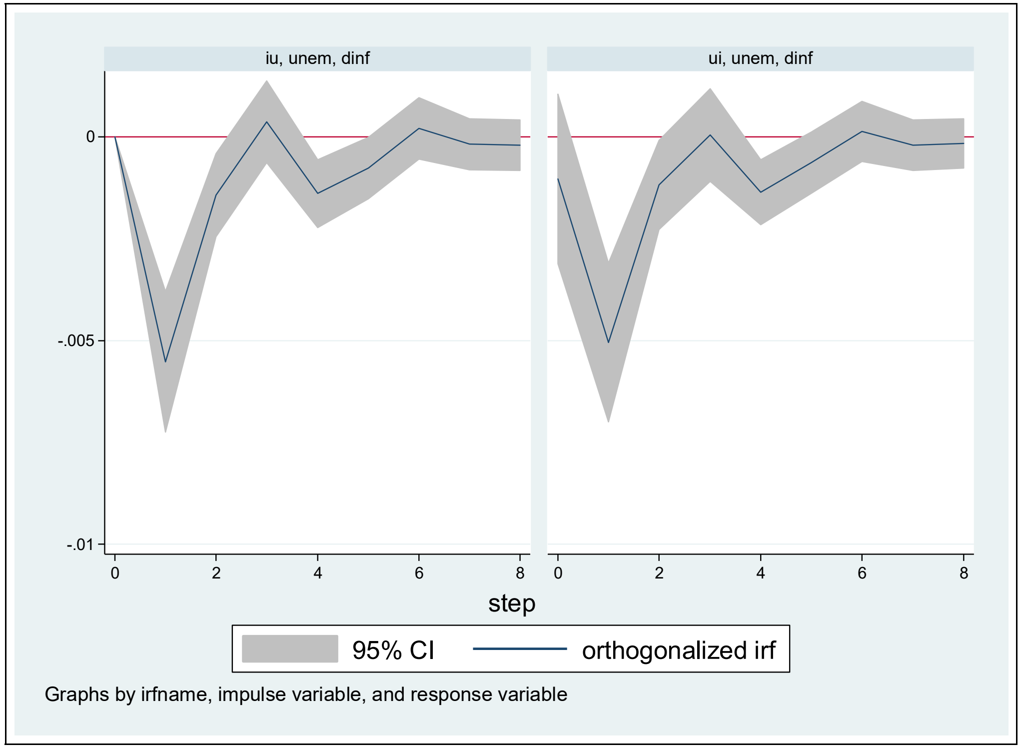

图 13.11　比较两种变量排序下的脉冲响应图(unem→ dinf)

在不同变量排序下，变量 *dinf* 对于 *unem* 冲击的脉冲响应差别不大，只是在反应幅度上略有不同。

比较在两种变量排序下，变量 *unem* 对于 *dinf* 冲击的脉冲响应(结果参见图 13.12)。

```
. irf graph oirf,i(dinf) r(unem) yline(0)
```

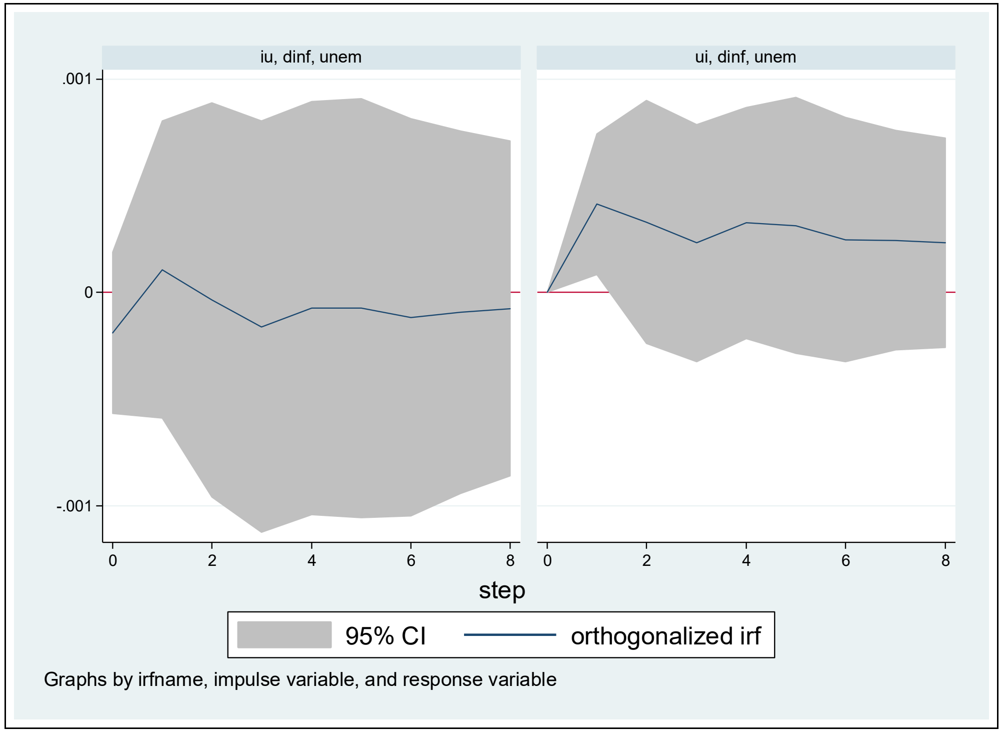

图 13.12　比较两种变量排序下的脉冲响应图(dinf→unem)

变量排序对于从 *dinf* 至 *unem* 的脉冲响应幅度有较大影响，但二者在变动方向上依然类似。

估计 VAR 模型后，可以用它进行预测。

假设仅用 1999 年以前的数据来估计 VAR 模型，然后预测 1999 年 1 季度-2002 年 1 季度的 13 个季度，并与实际观测值比较。

```
. varbasic dinf unem if
quarter<tq(1999q1),lags(1/2) nograph
```

其中，“`tq(1999q1)`”表示季度数据格式；选择项“`nograph`”表示不画脉冲响应图。

```
Vector autoregression

Sample:  1961q2 - 1998q4                        No. of obs      =       151
Log likelihood =  1118.534                      AIC             = -14.68256
FPE            =  1.44e-09                      HQIC            = -14.60139
Det(Sigma_ml)  =  1.26e-09                      SBIC            = -14.48274

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
dinf                  5     .014258   0.3673   87.65165   0.0000
unem                  5     .002585   0.9705   4971.522   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
dinf         |
        dinf |
         L1. |  -.4718995   .0702018    -6.72   0.000    -.6094925   -.3343064
         L2. |  -.4066509   .0682215    -5.96   0.000    -.5403626   -.2729393
             |
        unem |
         L1. |  -2.279801   .3502631    -6.51   0.000    -2.966304   -1.593298
         L2. |   2.050438   .3471149     5.91   0.000     1.370105    2.730771
             |
       _cons |   .0139506    .005028     2.77   0.006     .0040959    .0238053
-------------+----------------------------------------------------------------
unem         |
        dinf |
         L1. |   .0302607    .012728     2.38   0.017     .0053143    .0552072
         L2. |  -.0120022    .012369    -0.97   0.332    -.0362449    .0122405
             |
        unem |
         L1. |   1.636296   .0635048    25.77   0.000     1.511829    1.760763
         L2. |  -.6760615    .062934   -10.74   0.000    -.7994099   -.5527131
             |
       _cons |   .0023965   .0009116     2.63   0.009     .0006098    .0041832
------------------------------------------------------------------------------
```

子样本的样本容量减少为 151。

预测未来 13 个季度的变量取值，分别记为“*f_dinf*”与“*f_unem*”。

```
. fcast compute f_,step(13)
```

此命令将生成两个新变量，“*f_dinf*” 与 “*f_unem*”，分别为对 *dinf* 与 *unem* 的预测值。

这两个预测变量的标准误与置信区间也作为新变量出现在变量窗口。

将 *dinf* 与 *unem* 的预测值画图，并与实际观测值比较(结果参见图 13.13)。

```
. fcast graph f_dinf f_unem,observed lp(dash)
```

其中，选择项“observed”表示显示变量的实际观测值

选择项“lp(dash)”表示以虚线来画变量的预测值(以便区别于实际观测值)。

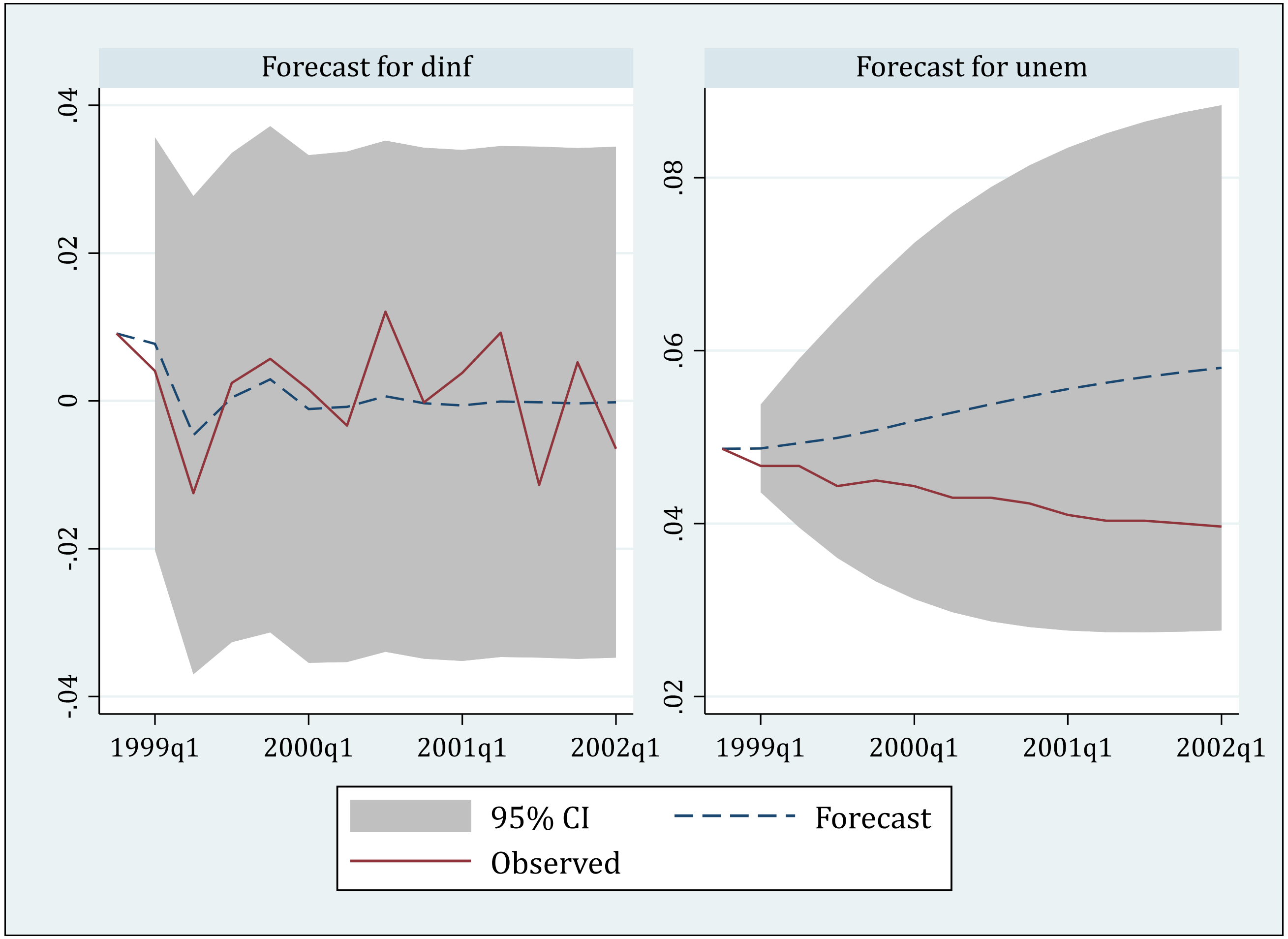

图 13.13　对未来 10 个季度的预测

对通胀率变动的预测准确度优于对失业率的预测(须注意二者的纵坐标单位与绝对位置都不同)，而二者的实际观测值均落在预测值 95%的置信区间内。

可以看出，预测的时期越长，则预测的精确度越低。

13.12　时间趋势项

时间序列(比如，宏观经济变量)常常包含某种时间趋势，比如 GDP 的指数增长趋势(参见图 13.1)，或 GDP 对数的线性增长趋势(参见图 13.2)。

假设时间序列$\left\{y_t\right\}$包含时间趋势，则不是平稳过程(期望随时间而变)。

一种处理方法为做差分(或者先取对数，再差分)，将其变为平稳过程(参见本章第 13.1 节)。

但有时我们想直接对原变量y_t建模，则一种常见做法为，在回归方程中引入“线性时间趋势项”(linear time trend)：

$$y_t = \alpha + \beta t + \varepsilon_t \qquad (13.40)$$

其中，$t = 1, 2, \cdots, T$为时间趋势项(T为样本容量)。

如果$\{y_t\}$存在指数增长趋势，则可对其对数建模：

$$\ln y_t = \alpha + \beta t + \varepsilon_t \qquad (13.41)$$

其中，系数β的经济含义为y_t的每期增长率，即

$$\beta = \frac{d \ln y}{dt} = \frac{dy/y}{dt} \qquad (13.42)$$

如果y_t的增长率并非常数，可以考虑加入时间趋势的平方项：

$$\ln y_t = \alpha + \beta t + \gamma t^2 + \varepsilon_t \qquad (13.43)$$

若忽略扰动项，则y_t的平均增长率为

$$\frac{dy/y}{dt} = \frac{d\ln y}{dt} = \beta + 2\gamma t \qquad (13.44)$$

在上式中，如果系数$\gamma > 0$，则y_t的增长率不断上升；反之，如果系数$\gamma < 0$，则y_t的增长率不断下降。

是否应引入时间趋势的平方项t^2，可通过检验“$H_0: \gamma = 0$”来判断。

但引入时间趋势的平方项，可能导致多重共线性。

假设样本容量为 100，可在 Stata 中计算时间趋势t及其平方项t^2的相关系数如下。

```
. clear
. set obs 100
. gen t=_n
. gen t2=t^2
. corr t t2
```

	t	t2
t	1.0000	
t2	0.9689	1.0000

t与t^2的相关系数接近 0.97，故存在严重的多重共线性。

最常见的做法是仅包含线性趋势项，以避免多重共线性。

如果$\{y_t\}$包含时间趋势，但被遗漏了，则可能导致遗漏变量偏差。考虑以下简单模型：

$$y_t = \alpha + \beta x_t + \varepsilon_t \qquad (13.45)$$

其中，变量y_t包含时间趋势γt(但被遗漏了)，故可将上式的扰动项写为

$$\varepsilon_t = \gamma t + u_t \qquad (13.46)$$

其中，u_t为不含时间趋势的扰动项。假设变量x_t包含时间趋势δt，并可写为

$$x_t = x_t^* + \delta t \qquad (13.47)$$

其中，x_t^*为x_t中不包含时间趋势的部分。将表达式(13.46)与(13.47)代入原模型(13.45)可得

$$y_t = \alpha + \beta(\underbrace{x_t^* + \delta t}_{x_t}) + (\underbrace{\gamma t + u_t}_{\varepsilon_t}) \qquad (13.48)$$

在上式中，解释变量x_t与扰动项ε_t相关，故 OLS 不一致。

由于宏观经济变量通常都有时间趋势，比如y_t与x_t都有时间趋势，故简单地将y_t对x_t进行回归将发现二者存在显著的关系，而事实上这只是因为二者都为共同的时间趋势所驱动。

这种现象是“伪回归”(spurious regression)的一种表现。

只要将遗漏的时间趋势加入回归方程(13.45)，即可消除此伪回归现象：

$$y_t = \alpha + \beta x_t + \gamma t + u_t \qquad (13.49)$$

其中，扰动项u_t不再包含时间趋势，不会与解释变量x_t相关，故 OLS 为一致估计。

继续以数据集 gdp_china.dta 为例。

考虑直接对 GDP 对数(lny)建模，即估计 *lny* 的 AR(2)模型，并加上时间趋势项。

```
. use gdp_china.dta,clear
. gen t=_n
. reg lny l(1/2).lny t if year<2013,r
```

```
Linear regression                               Number of obs     =         33
                                                F(3, 29)          =   38047.85
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9997
                                                Root MSE          =     .01627

------------------------------------------------------------------------------
             |               Robust
         lny | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         lny |
         L1. |   1.302829   .1150665    11.32   0.000     1.067491    1.538166
         L2. |  -.7357654   .0967217    -7.61   0.000    -.9335835   -.5379473
             |
           t |   .0411649   .0092175     4.47   0.000      .022313    .0600168
       _cons |   3.473641   .7552359     4.60   0.000     1.929011    5.018272
------------------------------------------------------------------------------
```

lny 的两阶滞后以及时间趋势项都高度显著。

预测 GDP 对数，并记为 *lny3*，然后计算 2013 年的预测误差。

```
. predict lny3
```

```
(option xb assumed; fitted values)
(2 missing values generated)

. dis exp(lny3[36])
95597.887

. dis y[36]-exp(lny3[36])
-508.67625
```

lny 的“AR(2)+时间趋势项”模型的预测误差为-508.67625 亿元，高估了 508.67625 亿元。

dlny 的 AR(2)模型的预测高估了 680.78688 亿元。

13.13 季节调整

1．季节效应

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

比如，考察中国的季度 GDP 数据。由于第一季度包含春节，故通常第一季度的 GDP 偏低。

如果直接以第二季度 GDP 除以第一季度 GDP 来计算环比增长率，则会高估第二季度的 GDP 增长率；同样道理，将第一季度 GDP 除以上年第四季度 GDP 则会低估第一季度的 GDP 增长率。

包含季节效应的时间序列不能直接计算环比增长率。

如果不进行季节调整，则只能计算同比增长率，即与去年同一季(月)相比。对于年度数据，则不需要进行季节调整。

可能导致季节效应的因素包括：

(1) 天气因素：比如，在冬季由于取暖而增加能源消耗。

(2) 行政因素：比如，学校开学与放假对交通的影响。

(3) 固定假日：比如，十一国庆节对旅游与交通的影响。

(4) 移动假日(moving holiday)：比如，春节期间，铁路运输量增加而 GDP 下降。

(5) 日历因素：比如，闰年与闰月的影响。

(6) 交易日效应：比如，五金店销售额在有五个周末的月份高于只有四个周末的月份。

所有这些季节因素共同构成一个时间序列的**季节要素**(seasonal component)。

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

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

2．季节调整的原理

季节调整通常通过估计**季节因子**(seasonal factor)来进行。

根据季节因子起作用的方式，季节因子主要分为两种，即“加法季节因子”(additive seasonal factor)与“乘法季节因子”(multiplicative seasonal factor)。

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

加法模型(additive model)的数学表达式如下：

$$Y_t = TC_t + S_t + I_t \qquad (13.50)$$

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

乘法季节因子则意味着对所有第 1 月(或第 1 季)都乘以相同的季节因子，以此类推。

乘法模型(multiplicative model)的数学表达式如下：

$$Y_t = TC_t \times S_t \times I_t \qquad (13.51)$$

使用乘法模型要求Y_t序列中不包含零或负数。对方程(13.51)两边同时取对数可得

$$\ln Y_t = \ln TC_t + \ln S_t + \ln I_t \qquad (13.52)$$

方程(13.52)在形式上与加法模型相同，故称为“对数加法模型”。

季节调整的目标就是将原序列Y_t分解为趋势循环要素、季节要素与不规则要素，然后去掉季节要素S_t，得到**季节调整序列** (seasonally adjusted series)，即趋势循环要素与不规则要素之和。

季节调整的具体方法有多种，使用不同方法，会得到不同的季节调整序列，带有一定的主观性；这是季节调整的局限性。

下面介绍最为简便的回归法。

3．回归法

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

以 airpassengers.dta 为例，该数据集包括 1949-1960 年国际航空旅客人数的月度数据(*airpassengers*)与时间变量(*time*)。

打开数据集后，首先看一下 *airpassengers* 的时间趋势图(结果参见图 13.14)。

```
. use airpassengers.dta,clear

. tsset time
```

```
Time variable: time, 1949m1 to 1960m12
        Delta: 1 month
```

```
. tsline airpassengers
```

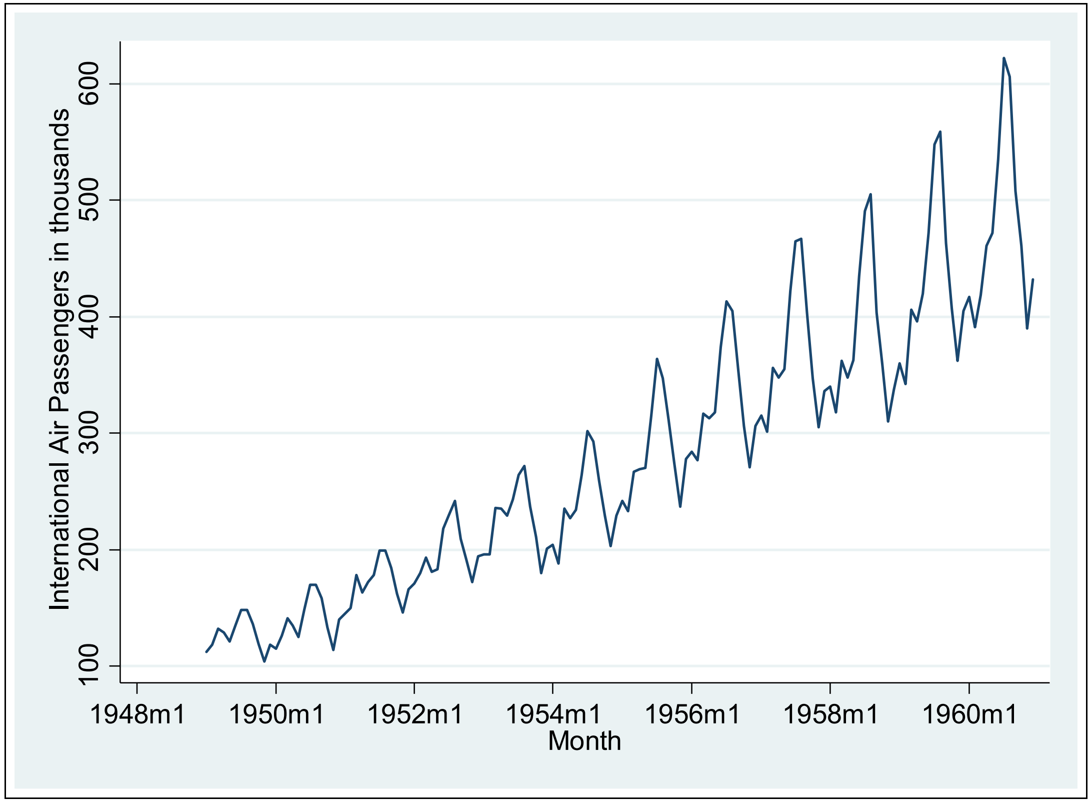

图 13.14　国际航空旅客人数的季节波动

国际航空旅客人数存在明显的季节波动，通常在夏季(七、八两月)达到高峰。

为了生成月度虚拟变量，首先从时间变量 *time* 中提取月度信息，记为变量 *month*。

```
. gen month=month((dofm(time)))
```

其中，`dofm(time)`根据将 *time* 转换为数值型，即从 1960 年 1 月 1 日至该月月初过了几天，而 `month((dofm(time)))`则进一步提取其中的月度信息。

此命令将生成变量 *month*，取值为 1, 2, ... ,12，对应于一年中的 12 个月。

其次，使用命令 `tab` 来生成季度虚拟变量。

```
. tab month,gen(m)
```

其中，选择项“gen(m)”表示，根据变量 *month* 的不同取值，生成相应的虚拟变量，记为 *m1*, *m2* , ... , *m12*，分别对应于 12 个月。

month	Freq.	Percent	Cum.
1	12	8.33	8.33
2	12	8.33	16.67
3	12	8.33	25.00
4	12	8.33	33.33
5	12	8.33	41.67
6	12	8.33	50.00
7	12	8.33	58.33
8	12	8.33	66.67
9	12	8.33	75.00
10	12	8.33	83.33
11	12	8.33	91.67
12	12	8.33	100.00
Total	144	100.00	

以 1 月份为参照值，把变量 *airpassengers* 对第 2-12 月的月度虚拟变量进行回归(也可通过在回归中放入 `i.m` 来实现)。

```
. reg airpassengers m2-m12
```

Source	SS	df	MS
Model	218382.243	11	19852.9312
Residual	1839661.92	132	13936.8327
Total	2058044.16	143	14391.9172

Number of obs	=	144
F(11, 132)	=	1.42
Prob > F	=	0.1690
R-squared	=	0.1061
Adj R-squared	=	0.0316
Root MSE	=	118.05

airpasseng~s	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
m2	-6.75	48.19549	-0.14	0.889	-102.0854	88.58545
m3	28.41667	48.19549	0.59	0.556	-66.91878	123.7521
m4	25.33333	48.19549	0.53	0.600	-70.00211	120.6688
m5	30.08333	48.19549	0.62	0.534	-65.25211	125.4188
m6	69.91667	48.19549	1.45	0.149	-25.41878	165.2521
m7	109.5833	48.19549	2.27	0.025	14.24789	204.9188
m8	109.3333	48.19549	2.27	0.025	13.99789	204.6688
m9	60.66667	48.19549	1.26	0.210	-34.66878	156.0021
m10	24.83333	48.19549	0.52	0.607	-70.50211	120.1688
m11	-8.916667	48.19549	-0.19	0.854	-104.2521	86.41878
m12	20.08333	48.19549	0.42	0.678	-75.25211	115.4188
_cons	241.75	34.07936	7.09	0.000	174.3377	309.1623

七、八两月的虚拟变量(*m7* 与 *m8*)均在 5%水平上显著为正；而其他月份的虚拟变量则不显著。

为了获得经季度调整的序列，下面使用命令 `predict` 来获得上述回归的残差项(记为 *air_sa*)。

```
. predict air_sa,r
```

但 OLS 残差项的平均值一定为 0，故需要把原序列的均值加回去，并记季节调整序列为 `airpassengers_sa`。

```
. sum airpassengers
```

Variable	Obs	Mean	Std. dev.	Min	Max
airpasseng~s	144	280.2986	119.9663	104	622

```
. gen airpassengers_sa = air_sa+r(mean)
```

将回归法的季节调整序列与原序列画图，结果参见图 13.15。

```
. tsline airpassengers_sa
airpassengers,lp(dash)
```

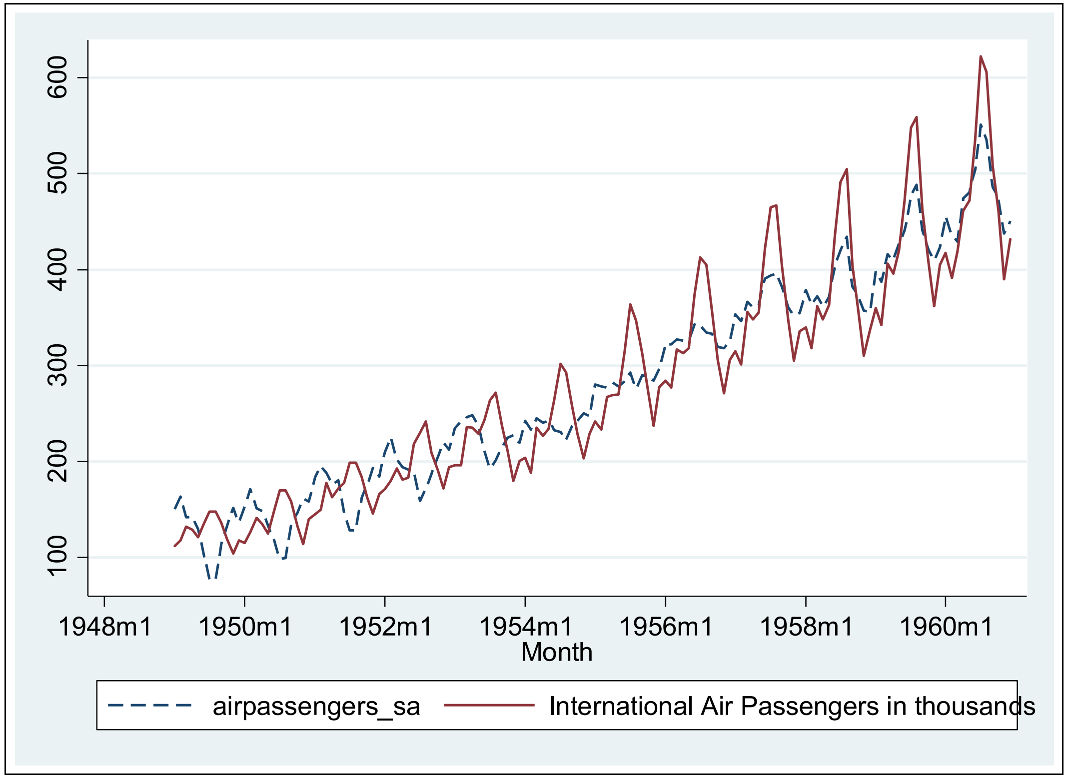

图 13.15　季节调整序列与原序列

经回归法进行季节调整后，序列变得更为光滑，季节性波动的特征有所减弱。

13.14 日期数据的导入

如果时间序列为年度数据，将时间数据导入 Stata 并无特殊之处，只要使用命令“`tsset year`”将变量 *year* 设为时间变量即可(假设时间变量为 *year*)。

对于日度数据、月度数据以及季度数据，其导入方法略为复杂。

如果数据中含有格式为“1949-10-01”或“1949/10/01”的时间变量，在导入 Stata 后，可能被视为“字符串”(string)，而非“数字型”(numeric)，无法直接对其进行运算(比如，滞后一期)。

对于日度数据(daily data)，可使用如下命令将其转换为“整数日期变量”(integer date variable)：

```
. gen newvar=date(varname,"YMD")
```

其中，函数“date”表示转换为日期变量；“varname”为原来的时间变量，而“newvar”为新定义的时间变量。

"YMD"告诉 Stata，原始数据的格式为“年-月-日”。如果数据格式为“月-日-年”，则应该为"MDY"，以此类推。

如此定义之后，新的时间变量 newvar 将以“整数日期”的形式显示。

在 Stata 内部，所有日期变量的存储格式均为“elapsed dates”(Stata 称之为 Stata Internal Format，简记 SIF)，即从 1960 年 1 月 1 日以来过了多少天。

为了让新的时间变量 `newvar` 仍以通常的日期格式(Human Readable Format，简记 HRF)在 Stata 中显示，可输入此命令：

```
. format newvar %td
```

其中，“`%td`”中的“`d`”即表示“date”。

对于月度数据(monthly data)，可使用如下命令进行转换

```
. gen newvar=monthly(varname,"YM")
```

其中，函数“`monthly`”表示转换为月度变量；"YM"告诉 Stata，原始数据的格式为“年-月”。

对于新定义的时间变量 `newvar`，Stata 内部的月度变量存储格式为“elapsed months”，即从 1960 年 1 月以来过了多少月。

如果仍想以日期格式在 Stata 中显示，可输入命令：

```
. format newvar %tm
```

其中，“`%tm`”中的“`m`”即表示“month”。

对于季度数据(quarterly data)，可使用以下命令进行转换

```
gen newvar=quarterly(varname,"YQ")
```

其中，函数“`quarterly`”表示转换为季度变量；"YQ"告诉 Stata，原始数据的格式为“年-季”。

对于新定义的时间变量 `newvar`，Stata 内部的季度变量存储格式为“elapsed quarters”，即从 1960 年 1 月以来过了多少季度。

如果仍想以日期格式在 Stata 中显示，可输入命令：

```
. format newvar %tq
```

其中，“`%tq`”中的“`q`”即表示“quarter”。

如果在原始数据中，年、月、日分别以数字型(numeric)变量“Y, M, D”来表示，可使用以下命令将其合成为单一的日期变量，

```
. gen newvar=mdy(M,D,Y)
```

有关日期数据的更多说明，参见“`help date`”。