Ch7：一元时间序列分析

SAS/ETS 中一元时间序列分析的过程主要包括 proc arima 和 proc autoreg，可用于时间序列的平稳性检验、模型识别、模型估计及预测等。proc autoreg 可以估计条件异方差中的 ARCH 和 GARCH 类模型。本章以蔡瑞胸(2012)中的符号和模型表示为基础。

为方便结果查看，在SAS主界面点击菜单：工具→选项→参数选择→结果，在弹出的对话框中勾选“创建列表”、“创建HTML”、“生成后立即查看结果”和“使用ODS图形”，并在“样式”后的下拉列表中选择“printer”。

7.1 Proc arima 过程

proc arima对平稳时间序列建立自回归移动平均（ARMA）模型，对非平稳（单位根）过程建立单整（integrated）自回归移动平均（ARIMA）模型。模型一般形式为

$$y_t=\phi_0+\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}-\theta_1 a_{t-1}+\cdots-\theta_q a_{t-q}+a_t, a_t\sim i.i.N(0,\sigma^2) \qquad (1)$$

$\{a_t\}_{t=1}^T$为正态白噪声序列，a_t称为新息（innovation）或者冲击（shock）。用滞后算子B可将ARIMA(p,q)表示为

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)y_t=\phi_0+\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)a_t \qquad (2)$$

其中滞后多项式$\Phi(B)\equiv 1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$和$\Theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$称为模型的自回归因子和移动平均因子。需要注意，proc arima按（1）式中形式报告参数估计结果，即移动平均系数报告为θ_i而不是$-\theta_i$！

另外，从（1）式给出的ARMA模型看出，proc arima过程不考虑a_t的条件异方差，不能建立ARCH和GARCH类模型。

proc arima可用于模型分析的全过程，从平稳性检验、模型识别到模型估计诊断和预测，过程基本语法为

```
PROC ARIMA options;
IDENTIFY VAR=variable options;
ESTIMATE options;
FORECAST options;
By variables;
run;
```

语句Identify、Estimate和Forecast分别用于模型识别、估计和预测。模型识别包括平稳性检验和模型阶数(p、q)确定。proc arima的运行为交互run组执行方式(见Ch5)，可以多次提交run组运行，直到quite语句或者运行其他步才退出。

首先对PROC ARIMA及其选项进行介绍，然后按功能介绍其他语句。PROC ARIMA语句用于过程的启动，选项（options）包括

（1） 数据集选项：

（i）data=数据集名：指定输入数据集

（ii）out=数据集名：指定输出数据集，存放预测值。

（2） 图形输出选项：plots=options，其中选项（options）指定要输出的图形，例如plots=all输出所有图形，而plots=none则不输出图形。

7.1.1 单位根（平稳性）检验

时间序列是否平稳决定着模型类型，建立模型前需要进行平稳性检验。平稳性检验方法可参考R.Tsay（2012）或者SAS帮助。平稳性检验属于模型识别，用identify语句进行。

对原序列进行单位根检验

语句格式为

IDENTIFY VAR=variable STATIONARITY=(options);

VAR=指定要检验的变量名，variable为变量名，STATIONARITY=(options)设定检验细节，包括检验方法、检验模型阶数等，通过options来确定。

STATIONARITY=(ADF=n):采用ADF（增广Dickey-Fuller）检验，检验模型分别1到n阶的自回归模型；

STATIONARITY=(ADF=(n1,n2,…,nk):分别用n1、n2、……、nk阶自回归模型作为检验模型进行ADF检验。

只给出检验方法不指定模型阶数时，默认检验模型自回归阶数为0（白噪声）、1（随机游动）和2。其它检验方法中检验模型阶数的设定方法相同。

STATIONARITY=(PP=n):采用Phillips-Perron检验，检验模型为n阶自回归模型；

采用给定阶数自回归模型进行ADF和PP单位根检验时，给出检验模型不带常数项（零均值）、只带常数项（单均值）和既带常数项又带趋势项（趋势）三种检验结果供参考。

数据集sashelp.stocks包含了IBM、Intel和Microsoft从1996年到2005年的月度股票交易数据，包括date（交易日期）、open（开盘价）、close（收盘价）、low（最低价）、high（最高价）、volum（交易量：手数）等变量。程序7-1对变量close进行单位根检验，每个run组中用where语句指定参与分析的观测。

程序7-1

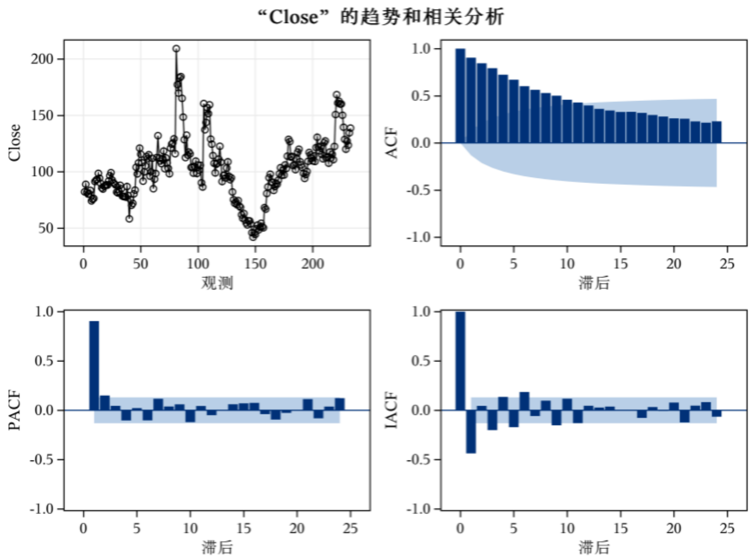
```
proc arima data=sashelp.stocks;
where stock="IBM";
identify var=close stationarity=(ADF);
run;
where stock="Intel";
identify var=close stationarity=(ADF=(5));
run;
where stock="Microsoft";
identify var=close stationarity=(PP);
run;
quit;
```

先选中第一个run组提交运行，结果窗口中“Arima:SAS系统”下的“标识1”（标识是identify的汉化翻译）中存储了检验结果，点开后出现“描述性统计量”、“白噪声的自相关检验”等分目录，点击“增广Dickey-Fuller单位根检验”出现以HTML格式和文本格式呈现的两种检验结果，点击HTML格式，在浏览器窗口显示相关结果。IBM的收盘价单位根检验结果呈现为

增广 Dickey-Fuller 单位根检验							
类型	滞后	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
零均值	0	-0.9268	0.4850	-0.54	0.4827		
	1	-0.5011	0.5692	-0.35	0.5574		
	2	-0.3399	0.6050	-0.25	0.5937		
单均值	0	-20.7252	0.0086	-3.23	0.0198	5.27	0.0305
	1	-14.8650	0.0387	-2.64	0.0867	3.54	0.1677
	2	-13.6681	0.0525	-2.50	0.1180	3.19	0.2580
趋势	0	-21.9655	0.0419	-3.34	0.0622	5.60	0.0890
	1	-15.9333	0.1469	-2.76	0.2150	3.83	0.4118
	2	-14.6204	0.1893	-2.60	0.2828	3.39	0.5007

没有设定检验模型阶数，默认采用0阶、1阶和2阶进行检验，三种不同检验模型（零均值、单均值、趋势）的结果一并列出。“零均值”检验模型使用的检验统计量为ρ（Rho）检验统计量和τ（Tau）检验统计量，而“单均值”和“趋势”检验模型使用的检验统计量还包括F检验统计量。具体定义可参考SAS帮助中SAS/ETS 14.3 User’s Guide→General Information→SAS Macros and Functions→Functions。

执行identify后，输出结果中除各种单位根检验结果外，还包括IBM收盘价序列白噪声检验结果以及序列的趋势和相关分析，后者以图形的形式呈现：



包括时序图、自相关函数（ACF）图、偏自相关函数（PACF）图以及逆自相关函数（IACF）图。

点击“增广Dickey-Fuller单位根检验”下文本格式结果，在弹出的SAS输出窗口中以文本方式显示检验结果。

此时程序编辑器窗口题头栏显示“PROC ARIMA正在运行”。选中第二个run组并提交

```
where stock="Intel";
identify var=close stationarity=(ADF=(5));
run;
```

结果窗口中“标识2”下给出各种检验结果，包括检验模型为5阶自回归模型的单位根ADF检验结果：

增广 Dickey-Fuller 单位根检验							
类型	滞后	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
零均值	5	-3.4950	0.1982	-1.29	0.1818		
单均值	5	-17.2757	0.0208	-2.68	0.0794	3.59	0.1553
趋势	5	-16.9731	0.1193	-2.65	0.2575	3.73	0.4329

从结果看，三种检验模型的检验结果均不能拒绝原假设，表明Intel的close序列为趋势平稳序列。

此时程序编辑器窗口题头栏显示“PROC ARIMA正在运行”。选中第三个run组并提交

```
where stock="Microsoft";
identify var=close stationarity=(PP);
run;
```

可以得到Phillips-Perron单位根检验结果：

Phillips-Perron 单位根检验					
类型	滞后	Rho	Pr < Rho	Tau	Pr < Tau
零均值	0	-3.3538	0.2076	-1.29	0.1803
	1	-2.8346	0.2469	-1.19	0.2138
	2	-2.6387	0.2640	-1.15	0.2282
单均值	0	-21.7678	0.0066	-3.37	0.0132
	1	-19.4620	0.0119	-3.20	0.0217
	2	-19.0042	0.0134	-3.16	0.0239
趋势	0	-22.4713	0.0376	-3.29	0.0709
	1	-19.9625	0.0645	-3.09	0.1108
	2	-19.5055	0.0710	-3.05	0.1200

三种模型和同一种模型不同阶数的检验结果不一致。零均值模型检验结果不能拒绝原假设，即Microsoft的收盘价close是平稳序列，而单均值模型检验结果则拒绝原假设，趋势模型不同滞后阶数的检验结果不一致。

最后在编辑器窗口输入quite并选中提交退出proc arima过程。

随机游动是一种特殊的单位根过程，语句identify可以对变量是否是带漂移项的随机游动进行检验，检验方法选择为RW，即STATIONARITY=(RW)可进行随机游动检验。

对差分序列进行单位根检验

一元时间序列如果非平稳，需要进行差分转换为平稳时间序列后才能建立arma模型。非平稳时间序列y_t差分$k-1$次后非平稳，差分k此后平稳，称为k阶单整序列，记为$y_t \sim I(k)$。对k次差分后的序列建立的arima(p,q)模型称为arima(p,k,q)模型。

要检验变量差分后是否平稳，只需在identify var=变量名后面添加小括号，并在括号内给出差分阶数即可，即

```
IDENTIFY VAR=variable(1<,1,…,1>) STATIONARITY=(options);
```

表示先对变量进行一阶差分，然后对差分后的变量再进行一阶差分，……。例如y(1)表示对y进行一阶差分，y(1,1)表示对y进行二阶差分，即$(y_t-y_{t-1})-(y_{t-1}-y_{t-2})$以此类推。注意y(2)表示滞后2期的一阶差分，即$y_t-y_{t-2}$，y(d)表示$y_t-y_{t-d}$等，因此语句

```
IDENTIFY VAR=variable(d1,d2) STATIONARITY=(options);
```

要检验的序列由原变量（假设为y）先做滞后d1期的一阶差分，对差分后的变量再做滞后期为d2的一阶差分，即（y_t-y_{t-d1}）-（$y_{t-d2}-y_{t-d1-d2}$）。程序7-2对IBM股票收盘价close的一阶差分进行单位根检验。

程序7-2

```
proc arima data=sashelp.stocks(where=(stock="IBM")) ;
identify var=close(1) stationarity=(ADF=2);
run;
quit;
```

输出窗口第一部分信息显示“变量名称=close，差分期间 1”等信息，表明是对一阶差分后的变量进行各种检验。

7.1.2 模型定阶

定阶是用样本信息确定合适的模型阶数，包括自回归阶数p和移动平均阶数q。只有确定阶数才能对模型进行估计。

常用的定阶指标包括自相关函数ACF（AutoCorrelation Function）、偏自相关函数PACF（Partial ACF）以及AIC、BIC信息准则等。过程proc arima采用扩展ACF（Extended ACF）EACF、最小信息准则MINIC（Minimum Information Criteria）和最小典型相关系数SCAN（Smallest CANonical Correlation）确定模型尝试性(tentative order) 阶数，同时画出自相关函数ACF图、偏自相关函数PACF图和逆自相关函数IACF图供参考。

(1)定阶方法

ESACF 方法

ARMA 模型需要确定的阶数分为自回归阶数和移动平均阶数，EASCF 方法首先选定阶数范围(1：p+d:，0：q+1)，然后对阶数从 1 到 p+d 的自回归模型进行估计，得出回归残差，将残差看做移动平均成分并用 ACF 确定阶数。计算步骤如下

（i）给定自回归阶数 m=1，…，p+d，进行如下回归：

Step1：取 q=0，用 OLS 方法拟合模型 AR(m)，

$$y_t = \phi_1^{(m,0)} y_{t-1} + \phi_2^{(m,0)} y_{t-2} + \cdots + \phi_m^{(m,0)} y_{t-m} + a_t$$

得出自回归系数估计值$\hat{\phi}_i^{(m,0)}, i = 1,2,\dots,m, m = 1,2,\dots,p+q$。

Step2：当移动平均阶数 q=j 不等于 0 时，采用如下递推公式计算模型 ARMA 中自回归系数的估计值（注意：没有用到移动平均系数）：

$$\hat{\phi}_i^{(m,j)} = \hat{\phi}_i^{(m+1,j-1)} - \hat{\phi}_{i-1}^{(m,j-1)} \frac{\hat{\phi}_{m+1}^{(m+1,j-1)}}{\hat{\phi}_m^{(m,m-1)}}$$

$$m = 1,2,\dots,p+d, j = 1,2,\dots,q+1$$

Step3：计算模型残差序列

$$\hat{a}_t^{(m,j)} = y_t - \hat{\phi}_1^{(m,j)} y_{t-1} + \hat{\phi}_2^{(m,j)} y_{t-2} + \cdots + \hat{\phi}_m^{(m,j)} y_{t-m}$$

（ii）计算残差序列$\hat{a}_t^{(m,j)}$的 j 阶样本自相关函数值$r_j(m)$

$$r_j(m) = \frac{\frac{1}{T-j-1}\sum_{t=j+1}^{T} \hat{a}_t \hat{a}_{t-j}}{\frac{1}{T-1}\sum_{t=1}^{T} \hat{a}_t^2}$$

并对$r_j(m)$进行显著性检验，$r_j(m)$的方差采用样本自相关函数方差的 Bartlett 近似公式计算，即$Var\left(r_j(m)\right) \approx (1 + \sum_{t=1}^{j} r_j^2(m))$。

如果数据服从的真实模型是 ARMA(p+d,q)，则当$j \geq q$时残差序列$\hat{a}_t^{(m,j)}$服从 MA(q)模型，因此

$$r_j(p+q) \geq 0, \quad j > q$$
$$r_j(p+q) \neq 0, \quad j = q$$

Tsay 和 Tiao（1984）证明 ESACF 满足

$$r_j(m) \approx 0, \quad j-q > m-p-d \leq 0$$
$$r_j(m) = c(m-p-d, j-q), \quad 0 \leq j-q \leq m-p-d$$

其中$c(m-p-d,j-q)$为非 0 常数，或者下界为-1、上界为+1 的连续随机变量。

采用 ESACF 方法可以确定 ARMA 模型的阶数。选定自回归阶数范围p_{min}:p_{max}和自移动平均回归阶数q_{min}:q_{max}，用$m = p_{min}, p_{min}+1, \ldots, p_{max}$和$j = q_{min}, q_{min}+1, \ldots, q_{max}$对应的$r_j(m)$值构造 ESACF 表，如表 7.1。

表 7.1 ESACF 表

	MA					
AR	0	1	2	3	.	.
0	$r_{1(0)}$	$r_{2(0)}$	$r_{3(0)}$	$r_{4(0)}$	.	.
1	$r_{1(1)}$	$r_{2(1)}$	$r_{3(1)}$	$r_{4(1)}$	.	.
2	$r_{1(2)}$	$r_{2(2)}$	$r_{3(2)}$	$r_{4(2)}$	.	.
3	$r_{1(3)}$	$r_{2(3)}$	$r_{3(3)}$	$r_{4(3)}$	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

用 t 检验判断 ESACF 表中的$r_j(m)$是否显著不为 0，表中 0 值形成的楔形区域左上角对应的阶数就是 ARMA 模型的阶数。表 7.2 给出的 ESCAF 表将模型阶数确定为 ARMA（1,2）。

表 7.2 ARMA(1,2)的 ESCAF 表

	MA							
AR	0	1	2	3	4	5	6	7
0	*	**X**	**X**	**X**	**X**	**X**	**X**	**X**
1	*	**X**	0	0	0	0	0	0
2	*	**X**	**X**	0	0	0	0	0
3	*	**X**	**X**	**X**	0	0	0	0
4	*	**X**	**X**	**X**	**X**	0	0	0
X=显著（不为 0）;0=不显著;*=没有规律性								

ESACF 方法既可以确定平稳序列的 ARMA 模型阶数，也可以确定非平稳序列的阶数。

MINIC 方法

MINIC 方法可以尝试性确定平稳可逆（stationary and invertable）时间序列的 ARMA 模型阶数。设z_t为零均值化的平稳可逆时间序列，服从 ARMA(p,q)模型，滞后算子表达式为

$$\Phi(B)z_t = \Theta(B)\epsilon_t$$

平稳可逆 ARMA(p,q)模型可以转化为 AR(∞)模型，如果仅建立 AR(p)模型（即 q=0），可在设定的自回归阶数范围$[p_{Amin}, p_{Amax}]$内用 AIC 准则确定模型阶数，记为p_A。

采用 AR 模型拟合时间序列数据时，往往需要较大的自回归阶数才能充分捕捉数据中的自相关性，采用 ARMA 模型则能减少模型阶数。为确定 ARMA 模型的阶数，在设定的自回归阶数范围$[0, p_{max}]$和移动平均阶数范围$[0, q_{max}]$内，选择$m \in [0, p_{max}]$和$j \in [0, q_{max}]$，估计模型

$$z_t = \phi_1^{(m,j)} z_{t-1} + \phi_2^{(m,j)} z_{t-2} + \cdots + \phi_m^{(m,j)} z_{t-m}$$
$$+\theta_1^{(m,j)} \hat{\epsilon}_{t-1} + \theta_2^{(m,j)} \hat{\epsilon}_{t-2} + \cdots + \theta_j^{(m,j)} \hat{\epsilon}_{t-j} + e_t$$

计算 BIC（Bayes 信息准则）值。不同(m,j)对应的 BIC 值形成表 7.3

表 7.3 MINIC 表

	MA					
AR	0	1	2	3	.	.
0	BIC(0,0)	BIC(0,1)	BIC(0,2)	BIC(0,3)	.	.
1	BIC(1,0)	BIC(1,1)	BIC(1,2)	BIC(1,3)	.	.
2	BIC(2,0)	BIC(2,1)	BIC(2,2)	BIC(2,3)	.	.
3	BIC(3,0)	BIC(3,1)	BIC(3,2)	BIC(3,3)	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

选择最小 BIC 值对应的阶数作为 ARMA 模型阶数。

当$p_{max} > p_{Amin}$时，表明p_{Amin}设定过小，AR(p_A)不能将数据中的自相关完全滤波干净，导致$\hat{\epsilon}_t$序列存在自相关性。如上定阶方法失效。

除了 ESACF 法和 MINIC 法之外，proc arima 还提供了基于样本典型（Sample Canonical）相关系数的定阶方法 SCAN，详情可参考 SAS 帮助。

(2)用identify var=确定模型阶数

模型定阶属于模型识别，用 identify 语句进行。语句语法为

```
IDENTIFY VAR=variable Option
```

Option 是 ESACF、MINIC 和 SCAN 中的一个，对应三种定阶方法。三种定阶选项后面均可以添加更多选项设定阶数的范围，p=(p1:p2)和 q=(q1:q2)分别用于设定 ARMA 模型的自回归阶数范围和移动平均阶数范围。MINIC 语句中添加 PERROR=（pe1:pe2）可以设定计算自回归模型估计残差$\hat{\epsilon}_t$的自回归模型阶数范围。

程序 7-3 对股票 IBM 的收盘价 close 一阶差分采用 ESCAF 和 MINIC 定阶。

程序7-3

```
proc arima data=sashelp.stocks(where=(stock="IBM"));
identify var=close(1) ESACF;
identify var=close(1) MINIC;
identify var=close(1) ESACF q=(0:0);
identify var=close(1) MINIC q=(0:0);
run;
```

第一个 identify 语句计算的 ESACF 值和对应的显著性水平为

扩展样本自相关函数						
滞后	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	-0.1924	-0.0330	0.0887	-0.0799	0.0773	-0.1842
AR 1	-0.3329	-0.0763	0.0403	0.0036	-0.0183	-0.1799
AR 2	0.4994	0.0389	0.0523	0.0070	-0.0009	-0.1583
AR 3	0.4795	-0.4384	0.0097	-0.0392	0.0107	-0.0952
AR 4	0.4931	0.0728	0.0745	-0.0044	-0.0347	-0.0751
AR 5	0.3100	-0.1243	0.1544	-0.0802	-0.1039	-0.1304

ESACF 概率值						
滞后	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.0034	0.6281	0.1928	0.2440	0.2626	0.0079
AR 1	<.0001	0.2554	0.5876	0.9620	0.7896	0.0173
AR 2	<.0001	0.6350	0.4739	0.9261	0.9902	0.0389
AR 3	<.0001	<.0001	0.9117	0.6446	0.8817	0.2716
AR 4	<.0001	0.3869	0.3431	0.9507	0.6542	0.4551
AR 5	<.0001	0.0851	0.0277	0.2565	0.2034	0.2343

得出的尝试性阶数为（3,2）和（4,2）供选择。可采用 model checking 方法根据阶数充足性（adequacy）原则确定模型阶数。具体做法是对两种阶数的模型进行估计，对估计残差进行白噪声检验，残差为白噪声的模型阶数为最终选择。

输出结果还包括序列 close(1)的时间序列散点图、ACF 图、PACF 图和 IACF（逆自相关函数）图。IACF 的定义和使用可参考 SAS 帮助。ACF、PACF 等图中的灰色带状区域

是以 0 为中心的 95%置信区间，对应的值落入区域时，可认为该滞后阶数对应的 ACF 值、PACF 值为 0，据此可以对 close 一阶差分的自相关性作出判断，在确定模型阶数时参考。

第二个 identify 语句得出计算残差的 AR 模型阶数为 9，BIC 值表为

最小信息准则						
滞后	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	4.953627	4.943954	4.965033	4.973763	4.987187	5.006394
AR 1	4.938896	4.962341	4.973961	4.986106	5.005548	5.026738
AR 2	4.956843	4.977058	4.997229	4.999986	5.019614	5.041532
AR 3	4.974022	4.996045	5.006809	5.022472	5.034634	5.056264
AR 4	4.994091	5.016969	5.029677	5.036906	5.05811	5.079693
AR 5	5.012055	5.0354	5.050255	5.058307	5.075353	5.089308

其中 BIC(0,1)=4.938896 最小，建议对数据建立 ARMA(1,0)模型。

由此看出，ESACF 和 MINIC 不带选项 p=(p1:p2)和 q=(q1:q2)时默认的阶数范围均为 p=(0:5)和 q=(0:5)。此外，两种方法得出的模型建议阶数并不一样，可根据实际情况选择，也可通过模型残差白噪声检验来确定。

第三个和第四个 identify 语句通过添加 q=(0:0)将备选模型限定在自回归模型(MA 阶数限定为 0），此时 ESACF 和 MINIC 方法得出的自回归阶数均为 1。

7.1.3 模型估计和诊断

ARIMA 模型估计方法有三种，分别是条件最小二乘 CLS（Conditional Least Square)、无条件最小二乘 ULS（Unconditional Least Square）和极大似然 ML。proc arima 中语句 ESTIMATE 用于模型估计，通过选项设定模型阶数、估计方法、输出数据集、输出图形等。语句语法为

```
ESTIMATE options;
```

语句选项（options）包括

（1）模型阶数和估计方法选项

模型阶数设定

p=n：将模型自回归阶数设定为 n，模型为

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + a_t$$

p=(n1,n2,…,nk)：模型中只包括括号内给出的滞后阶数，模型为

$$y_t = \phi_0 + \phi_1 y_{t-n_1} + \phi_2 y_{t-n_2} + \cdots + \phi_k y_{t-n_k} + a_t$$

q=m：将模型移动平均阶数设定为 m，模型为

$$y_t = \mu - \theta_1 a_{t-1} + \cdots - \theta_q a_{t-m} + a_t$$

q=(m1,m2,…,ml)：模型中只包括括号内给出的滞后阶数，估计模型为

$$y_t = \mu + \theta_1 a_{t-m_1} + \theta_2 a_{t-m_2} + \cdots + \theta_l a_{t-m_l} + a_t$$

一般情况下同时使用选项 p=和 q=，将模型设定为自回归移动平均模型。

noint：模型不带常数项（截距）。

估计方法设定

method=option

option 有三个选项 CLS、ML 和 ULS，分别对应条件最小二乘、极大似然和无条件最小二乘估计方法。

CLS 将样本前值（pre-sample）$y_0, y_{-1}, \ldots, y_{-p}$ 和 $a_0, a_{-1}, \ldots, a_{-q}$ 设为 0，带入 $y_1, y_2, \ldots, y_{\min(p,q)}$ 的 ARMA(p,q)模型中，以此进入目标函数对参数进行估计，模型系数最

小二乘估计具有解析表达式。ULS 和 ML 则将$y_0, y_{-1}, \ldots, y_{-p}$和$a_0, a_{-1}, \ldots, a_{-q}$看做未知参数和模型系数一起估计，估计没有解析表达式，用数值算法极小化残差平方得出数值解。

估计结果输出

以程序 7-4 运行结果进行说明。程序 7-4 在程序 7-3 识别的基础上，运行两个 run 组，采用 ML 方法对 IBM 的 close 一阶差分序列拟合 ARIMA(2,2)模型，采用 uls 方法对 Intel 的 close 一阶差分序列拟合 ARIMA(1,4)。

程序7-4

```
proc arima data=sashelp.stocks(where=(stock="IBM"));
identify var=close(1) ESACF;
estimate p=3 q=2 method=ml;
run;
estimate p=4 q=2 method=uls;
run;
quit;
```

第一个 run 组运行结果为

最大似然估计					
参数	估计	标准误差	t 值	近似 Pr > |t|	滞后
MU	0.14373	0.21139	0.68	0.4966	0
MA1, 1	0.06127	4.61130	0.01	0.9894	1
MA1, 2	0.93861	4.32439	0.22	0.8282	2
AR1, 1	-0.14601	0.09069	-1.61	0.1074	1
AR1, 2	0.83607	0.08332	10.03	<.0001	2
AR1, 3	0.21387	0.06844	3.13	0.0018	3

常数估计	0.013808
方差估计	133.426
标准误差估计	11.55102
AIC	1801.636
SBC	1822.317
残差数	232

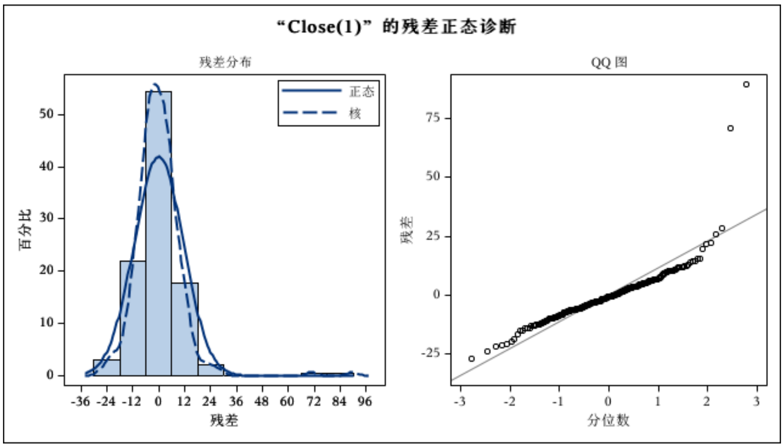
为方便叙述，写出对应的 ARIMA(2,2)模型

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} - \theta_1 a_{t-1} - \theta_2 a_{t-2} + a_t \quad (3)$$

第一个表格中给出了模型均值和回归系数估计值、标准误、t-检验统计量值以及概率值（p-value）。mu 为均值，MA1,1 和 M1,2 为移动平均系数θ_1和θ_2的估计值（不是$-\theta_1$和$-\theta_2$的估计值），AR1,1-AR1,为自回归系数ϕ_1、ϕ_2和ϕ_3的估计值。第二个表格为模型整体估计效果指标，常数估计为常数项ϕ_0的估计值，方差估计为误差项方差（无条件方差）σ^2的估计$\hat{\sigma}^2$，标准误差估计为$\hat{\sigma}$。AIC 和 SBC 分别为赤池信息准则和贝叶斯（也叫许瓦兹）信息准则值。注意模型均值和模型常数项和的关系为$mu \triangleq Ey_t = \phi_0/1 - (\phi_1 + \cdots + \phi_p)$。

估计结果输出的另一部分是模型估计残差的相关性信息，包括残差自相关性检验、残差 ACF、PACF 和 IACF 图以及残差分布的拟合密度曲线和 QQ 图，用于判断残差是否白噪声，是否服从正态分布。ARIMA(2,2)估计结果中，滞后阶数 6、12、18 对应的残差相关性检验统计量 p 值分别为 0.0197、0.2415 和 0.2999，不能得出统一的结论。残差相关诊断给出的 ACF、PACF 图形表明，不能将残差看做白噪声。ARIMA(4,2)估计结果中，滞后阶数 6、12、18 对应的残差相关性检验统计量 p 值分别为 0.0547、0.0957 和 0.0014，拒绝残差为白噪声的假设。从 ACF 和 PACF 图看出，滞后 24 期处的 ACF 和 PACF 明显不为 0，表明数据有很强的周期性，周期长度为 24 个月。

为了判断模型误差项是否服从正态分布，输出结果给出残差正态诊断图：



残差正态诊断用于判断残差是否服从正态分布，据此推断模型误差项是否服从正态分布。左边图形给出核估计得出的残差密度函数曲线，并配以正态分布密度曲线，右边给出残差的 Q-Q 图。从图形看出，模型估计残差分布表现出明显的尖峰后尾特征，明显偏离正态分布。

输出结果最后一项是用滞后算子表示的模型估计结果：

自回归因子	
因子 1:	1 + 0.14601 B**(1) - 0.83607 B**(2) - 0.21387 B**(3)

移动平均因子	
因子 1:	1 - 0.06127 B**(1) - 0.93861 B**(2)

为便于比较，将（3）式中的 ARMA(2,2)模型用滞后算子表示为

$$(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)y_t=\phi_0+(1-\theta_1 B-\theta_2 B^2)\ a_t$$

自回归因子（多项式）为$\Phi(B)=1+1.146B-0.836B^2-0.214B^3$，移动平均因子（多项式）为$\Theta(B)=1-0.061B-0.939B^2$。

（2）输出数据集选项

通过设定输出数据集，可将定阶和估计过程中产生的结果输出为 SAS 数据集。

```
OUTMODEL=SAS-data-set
```

将模型估计结果输出到 SAS 数据集，包括参数估计值、参数估计标准误和其它信息。输出 SAS 数据集包含如下自动变量

NAME：模型变量名；

TYPE：估计方法；

STATUS：估计模型时算法的收敛状态。当采用 ML 和 ULS 估计方法时，参数估计采用数值算法极大或者极小化化目标函数得出数值解。当_STATUS_值为“0 收敛”时，表明目标函数的极大化算法是收敛的，值为“1 警告”时是不收敛的。

PARM：参数名称，NUMRESID 表明残差个数，NPARMS 为模型回归系数个数，NDIFS 为建立模型前（为达到平稳）对变量的差分次数；ERRORVAR 为模型新息项方差；MU 为模型均值；MA 为移动平均系数，对应的滞后阶数由变量_LAG_值确定；AR 为自回归系数，对应变量的滞后阶数由变量_LAG_值确定。

VALUE：变量_PARM_给出的各参数的取值。

STD：回归系数估计的标准误，回归系数包括常数项 MU、自回归系数 AR 和移动平均系数 MA。

LAG：自回归系数和移动平均系数对应的滞后阶数。

OUTSTAT=SAS-data-set

将模型诊断统计量的值输出到 SAS 数据集，包括估计方法、模型诊断统计量等。输出 SAS 数据集包含如下自动变量

TYPE:估计方法，取值为 ML、CLS 和 ULS 之一；

STAT:模型诊断统计量。AIC 和 SBC 表示赤池信息准则和贝叶斯信息准则；SSE 为残差平方和；LOGLIK 为采用 ML 或者 UCL 方法时的对数似然函数值（UCL 方法和误差项正态分布假设下的 ML 方法等价）；CONV 为极大化或者极小化目标函数数值算法的收敛状况，取 0 值表示收敛，其它值表明收敛存在问题；NITER 为数值算法的迭代次数（Number of ITERation）。其它变量名的含义与 **OUTMODEL**=SAS-data-set 选项生成数据集中 _PARM_取值的参数相同。

程序 7-5 在两个 estimate 语句中添加 outmodel=和 outstat=选项，分别将第一次模型估计的参数估计结果和第二次模型估计的诊断统计量值输出为 SAS 数据集 s 和 s1。

程序7-5

```
proc arima data=sashelp.stocks(where=(stock="Intel"));
identify var=close(2) ESACF;
estimate p=2 q=2 method=ml outmodel=s;
run;
estimate p=1 q=4 method=uls outstat=s1;
run;
quit;
```

需要注意，只有退出proc arima后输出数据集才能形成。

7.1.4 预测

模型估计和诊断完成之后，用forecast语句对时间序列进行预测。forecast语句必须位于estimate语句之后，且预测方法依赖于模型的估计方法，采用ULS和ML方法估计的模型，对应的预测方法称为无条件预测（unconditional forecat），采用CLS估计的模型，对应的预测方法称为条件预测（conditional forecast）。下面以ARMA（2,2）模型为例进行说明。

条件预测

采用CLS估计方法得出的模型为

$$y_t = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2} - \hat{\theta}_1 a_{t-1} - \hat{\theta}_2 a_{t-2} + a_t$$

1步预测$y_t(1)$为

$$y_t(1) = E(y_{t+1}|y_1, \dots, y_t) = \hat{\phi}_0 + \hat{\phi}_1 y_t + \hat{\phi}_2 y_{t-1} - \hat{\theta}_1 a_t - \hat{\theta}_2 a_{t-1}$$

其中a_{t-1}和a_{t-1}通过下式得出

$$\begin{aligned} a_t &= y_t - \hat{\phi}_0 - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2} - \hat{\theta}_1 a_{t-1} \\ a_{t-1} &= y_{t-1} - \hat{\phi}_0 - \hat{\phi}_1 y_{t-2} - \hat{\phi}_2 y_{t-3} - \hat{\theta}_1 a_{t-2} \\ &\dots \\ a_1 &= y_1 - \hat{\phi}_0 \end{aligned}$$

2步预测$y_t(2)$为

$$\begin{aligned} y_t(2) &= E(y_{t+2}|y_1, \dots, y_t) \\ &= E(E(y_{t+2}|y_1, \dots, y_{t+1})|y_1, \dots, y_t) \\ &= E(\hat{\phi}_0 + \hat{\phi}_1 y_{t+1} + \hat{\phi}_2 y_t - \hat{\theta}_1 a_{t+1} - \hat{\theta}_2 a_t | y_1, \dots, y_t) \\ &= \hat{\phi}_0 + \hat{\phi}_1 E(y_{t+1}|y_1, \dots, y_t) + \hat{\phi}_2 y_t - \hat{\theta}_1 E(a_{t+1}|y_1, \dots, y_t) - \hat{\theta}_2 a_t \\ &= \hat{\phi}_0 + \hat{\phi}_1 y_t(1) + \hat{\phi}_2 y_t - \hat{\theta}_2 a_t \end{aligned}$$

3步预测$y_t(3)$为

$$\begin{aligned} y_t(3) &= E(y_{t+3}|y_1,\dots,y_t) \\ &= E(E(y_{t+3}|y_1,\dots,y_{t+2})|y_1,\dots,y_t) \\ &= \hat{\phi}_0+\hat{\phi}_1 E(y_{t+2}|y_1,\dots,y_t)+\hat{\phi}_2 E(y_{t+1}|y_1,\dots,y_t)y_t-\hat{\theta}_1 E(a_{t+2}|y_1,\dots,y_t) \\ &\qquad -\hat{\theta}_2 E(a_{t+1}|y_1,\dots,y_t) \\ &= \hat{\phi}_0+\hat{\phi}_1 y_t(2)+\hat{\phi}_2 y_t(1) \end{aligned}$$

h(>3)步预测$y_t(h)$为

$$y_t(h)=\hat{\phi}_0+\hat{\phi}_1 y_t(h-1)+\hat{\phi}_2 y_t(h-2)$$

无条件预测

设$(y_1,y_2,\dots,y_t)$的协方差矩阵为V_t ，y_{t+h}和$(y_1,y_2,\dots,y_t)$ 协方差矩阵（行向量）为$C_{t,h}$。y的h步无条件预测$y_t(h)$为

$$y_t(h)=C_{t,h}V_t^{-1}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{pmatrix}$$

V_t和$C_{t,h}$可用样本和参数估计计算，具体可参考Brockwell和Davis(1991)。

forecast语句及其选项

语句格式为

forecast options;

选项options包括

ID=Vname：指定输入数据集中的变量Vname为ID变量，通常用日期时间变量作为ID变量。ID变量用于样本外预测值标识，通过外推获得样本外ID变量值。

Interval=Intname:设定时间区间类型，Intname为区间名。proc arima按给定的区间对ID变量进行外推。

Align=options：指定日期ID变量按区间取值的方式。options可以是B、M和E，对应区间的开始、中间和结束，默认值为E。该选项要配合ID=和Interval=选项使用。

Lead=n：设定预测步数h的最大值，默认24。

Out=SAS-data-set：将预测值和其他值输出为SAS数据集。输出数据集除包含输入数据集中的ID变量、By变量和模型变量（identify语句中var=给出的变量）外，还包括

Forecast：模型变量的一步和多步预测；

STD：预测标准误；

Residual：实际值和预测值之差（残差），只有样本内观测才能计算。

如果在不同的forecast语句中的out=选项中给出同一个输出数据集，则最后一个forecast语句的预测结果覆盖前面语句的预测结果。如果打算在一个数据集中存放不同forecast语句的预测值，可通过proc arima语句选项out=指定输出数据集。

输出预测图形

在proc arima语句中添加选项

plots=**forecast**(all)输出一步预测和多步预测值形成的图形。

需要注意，当identify语句中以var=x(k)设定变量x的k次差分为模型变量时，定阶、估计和预测都是按差分后变量x(k)进行，但最终的预测值是原变量x的预测值。x的预测值通过x(k)预测值多次求和得到。例如设$\Delta y_{t+1}(1)$为一阶差分变量$\Delta y_t=y_t-y_{t-1}$的一步预测，则y_t一步预测$y_t(1)$可计算为

$$y_t(1)=\Delta y_t(1)+y_t$$

类似得出多步预测计算方法。

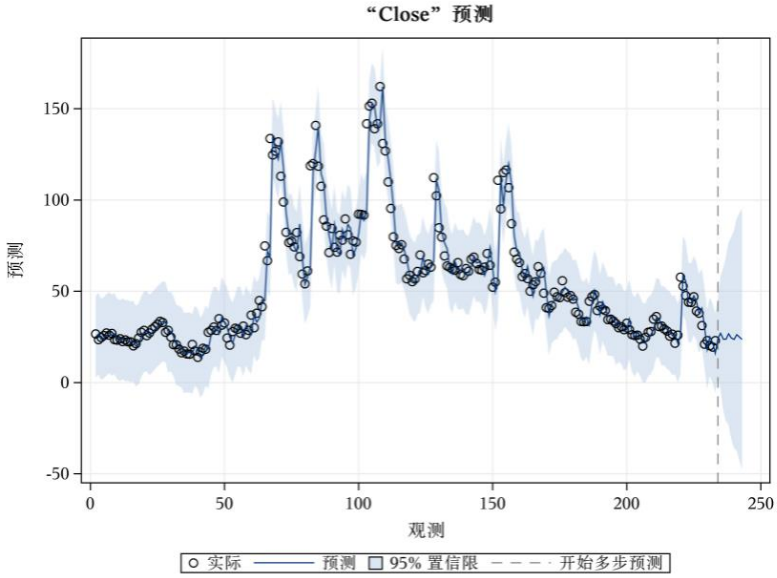
程序7-6对程序7-5中的close一阶差分变量建立并估计ARIMA(2,2)模型的基础上，

用forecast语句进行一步到10步预测，预测结果输出到SAS数据集中，并输出预测图形。

程序7-6

```
proc arima data=sashelp.stocks plots=forecast(all);
identify var=close(1) ESACF;
estimate p=2 q=2 method=ml outmodel=s;
forecast lead=10 out=f align=m;
where stock="Intel";
run;
quit;
```

输出结果预测输出分三部分：第一部分为一步预测到十步预测值及其标准误95%的置信区间上下限，第二部分为预测值和置信区间图，第三部分为预测值和实际值散点图：



图的题头是“close”预测，表明是对原始变量close的预测结果，不是其一阶差分close(1)的预测结果。散点图中以竖虚线将样本内和样本外区分开，左边为样本内预测，既有预测值又有实际值，右边为样本外预测值，没有实际值。

数据集f保存了预测结果，变量L95和U95表示预测值95%的置信区间下限和上限。

7.2 Proc AutoReg过程

proc autoreg对时间序列变量进行线性回归。时间序列回归模型有两个特点：模型误差项存在序列相关和条件异方差，proc autoreg用误差项自回归模型来捕捉其序列相关性，提高估计效率并对检验统计量进行修正，用ARCH或者GARCH类模型捕捉误差项的条件异方差，提高估计效率的同时建立因变量的波动模型，据此对因变量及其波动进行预测。

当回归模型解释变量为因变量的滞后变量时，proc autoreg处理的模型变成自回归模型。因此proc autoreg可以为时间序列建立具有GARCH误差项的AR模型，是对proc arima的补充和推广。

7.2.1 时间序列回归模型

设$y_t, x_{1t}, x_{2t}, \ldots, x_{kt}$为平稳时间序列变量，回归模型

$$y_t = c + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + u_t \tag{4}$$

是时间序列回归模型。对u_t建立m阶自回归模型捕捉误差项的自相关性，即

$$u_t = \epsilon_t - \varphi_1 u_{t-1} - \varphi_2 u_{t-2} - \cdots - \varphi_m u_{t-m} \tag{5}$$

ϵ_t为白噪声。

当$\beta_i=0,i=1,2,\dots,k$时，模型

$$\begin{aligned} y_t &= c+u_t \\ u_t &= \epsilon_t-\varphi_1 u_{t-1}-\varphi_2 u_{t-2}-\cdots-\varphi_m u_{t-m} \end{aligned} \tag{6}$$

等价于y_t的 m 阶自回归模型

$$y_t+\varphi_1 y_{t-1}+\varphi_2 y_{t-2}+\cdots+\varphi_m y_{t-m}=c(1+\sum_{i=1}^{m}\varphi_i)+\epsilon_t \tag{7}$$

当$x_{lt}=x_{t-l},l=0,1,\dots,p$时，模型

$$y_t=c+\beta_1 x_t+\beta_2 x_{t-1}+\cdots+\beta_p x_{t-p}+u_t \tag{8}$$

为自回归分布滞后模型。

（4）式中$c+\beta_1 x_{1t}+\beta_2 x_{2t}+\cdots+\beta_k x_{kt}$称为结构化部分（structural part），是y_t的确定性趋势部分，$u_t=\epsilon_t-\varphi_1 u_{t-1}-\varphi_2 u_{t-2}-\cdots-\varphi_m u_{t-m}$称为自回归部分（时间序列部分)。模型估计后的残差分也为两种：结构化残差(ResidualM)和总残差(Residual)。结构化残差为

$$RM=y_t-\hat{c}-\sum_{l=1}^{k}\hat{\beta}_l x_{lt} \tag{9}$$

总残差为

$$R=y_t-\hat{c}-\sum_{l=1}^{k}\hat{\beta}_l x_{lt}-\sum_{s=1}^{m}\hat{\varphi}_s\hat{u}_{t-s} \tag{10}$$

实际上结构化部分就是y_t对解释变量$x_{it},i=1,2,\dots,k$的回归函数，不考虑误差项中存在的自相关。

ARCH和GARCH类模型是误差项的条件异方差动态模型，具有如下形式（以GARCH(p,q)为例）

$$\begin{aligned} y_t &= c+\beta_1 x_{1t}+\beta_2 x_{2t}+\cdots+\beta_k x_{kt}+u_t \\ u_t &= \epsilon_t-\varphi_1 u_{t-1}-\cdots-\varphi_m u_{t-m} \\ \epsilon &= \sqrt{h_t}e_t, \qquad e_t\sim i.i.N(0,1) \\ h_t &= \omega+\sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2+\sum_{i=1}^{p}\gamma_j h_{t-j} \end{aligned} \tag{11}$$

Proc Autoreg 可以建立的 GARCH 类型包括

- GARCH （广义 ARCH：Generalized ARCH）
- IGARCH（单整 GARCH：Integrated GARCH）
- EGARCH（指数 GARCH：Exponential GARCH）
- QGARCH（二次 GARCH：Quadratic GARCH）
- TGARCH（门限 GARCH：Threshold GARCH）
- PGARCH（指数 GARCH：Power GARCH）
- GARCH-M（均值中 GARCH GARCH-in-mean）

可以在各种 GARCH 模型中添加外生变量。

7.2.2 proc autoreg语法

Proc autoreg 语句构成及其语法为

```
PROC AUTOREG options;
MODEL dependent=<regressors>/<options>;
HETERO variables/<options>;
```

```
OUTPUT <OUT=SAS-data-set><options><keyword=name>;
RESTRICT equation,…,equation ;
TEST equation,…,equation/<option>;
Run;
```

Proc autoreg 功能包括

检验

proc autoreg 提供了丰富的时间序列检验和模型检验方法，包括

- 单位根检验：提供 ADF、ERS、NP、KPSS 和 PP 等常见的单位根检验方法，比 proc arima 提供的检验方法更为丰富。
- 误差项自相关检验：包括 Durbin-Watson 检验和 h-Durbin 检验等。
- ARCH 检验：提供包括 Engle 拉格朗日乘子检验在内的四种 ARCH 检验方法。
- 变点检验：包括 Chow 变点检验和 Bai-Perron 结构变化检验。
- 参数约束检验：对参数之间的线性约束进行检验，包括回归系数约束检验和 GARCH 模型系数检验。

估计

proc autoreg 提供了四种模型估计方法

- 尤勒-沃尔克(Yule-Walker) 方法
- 迭代尤勒-沃尔克(iterated Yule-Walker)方法
- 无条件最小二乘（unconditional least squares）方法
- 精确极大似然（exact maximum likelihood）

其中无条件最小二乘和精确极大似然方法与 Proc arima 中的 ULS 和 ML 类似。尤勒-沃尔克方法首先对（4）进行 OLS 估计得出估计残差$\hat{u}$，用残差$\hat{u}$代替误差u计算出$\hat{u}$的样本自相关函数并带入误差项u的尤勒-沃尔克方程，从中计算出(5)中的自回归系数$\varphi_i, i = 1,2,\ldots,m$，进而计算出误差向量$\boldsymbol{u} = (u_1, u_2, \ldots, u_T)'$的方差协方差矩阵，以此对（4）式进行广义最小二乘法求出回归系数$\beta_i, i = 1,2,\ldots,k$的估计值。迭代尤勒-沃尔克则是将以上方法迭代直至收敛。

proc autoreg 提供四种参数估计方差和标准误的稳健估计，包括 White 估计和 Newey-West 估计。

诊断

这里是指模型估计后对模型设定正确性和估计效果的评价。

- 模型自回归阶数和 GARCH 阶数(p,q)设置是否充分（adequate），可用过检验残差或者标准化残差是否为不相关序列或者独立序列判断。标准化残差是用条件标准误$\hat{h}_t$对残差进行标准化，即$\hat{e}_t = \hat{\epsilon}_t/\hat{h}_t$ 。检验方法包括 BDS 检验、游程检验（run test）、反转点检验（turning point test）和秩形式的 von Neumann 比检验（rank version of von Neumann ratio test）。
- 误差项正态分布检验，检验ϵ_t（没有 GARCH 效应）或者$e_t = \epsilon_t/h_t$ （有 GARCH 效应）是否服从正态分布。检验方法为 J-B 检验。

预测

用估计出的模型对变量及其条件方差进行预测。

- **变量预测**：变量预测分为两种：基于结构成分的预测和基于完整模型的预测。

（i）基于结构成分的预测$\hat{y}_t$及其置信区间上下限 $LCLM_t$ 和 $UCLM_t$ 为

$$\begin{aligned}\hat{y}_t &= \hat{c} + \textstyle\sum_{i=1}^{k} \hat{\beta}_i x_{it} \\ LCLM_t &= \hat{y}_t - t_{\alpha/2} v \\ UCLM_t &= \hat{y}_t + t_{\alpha/2} v\end{aligned} \tag{12}$$

v为$\hat{y}_t$的标准误，$t_{\alpha/2}$为t分布的$(1-\alpha/2)$分位点。

（ii）基于完整模型的预测$\tilde{y}_t$及其置信区间上下限$\widehat{LCLM}_t$和$\widehat{LCLM}_t$为

$$\begin{aligned}\tilde{y}_t &= \hat{c} + \sum_{i=1}^{k}\hat{\beta}_i x_{it} + u_{t|t-1} \\ \widehat{LCLM}_t &= \tilde{y}_t - t_{\alpha/2}v \\ \widehat{UCLM}_t &= \tilde{y}_t + t_{\alpha/2}v\end{aligned} \tag{13}$$

$u_{t|t-1}$为误差项的一步预测。

■ **条件方差（波动）预测**：以 GARCH 模型为例，将 GARCH(p,q)模型转化为误差项ϵ_t^2的 ARMA 模型后推导出

$$\epsilon_t^2 = \omega + \sum_{i=1}^{n}(\gamma_i + \alpha_i)\,\epsilon_{t-i}^2 + \sum_{i=1}^{p}\gamma_j\eta_{t-j} + \eta_t \tag{14}$$

其中$\eta_t = \epsilon_t^2 - h_t$，$n = \max(p,q)$，按 ARMA 模型预测方法对条件方差$h_t \triangleq E(\epsilon_t^2|F_{t-1})$进行预测。

与 proc arima 的搭配

proc arima 过程可以确定时间序列自回归阶数，proc autoreg 不能；proc autoreg 可以对时间序列建立 GARCH 类模型，proc arima 模型不能。可以先采用 proc arima 确定序列自回归阶数 m，再采用 proc autoreg 估计具有 GARCH 误差项的自回归模型。

proc autoreg 不能采用交互式 run 组运行方式，遇到 run 语句就结束运行并退出。

7.2.3 单位根（平稳性）检验

proc autoreg在语句model中添加stationarity=选项对序列进行单位根检验，比proc arima提供更多单位根检验方法。语句格式为

```
model vname=/STATIONARITY(test<=test-options>);
```

model语句选项在斜杠“/”之后添加。选项与proc arima中语句identify中单位根检验选项stationarity使用方法类似，只是可选择的检验方法更多。这里仅对新的检验方法设定进行介绍，proc arima中的ADF检验和PP检验不再介绍。

■ STATIONARITY=(ERS):选择ERS（Elliott-Rothenberg-Stock）单位根检验方法。ERS检验用贝叶斯信息准则确定检验模型的滞后阶数，最小滞后阶数为3，默认最大滞后阶数为8。STATIONARITY=(ERS=n)形式的选项可设定检验模型最大滞后阶数。

■ STATIONARITY=(NG):选择NG（Ng-Perron）单位根检验方法。NG检验用AIC信息准则确定检验模型的滞后阶数并用广义最小二乘方法消除数据中的趋势。默认最大滞后阶数为8。STATIONARITY=(NG=n)形式的选项可设定检验模型最大滞后阶数。

■ STATIONARITY=(KPSS):选择KPSS（Kwiatkowski-Phillips-Schmidt-Shin）单位根检验方法。KPSS检验的最大特点是以变量不存在单位根（即平稳）为原假设，拒绝原假设得出存在单位根的结论。假设检验两类错误问题和非平稳时间序列的普遍性使KPSS检验具有特殊意义：能够将变量存在单位根结论的错误（即第一类错误）概率控制在显著水平以下。其它单位根检验方法均以存在单位根为原假设。

如果model语句中设定的回归模型带截距项,单位根检验的原假设模型只有均值以及均值和确定性趋势两种，报告的检验结果也只有两种，不包括既没有均值也没有确定性趋势的情况。回归模型不带截距项（model语句有noint选项）时，单位根检验的原假设模型分为0均值、均值以及均值和确定性趋势三种，报告三种检验结果。

程序7-7对sashelp.stocks中变量close进行单位根检验，检验方法选择kpss。采用by语句对IBM、Intel和Microsoft股票的月收益序列分别进行。

程序7-7

```
proc autoreg data=sashelp.stocks;
model close=/stationarity=(kpss);
model close=/noint stationarity=(kpss);
by stock;
run;
```

以IBM的输出结果（进行了编辑以节省空间）为例进行说明。第一个model语句的检验结果（结果窗口“模型1”中“平稳性检验”下“KPSS检验”）为

AUTOREG 过程

Stock=IBM

因变量	Close

KPSS 平稳性检验			
类型	滞后	Eta	Pr > Eta
单均值	14	0.1978	0.2724
趋势	14	0.1518	0.0456

分为单均值和趋势两种检验模型列示，单均值模型检验不能拒绝原假设，带趋势模型检验拒绝原假设。结果窗口中一同给出的还有检验模型的估计结果等相关信息。

第二个model语句的检验结果（“模型2”）为

KPSS 平稳性检验			
类型	滞后	Eta	Pr > Eta
零均值	14	4.7788	0.0003
单均值	14	0.1978	0.2724
趋势	14	0.1518	0.0456

由于添加了选项noint，检验模型有三类，包括“零均值”检验模型。

model语句中不能用close(1)表示close的一阶差分。要检验close一阶差分的平稳性，需先用data步生成变量一阶差分变量close_1。程序7-8采用ERS方法检验IBM股票close一阶差分的平稳性。

程序7-8

```
data test(where=(stock="IBM"));
set sashelp.stocks;
dclose_1=dif(close);
run;
proc autoreg;
model dclose_1=/noint stationarity=(ers=5);
run;
```

用ers=5将最大滞后阶数设定为5。检验结果为

ERS 单位根检验						
类型	滞后	方差	PT	Pr < PT	DF-GLS	Pr < DFGLS
单均值	3	211.8126	11.9198	0.3937	-0.4202	<.0001
趋势	3	211.8126	10.8138	1.0000	-1.6891	0.5126

最优滞后阶数选择结果为3，检验的概率值小于0.01，强烈拒绝原假设，close一阶差分后不存在单位根，为平稳序列。

7.2.4 自回归模型设定和估计

用model语句可以设定和估计自回归模型，分没有回归自变量和有回归自变量两种方法。下面以变量y为例进行说明

没有回归自变量形式的模型设定

model语句的语法格式为

格式1：model y=/nlag=n;

格式2：model y=/nlag=(n1,n2,…,nk);

两种格式中nlag=都是设定（6）式中模型误差项u_t的自回归模型阶数。格式1设定的是（6）式中的模型，可转换为（7）式中y_t的自回归模型。格式2设定的u_t的自回归模型中只有指定滞后阶数n1,n2,…,nk的变量出现，转换为y_t自回归模型为

$$y_t+\varphi_{n_1}y_{t-n_1}+\varphi_{n_2}y_{t-n_2}+\cdots+\varphi_{n_k}y_{t-n_k}=c\left(1+\sum_{i=1}^{k}\varphi_{n_i}\right)+\epsilon_t$$

选项nlag=n等价于nlag=(1,2,… ,n)。

需要注意，没有回归自变量的model语句中，模型变量后面的等号“=”不能省略。

以滞后变量作为回归自变量的模型设定

将变量y的各阶滞后变量y_{-1}、y_{-2}、…、y_{-m}，将作为回归自变量，即

model $y=y_{-1}$ y_{-2} … y_{-m}/options;

options中不再出现nlag=选项，proc Autoreg不再估计误差项自回归模型。此时需要用data步在输入数据集中生成因变量的各阶滞后变量。

两种方法得出的回归系数ϕ_i估计符号相反。由于估计方法不一样，估计值可能存在差异。程序7-9对IBM股票的close的一阶差分dclose建立2阶自回归模型

程序7-9

```
data test;
set sashelp.stocks;
where stock='IBM';
dclose=dif(close);
dclose1=lag(dclose);
dclose2=lag2(dclose);
run;
proc autoreg;
model dclose=/nlag=2;
model dclose=dclose1 dclose2;
run;
```

第一个model语句没有回归自变量，通过误差项的AR(2)模型间接设定变量dclose的AR(2)模型，proc autoreg将dclose作为时间序列变量处理，输出结果中包括dclose和回归残差的自相关图形。参数估计分三步进行，首先对回归函数中的回归系数（这里只有截距项intercept）采用OLS进行初步估计，“参数估计”显示为

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	0.2438	0.7881	0.31	0.7574

第二步计算第一步回归残差的样本自相关函数，带入误差项u_t的Yuler-Walker方程解出自回归系数。“自回归参数的估计”显示为

自回归参数的估计			
滞后	系数	标准误差	t 值
1	0.206419	0.065907	3.13
2	0.072683	0.065907	1.10

最后一步是用估计出的u_t的自回归系数计算误差项向量的方差协方差矩阵，再对回归模型系数（这里是常数项intercept）进行广义最小二乘估计，最终的估计结果以“参数估计”显示为

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	0.2331	0.6063	0.38	0.7011

估计结果与初始估计相差不大。

第二个model语句用dclose的滞后作为回归解释变量，不再设定误差项的自回归模型，proc autoreg 将dclose模型作为一般线性回归模型处理，采用OLS估计模型参数，“参数估计”显示为

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	0.3012	0.7807	0.39	0.7000
dclose1	1	-0.2055	0.0662	-3.10	0.0021
dclose2	1	-0.0731	0.0663	-1.10	0.2714

比较发现，两种方法得出的自回归系数估计符号相反，绝对值近似相同。截距项含义不同，两种估计结果不具有可比性。

7.2.5 时间序列回归模型估计

将（4）和（5）放在一起，得出具有自回归误差项的时间序列回归模型

$$\begin{aligned} y_t &= c + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + u_t \\ u_t &= \epsilon_t - \varphi_1 u_{t-1} - \varphi_2 u_{t-2} - \cdots - \varphi_m u_{t-m} \end{aligned} \tag{15}$$

时间序列回归与横截面回归的最大不同在于误差项的序列相关。序列相关性不影响OLS估计的一致性和渐进正态性，但参数估计渐进分布的方差（从而标准误）计算公式需要修正。序列相关的处理方式分为两种：第一种方法采用OLS估计回归系数并采用异方差和自相关一致（HAC：Heteroskedesticity-Autoregession-Consistent）的方法计算参数估计标准误，据此构造显著性检验的t-统计量或者z-统计量。第二种方法对误差项的序列相关建立模型，将其参数连同回归系数一同估计。

OLS估计+Newey-West标准误

采用第一种方法时，在model语句中添加选择covest=neweywest(options)选项，SAS关键词covest是协方差估计（covariance estimation）的缩写，具体语法格式为

```
model  y=x1  x2  …  xk/covest=neweywest;
```

贝塔系数：数据集sh_trt_r中变量r_m和r_trt分别为上证指数（000001）和上市公司同仁堂（60085）的股票2016年1月4日到2020年12月31日的日收益率。程序7-11将变量r_trt对r_m进行回归，计算同仁堂日收益的贝塔系数。这里略去对r_trt和r_m平稳性

检验环节，默认两个变量为平稳时间序列。采用OLS估计回归系数，并用HAC计算参数估计的方差和标准误。

程序7-11

```
proc autoreg data=sh_trt_r;
model r_trt=r_m/covest=neweywest;
run;
```

回归结果为

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	-0.000392	0.000393	-1.00	0.3189
r_m	1	0.8383	0.0771	10.87	<.0001

表明同仁堂这一时期的贝塔系数为0.8383。

newey-West方法只是多种HAC计算方法中的一种，通过covest=选项可以更为灵活地选择HAC计算方法，细节可参考SAS帮助。

带自回归误差项的回归模型

采用第二种方法时，在model语句中添加选项nlag=设定误差项自回归模型，即

model y=x1 x2 … xk/nlag=n;

程序7-12将变量r_trt对r_m进行回归，对误差项建立自回归模型以捕捉自相关性。S估计回归系数，并用HAC计算参数估计的方差和标准误。

程序7-12

```
proc autoreg data=sh_trt_r;
model r_trt=r_m/nlag=5;
run;
```

输出结果分为两部分，第一部分是采用OLS估计的结果，同程序7-11结果相同。第二部分中，首先输出的是用模型OLS估计残差计算的残差样本自相关函数带入u_t的Yuler-Walker方程解出的自回归系数，

自回归参数的估计			
滞后	系数	标准误差	t 值
1	0.051999	0.028788	1.81
2	0.051014	0.028810	1.77
3	-0.052863	0.028807	-1.84
4	0.034282	0.028810	1.19
5	0.089349	0.028788	3.10

然后输出的是用误差自回归系数计算误差项协方差矩阵后模型的GLS估计

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	-0.000393	0.000368	-1.07	0.2854
r_m	1	0.8425	0.0370	22.79	<.0001

尽管最终结果和初始结果差别不大，但最终结果0.8425更为准确可靠。

7.2.6 ARCH检验和GARCH模型估计

条件异方差是金融时间序列数据中普遍存在的现象，GARCH类模型为条件异方差建模和预测提供了灵活多样的工具。建立GARCH模型之前的ARCH检验，为是否需要建立GARCH模型提供实证依据。

ARCH检验

ARCH检验又称为ARCH效应检验（ARCH Effect Test）。ARCH模型来源是误差项平方的AR(m)模型

$$u_t^2=\omega+\alpha_1 u_{t-1}^2+\alpha_2 u_{t-2}^2+\cdots+\alpha_m u_{t-m}^2+e \tag{16}$$

ARCH检验的原假设是u_t^2不存在自相关（ARCH效应），即

$$H_0:\ \alpha_1=\alpha_2=\cdots=\alpha_m=0;\ \ H_0:\exists\quad \alpha_i\neq 0, i=1,2,\dots,n$$

由于u_t^2不可观测，采用模型回归残差平方$\hat{u}^2$代替进行回归

$$\hat{u}_t^2=\omega+\alpha_1\hat{u}_{t-1}^2+\alpha_2\hat{u}_{t-2}^2+\cdots+\alpha_m\hat{u}_{t-m}^2+e \tag{17}$$

回归后检验原假设H_0: $\alpha_1=\alpha_2=\cdots=\alpha_m=0$。

由此看出，ARCH检验依赖（4）式中的均值模型设定。

proc autoreg语句model通过选项ARTCH=提供三种检验方法。语句格式为

```
model y=x1 x2 … xk/ ARCHTEST=(options);
```

options确定检验方法，分别为

QLM：Engle（1982）的LM（拉格朗日乘子）ARCH检验方法；

LK：Lee和King（1993）的ARCH检验方法；

WL：Wong和Li（1995）的ARCH检验方法；

ALL：以上三种方法。

程序7-13检验sh_trt_r中的序列r_m是否存在ARCH效应

程序7-13

```
proc autoreg data=sh_trt_r;
model r_m=/nlag=2 ARCHTEST=(LK);
run;
```

显示两种检验结果，两种检验都采用LK统计量：

基于 OLS 残差的 ARCH 扰动检验			基于残差的 ARCH 扰动检验		
阶数	LK	Pr > |LK|	阶数	LK	Pr > |LK|
1	4.7464	<.0001	1	3.7805	0.0002
2	11.0672	<.0001	2	9.3700	<.0001
3	11.2841	<.0001	3	9.9706	<.0001
4	12.7126	<.0001	4	11.6033	<.0001
5	10.3328	<.0001	5	9.7338	<.0001
6	11.4243	<.0001	6	10.7663	<.0001
7	12.0956	<.0001	7	11.6708	<.0001
8	12.2670	<.0001	8	11.9861	<.0001
9	11.5098	<.0001	9	11.4125	<.0001
10	12.5680	<.0001	10	12.5727	<.0001
11	12.2862	<.0001	11	12.1090	<.0001
12	11.9991	<.0001	12	11.7729	<.0001

“基于OLS残差的ARCH扰动检验”，是指采用不考虑误差项u_t中的自相关，对模型进行OLS估计得出的残差进行的检验，“基于残差的ARCH扰动检验”则是基于尤勒-沃尔克方法估计模型得出的异方差的检验。从结果看出，两种方法检验结果都表明2016年1月4日到2020年12月31日间上证指数日收益存在明显的ARCH效应。

GARCH模型估计

一旦检验出ARCH效应，就应该对模型误差项建立GARCH模型。model语句中选项GARCH=对各种GARCH模型进行设定。语句格式为

model y=x1 x2 … xk/nlag=n GARCH=(options)

GARCH=(options)中的options内容和功能为

Q=q：设定GARCH模型中ARCH项的阶数；

P=p：设定GARCH模型中GARCH项阶数。

p和q的含义参考（8）式。

type=value：设定GARCH模型类型，value取值及对应的GARCH类型为

EGARCH：指数GARCH；

IGARCH：单整GARCH；

PGARCH：幂GARCH；

QGARCH：二次GARCH；

TGARCH|GJR：门限GARCH。

value取其它值可以对GARCH模型进行更细致的设定，包括

stationarity:将GARCH模型约束为平稳模型，即将GARCH项系数之和约束为小于1；

用选项dist=可以选择模型误差项分布类型，选项格式为

dist=t：误差项分布选择为t分布；

dist=normal：误差项分布选择为正态分布。

默认为正态分布。但有些模型不能选择t分布，例如EGARCH模型。具体细节可查看SAS帮助。

程序7-14对股票Intel的收盘价close建立AR(5)模型，对误差项建立EGARCH(1,1)模型。

程序7-14

```
proc autoreg data=sashelp.stocks(where=(stock='IBM'));
model close=/nlag=5 GARCH=(p=1,q=1,type=EGARCH);
run;
```

输出结果

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	89.9415	6.8193	13.19	<.0001
AR1	1	-0.7512	0.0808	-9.29	<.0001
AR2	1	-0.1640	0.0954	-1.72	0.0856
AR3	1	-0.1634	0.0750	-2.18	0.0293
AR4	1	0.1085	0.0481	2.26	0.0241
AR5	1	0.0261	0.0363	0.72	0.4720
EARCH0	1	2.1716	0.3334	6.51	<.0001
EARCH1	1	1.3850	0.1447	9.57	<.0001
EGARCH1	1	0.5600	0.0645	8.68	<.0001
THETA	1	-0.3655	0.0809	-4.52	<.0001

第一部分为AR(5)均值模型的常数项和误差项自回归系数估计值，第二部分为GARCH模型参数估计结果。其中EARCH0为常数项，EARCH1为ARCH项系数，EGARCH1为GARCH项系数，THETA为表示杠杆效应的参数。

SAS中的EGARCH(1,1)模型的形式与R.Tsay(2012)相同，即

$$\begin{aligned} \ln h_t &= \omega + \alpha_1 g(z_{t-1}) + \gamma_1 \ln h_{t-1} \\ g(z_t) &= \theta z_t + \gamma[|z_t| - E|z_t|] \\ z_t &= \epsilon_t/\sqrt{h_t} \end{aligned} \tag{18}$$

SAS将参数γ的值设为1，由此得出

$$g(z_t) = \begin{cases} (\theta+1)z_t - E|Z_t|, & \text{if } z \geq 0; \\ (\theta-1)z_t - E|Z_t|, & \text{if } z < 0; \end{cases}$$

参数θ捕捉了杠杆效应，为杠杆效应参数。

输出结果中的EARCH0、EARCH1、EGARCH1和THETA分别对应（18）式中参数ω、α_1、γ_1和θ 。THETA估计值为17，但t检验p-值0.45远大于0.05，可以认为$\theta=0$，不存在杠杆效应。

GARCH模型中引入外生变量

为研究某些变量对模型误差项条件方差的影响，需要在GARCH模型中加入外生变量。带外生变量w_{1t}, w_{2t}的GARCH模型（以GARCH(1,1)为例）具有如下形式

$$\begin{aligned} \epsilon_t &= \sqrt{h_t} e_t \\ h_t &= \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 h_{t-1} + \delta_1 w_{1t} + \delta_2 w_{2t} \end{aligned} \tag{19}$$

类似地可以添加多个外生变量。带外生变量w_{t1}, w_{t2}的EGARCH模型（以EGARCH(1,1)为例）具有如下形式

$$\begin{aligned} \ln h_t &= \omega + \alpha_1 g(z_t) + \gamma_1 \ln h_{t-1} + \delta_1 w_{1t} + \delta_2 w_{2t} \\ g(z_t) &= \theta z_t + \gamma(|z_t| - E|z_t|) \\ z_t &= \epsilon_t/\sqrt{h_t} \end{aligned} \tag{20}$$

类似地在其它GARCH模型中添加外生变量。

语句hetero能够为model语句选项GARCH=设定的GARCH模型中添加外生变量，语句语法格式为

```
hetero varlist/<COEF=value>;
```

varlist为外生变量名列表，选项COEF对GARCH模型中外生变量系数δ_1,δ_2的估计进行约束，value取值为

- nonneg：将δ_1,δ_2的值约束为非负；
- unit：将δ_1,δ_2的值取1；
- zero：将δ_1,δ_2值取0；

默认为nonneg。将η_1,η_2取值为1或者0，便于和无约束的估计结果进行比较，例如用对数似然值之差进行极大似然比检验等。

程序7-15对sashelp.stocks中股票Intel的交易量变量volume取对数，然后用proc autoreg中两个model语句分别对close建立AR(3)+GARCH(1,1)模型和AR(3)+EGARCH(1,1)模型，并在GARCH模型和EGARCH模型中引入外生变量lnvol。

程序7-15

```
data test(where=(stock='Intel'));
set sashelp.stocks;
lnvol=log(volume);
proc autoreg;
model close=/nlag=3 garch=(p=1,q=1);
hetero lnvol;
model close=/nlag=3 garch=(p=1,q=1,type=egarch);
hetero lnvol;
run;
```

第一个model语句采用GARCH模型，并用hetero语句在GARCH模型中引入外生变量lnvol，估计结果为

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	36.8964	9.5944	3.85	0.0001
AR1	1	-0.9430	0.0808	-11.67	<.0001
AR2	1	-0.1738	0.0774	-2.24	0.0248
AR3	1	0.1708	0.0474	3.60	0.0003
ARCH0	1	1.0537E-8	0	Infty	<.0001
ARCH1	1	0	0	.	.
GARCH1	1	0	0	.	.
HET1	1	6.6851	0.2631	25.40	<.0001

对数似然函数的最优化不理想，ARCH系数和GARCH系数都取到了下界0，标准误也是0，导致无法计算对应的t值还概率值，估计结果不具参考价值。

第二个model语句采用EGARCH拟合数据中的条件异方差，用hetero语句在EGARCH模型中引入外生变量lnvol，参数估计结果为

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	49.4290	2.2247	22.22	<.0001
AR1	1	-0.7996	0.0252	-31.71	<.0001
AR2	1	-0.2623	0.0294	-8.93	<.0001
AR3	1	0.2065	0.0202	10.21	<.0001
EARCH0	1	2.8290	2.7680	1.02	0.3068
EARCH1	1	-0.9341	0.1140	-8.19	<.0001
EGARCH1	1	0.1676	0.0528	3.18	0.0015
THETA	1	-0.9998	0.1154	-8.66	<.0001
HET1	1	0.0465	0.1507	0.31	0.7577

从中看出，EARCH项系数、EGARCH项系数和杠杆效应系数THETA均显著不为0，HET1系数不显著，表明外生变量lnvol对波动没有显著影响。

除GARCH类模型外，hetero语句也可以设定其它函数形式的误差项条件异方差，且可以对是否存在给定形式的异方差进行检验，具体内容可参考SAS帮助。

两个不同时间序列变量的回归模型也可以设定GARCH模型。程序7-12（续）在程序7-12设定的模型中添加GARCH(1,1)模型

程序7-12（续）

```
proc autoreg data=sh_trt_r;
model r_trt=r_m/nlag=5 garch=(p=1,q=1);
run;
```

参数估计结果输出为

参数估计					
变量	自由度	估计	标准误差	t 值	近似 Pr > |t|
Intercept	1	-0.000572	0.000386	-1.48	0.1383
r_m	1	0.7630	0.0264	28.92	<.0001
AR1	1	0.0528	0.0374	1.41	0.1586
AR2	1	0.0715	0.0382	1.87	0.0613
AR3	1	-0.0469	0.0372	-1.26	0.2071
AR4	1	0.0372	0.0346	1.08	0.2822
AR5	1	0.0620	0.0339	1.83	0.0679
ARCH0	1	0.0000168	2.9375E-6	5.73	<.0001
ARCH1	1	0.0920	0.0159	5.79	<.0001
GARCH1	1	0.8328	0.0255	32.71	<.0001

注：模型估计控制

在model语句中添加选项，可以对模型估计进行控制，包括估计方法选择、优化方法选择、收敛标准设定等。选项method=value用以确定估计方法，value取值ML、ULS、YW

和ITYW，分别对应极大似然、无条件最小二乘、尤勒-沃尔克和迭代尤勒-沃尔克方法。有GARCH=选项或者HETERO语句时，默认ML方法；没有GARCH=选项和HETERO语句，也没有LAGDEP选项，有NLAG选项，默认YW方法；没有GARCH=选项和HETERO语句，也没有NLAG选项时，采用OLS估计模型，method=选项被忽略。有GARCH=选项时需要用数值方法寻找对数似然函数的极值点，选项optmethod=value确定数值优化方法，value取值QN和TR分别对应拟牛顿（quasi-Newton）算法和信赖域（trust region）算法。

7.2.5 变点检验和变结构检验

时间序列往往在某个或者某些时点处发生结构性变化，所谓结构性变化是指不同时点之间数据所服从的模型发生不同，这里仅限于回归系数的变化。变结构检验（structure change test）分为Chow单点变结构检验和BP（Bai-Perron）多点变结构检验。

Chow检验

Chow检验是由邹至庄提出的检验方法。选定可能存在结构变化的时点T_1（$0 < T_1 < T$），（称为断点break point）将样本从T_1处分为两组，设定三个回归模型

$$\text{Model I: } y_t = c^{(1)} + \beta_1^{(1)} x_{1t} + \beta_2^{(1)} x_{2t} + \cdots + \beta_k^{(1)} x_{kt} + u_t^{(1)}, \quad 0 \le t \le T_1$$
$$\text{Model II: } y_t = c^{(2)} + \beta_1^{(2)} x_{1t} + \beta_2^{(2)} x_{2t} + \cdots + \beta_k^{(2)} x_{kt} + u_t^{(2)}, \quad T_1 \le t \le T$$
$$\text{Model III: } y_t = c + \beta_1^{(2)} x_{1t} + \beta_2^{(2)} x_{2t} + \cdots + \beta_k^{(2)} x_{kt} + u_t^{(2)}, \quad 0 \le t \le T$$

待检验假设为

$$H_0: \beta_i^{(1)} = \beta_i^{(2)}, i = 1,2,\ldots,k; \qquad H_1: \exists\ i \in \{1,2,\ldots,k\} \quad s.t. \beta_i^{(1)} \ne \beta_i^{(2)}$$

Chow检验采用参数约束F检验。对原假设成立时的模型进行回归，此时两个ModelI和ModelII模型合并为ModelIII，回归采用所有样本，回归残差平方和（即约束残差平方和）为SSR_R。分别对ModelI和ModelII进行回归，得出残差平方和SSR_I和SSR_{II}，相加得出无约束残差平方和$SSR_{UR} = SSR_I + SSR_{II}$。构造统计量F

$$F = \frac{(SSR_R - SSR_{UR})/k}{SSR_{UR}/(T - 2k)} \tag{21}$$

在原假设下F服从自由度为(k,T-2k)的F分布。

在moldel语句中添加选项Chow=可进行Chow检验。语法格式为

```
model y=x1 x2 … xk/chow=(n1 n2 … np);
```

要求在观测个数n1、n2、…和np处进行结构变点（断点）检验。n1是从因变量首个非缺失值算起的第n1个观测处。如果因变量的前10个观测为缺失值，chow=(20)要求对数据集的第30个观测处进行变点检验。

程序7-14对数据集sh_trt_r中变量r_m建立AR(3)+GARCH(1,1)模型，并对第15和第345观测处进行结构变点Chow检验。

程序7-14

```
proc autoreg data=sh_trt_r;
model r_m=/nlag=3 garch=(p=1,q=1) chow=(15 345);
run;
```

“结构更改检验”栏目显示

结构更改检验					
检验	断点	分子自由度	分母自由度	F 值	Pr > F
Chow	15	1	1202	8.13	0.0044
Chow	345	1	1202	0.21	0.6493

第15观测处断点检验拒绝原假设，认为发生结构变化。第345处Chow检验不能拒绝原假设，可以认为没有结构变化。

注意程序7-14设定的模型没有解释变量，实际上相当于检测模型常数项在给定时点处是否发生变化，并没有检验自回归系数是否发生变化。要检验包括自回归系数在内的所有回归系数是否发生变化，需在model语句中以因变量的滞后变量作为解释变量。程序7-15首先生成r_m的1至3阶滞后，用model语句显式建立r_m的AR(3)模型，并进行结构变点检验。

程序7-15

```
data test;
set sh_trt_r;
r_m_1=lag(r_m);
r_m_2=lag2(r_m);
r_m_3=lag3(r_m);
run;
proc autoreg;
model r_m=r_m_1 r_m_2 r_m_3/garch=(p=1,q=1) chow=(15 345);
run;
```

检验结果

结构更改检验					
检验	断点	分子自由度	分母自由度	F 值	Pr > F
Chow	15	4	1154	7.15	<.0001
Chow	345	4	1154	3.08	0.0154

显示第15和第345个观测处发生了结构变化。

BP检验

BP检验是白聚山和Perron给出的多变点结构变化检验方法，以线性回归模型为基础。BP检验将模型回归系数分为有结构变化和无结构变化两部分，设在时点$T_1,T_2,\dots,T_m$处可能存在结构变化，而在区间$(T_i,T_{i+1}], i=0,1,\dots,m-1$ 内不再有结构变化，则时间序列回归模型可以分段写成

$$\begin{array}{ll} y_t=\boldsymbol{x}_t'\boldsymbol{\beta}+\boldsymbol{z}_t'\boldsymbol{\delta}_1+u_{1t}, & t\in(1,T_1] \\ y_t=\boldsymbol{x}_t'\boldsymbol{\beta}+\boldsymbol{z}_t'\boldsymbol{\delta}_2+u_{2t}, & t\in(T_1,T_2] \\ \dots\quad\dots\quad\dots\quad\dots & \dots \\ y_t=\boldsymbol{x}_t'\boldsymbol{\beta}+\boldsymbol{z}_t'\boldsymbol{\delta}_m+u_{mt}, & t\in(T_m,T] \end{array} \tag{22}$$

$\boldsymbol{x}_t=(1,x_{1t},x_{2t},\dots,x_{pt})',\boldsymbol{\beta}=(c,\beta_1,\dots,\beta_p)'$，$\boldsymbol{x}_t'\boldsymbol{\beta}$为没有结构变化的部分，$\boldsymbol{\beta}$为不发生结构变化回归系数向量。$\boldsymbol{z}_t=(z_{1t},z_{2t},\dots,z_{qt})',\boldsymbol{\delta}_i=(\delta_{1i},\delta_{i2},\dots,\delta_{iq})',i=1,2,\dots,m$为可能存在结构变化的解释变量及其回归系数。$p=0$时解释变量中（除常数项外）不存在没有结构变化的部分。

对给定m个可能变点（断点）$(T_1,T_2,\dots,T_m)$，BP方法首先求解最小二乘问题

$$(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\delta}})=\arg\min_{\boldsymbol{\beta},\boldsymbol{\delta}} SSR(\boldsymbol{\beta},\boldsymbol{\delta}|T_1,T_2,\dots,T_m)$$

得出$\boldsymbol{\beta}$和$\boldsymbol{\delta}$的估计，其中$\boldsymbol{\delta}=(\boldsymbol{\delta}_1',\boldsymbol{\delta}_2',\dots,\boldsymbol{\delta}_m')'$，$SSR(\boldsymbol{\beta},\boldsymbol{\delta}|T_1,T_2,\dots,T_m)$为残差平方和

$$SSR(\boldsymbol{\beta},\boldsymbol{\delta}|T_1,T_2,\dots,T_m)$$
$$=\sum_{t=1}^{T_1}(y_t-\boldsymbol{x}_t'\boldsymbol{\beta}-\boldsymbol{z}_t'\boldsymbol{\delta}_1)^2+\sum_{i=1}^{m-1}\sum_{t=T_i+1}^{T_{i+1}}(y_t-\boldsymbol{x}_t'\boldsymbol{\beta}-\boldsymbol{z}_t'\boldsymbol{\delta}_i)^2+\sum_{t=T_m+1}^{T}(y_t-\boldsymbol{x}_t'\boldsymbol{\beta}-\boldsymbol{z}_t'\boldsymbol{\delta}_m)^2$$

。将参数估计$\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\delta}}$带入得出残差平方和$SSR(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\delta}}|T_1,T_2,\dots,T_m)$

选择$t=1,2,\dots,T$中任意m个时点进行如上回归，并求解如下极小化问题的解

$$(\hat{T}_1,\hat{T}_2,\dots,\hat{T}_m)=\arg\min_{\substack{(T_1,T_2,\dots,T_m)\\ \subseteq(1,2,\dots,T)}} SSR(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\delta}}|T_1,T_2,\dots,T_m) \quad (23)$$

得出最有可能的断点。

Bai和Perron采用基于动态规划算法计算估计值$\hat{T}_1,\hat{T}_2,\dots,\hat{T}_m$，并给出$\hat{T}_1,\hat{T}_2,\dots,\hat{T}_m$的渐近分布，据此构造三种断点检验统计量，给出三种检验方法。设可能的结构断点数为k。

- 第一种：SupF

用于检验有确定个数断点的模型。

原假设H_0：不存在断点；备择假设H_1：存在k个断点

等价地

$$H_0:\delta_1=\cdots=\delta_k=0；\ H_1:\exists\ \ i\in\{1,2,\dots,k\},\delta_i\neq 0\ \ )$$

检验统计量$supF(k)$为Wald检验统计量，是约束条件$R\hat{\boldsymbol{\delta}}$用协方差矩阵标准化的二次型，其定义为

$$supF(k)=\frac{1}{T}\left(\frac{T-(k+1)q-p}{kq}\right)(R\hat{\boldsymbol{\delta}})'\left(\boldsymbol{Var}(R\hat{\boldsymbol{\delta}})\right)^{-1}(R\hat{\boldsymbol{\delta}}) \quad (24)$$

其中

$$R\hat{\boldsymbol{\delta}}=(\hat{\boldsymbol{\delta}}_1'-\hat{\boldsymbol{\delta}}_2',\cdots,\hat{\boldsymbol{\delta}}_{k-1}'-\hat{\boldsymbol{\delta}}_k')'$$

$\hat{\boldsymbol{\delta}}$是（23）式中的估计值，即残差平方和最小的$\hat{T}_1,\hat{T}_2,\dots,\hat{T}_m$对应模型中$\boldsymbol{\delta}$的估计值。从（24）看出，supF是断点隔出的不同时间区间内回归系数估计值的差值加权平方和，值越大表明断点发生结构变化的可能性越大。

- 第二种：UDmaxF和WDmaxF

用于检验断点个数不确定，但不超过给定个数的模型。

原假设H_0：不存在断点；备择假设H_1：断点个数不超过k

检验统计量有两个：UDmaxF和WDmaxF。UDmaxF定义为

$$UDmaxF(M)=\max_{k\leq M} supF(k) \quad (25)$$

其对每个小于M的断点个数k计算supF(k)，然后取最大值。取最大值时对每个supF(k)取相同的权重1（Unite Weight）。WDmaxF定义为

$$WDmaxF(M,\alpha)=\max_{k\leq M}\frac{c_\alpha(1)}{c_\alpha(k)}supF(k) \quad (26)$$

$c_\alpha(k)$为显著水平α下检验统计量$supF(k)$的临界值，α可以取0.1、0.05、0.025和0.01四个值。取最大值时，WDmaxF对每个supF(k)取不同的权重（weight）。

- 第三种：SupF(l+1|l)

是一种序贯（sequential test）检验，总共有k+1个检验，第p个检验为。

原假设H_0：有l个断点；备择假设H_1：有l+1个断点

对给定的l个断点$(\hat{T}_1,\hat{T}_2,\dots,\hat{T}_l)$，寻找第$l+1$个断点$\hat{T}_{l+1}$极小化残差平方和$SSR(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\delta}}|\hat{T}_1,\hat{T}_2,\dots,\hat{T}_l;\hat{T}_{l+1})$，据此构造检验统计量

$$supF(l+1|l)=$$
$$(T-(l+1)q-p)\frac{SSR(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\delta}}|\hat{T}_1,\hat{T}_2,\dots,\hat{T}_l)-SSR(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\delta}}|\hat{T}_1,\hat{T}_2,\dots,\hat{T}_l;\hat{T}_{l+1})}{SSR(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\delta}}|\hat{T}_1,\hat{T}_2,\dots,\hat{T}_l)} \quad (27)$$

统计量临界值通过模拟方法得出。

在moldel语句中添加选项BP=可进行BP检验。语法格式为

model y=x1 x2 … xk/BP=(options);

选项options之间用逗号隔开。options包括

- EPS=a：设定两个相邻变点之间的最短间隔，默认值为0.05。
- M=n：设定变点个数，默认值为5。给定值n后，model语句将采用三种方法进行检验：

  （1）采用第一种方法中检验统计量$supF(k)$对

  H_0 ：没有变点；H_1:有i个变点

  进行检验，$i=1,2,\dots,n$。共检验n次。

  （2）采用第二种方法的两种检验统计量$UDmaxF(M)$和$WDmaxF(M,\alpha)$对假设

  H_0 ：没有变点；H_1:变点个数小于n

  进行检验。

  （3）采用第三种方法中检验统计量$supF(l+1|l)$对假设

  H_0 ：有l个断点；H_1:有l+1个断点

  进行检验，l=0,1,2,…,n。检验n+1次。
- P=k：设定模型中前k个解释变量的回归系数没有变点，默认值为0。
- 其它选项：参数估计协方差矩阵估计方法、核估计中核的选择、(22)式中$u_{1t},u_{2t},\dots,u_{mt}$是否同分布、是否同方差等等。详情可参考SAS帮助。

程序7-16对sh_trt_r中股票同仁堂日收益r_trt对上证指数日收益r_m的线性回归模型进行结构变点检验，检验同仁堂股票的贝塔系数以及常数项是否有结构变点，用bp=(m=3)选项指定检验3个变点。

程序7-16

```
proc autoreg data=sh_trt_r;
model r_trt=r_m/nlag=3 garch=(p=1,q=1) bp=(m=);
run;
```

输出结果中分别以“Bai-Perron多结构变化检验-中断日期”一栏给出的是k=1,2,3时按（23）式计算得出的断点对应的观测号，给定不同的断点数得出的断点位置不同。例如，断点个数为1时得出的断点是第230个观测处，而断点个数为2时得出的断点分别是503和577个观测。“Bai-Perron多结构变化检验-supF检验”一栏给出原假设为没有断点，备择假设为有1个、2个和3个断点的supF检验，检验结果均拒绝原假设。“Bai-Perron多结构变化检验-UDmaxF检验”一栏给出的对原假设为没有断点，备择假设为断点个数小于3的UDmaxF检验，检验结果拒绝原假设。

Bai-Perron 多结构变化检验

中断日期

中断数	中断	95% 置信限	
1	230	154	306
2	503	475	531
	577	510	644
3	230	118	342
	976	911	1041
	1086	1049	1123

Bai-Perron 多结构变化检验

supF 检验

中断数	supF	Pr > supF
1	27.26	0.0003
2	23.79	<.0001
3	14.08	0.0012

Bai-Perron 多结构变化检验

UDmaxF 检验

中断数	UDmaxF	Pr > UDmaxF
3	27.2623042	0.0003

“Bai-Perron多结构变化检验-WDmaxF检验”一栏给出的对原假设为没有断点，备择假设为断点个数小于3的WDmaxF检验，其中统计量的4个值分别对应α取0.1、0.05、0.025和0.01。检验结果均拒绝原假设。拒绝原假设。“Bai-Perron多结构变化检验-supF(l+1|l)检验”一栏给出的是supF(l+1|l)检验结果，共有4个检验，其中原假设为0个断点、备择假设为1个断点的检验中，原假设被拒绝，新增的断点为第230个观测。其余

三个检验的不能拒绝原假设，但仍然给出了新的可能断点。

Bai-Perron 多结构变化检验			
WDmaxF 检验			
中断数	Alpha	WDmaxF	Pr > WDmaxF
3	0.1000	27.2623042	0.0002
	0.0500	27.2623042	0.0003
	0.0250	27.2623042	0.0001
	0.0100	28.4493698	0.0001

Bai-Perron 多结构变化检验			
supF(l+1|l) 检验			
l	新断点	supF(l+1|l)	Pr > supF(l+1|l)
0	230	27.2623042	0.0002
1	960	8.12773747	0.6491
2	976	11.6077714	0.2197
3	561	13.5255216	0.1044

从各种检测结果综合判断，模型有1个断点，位于第230个观测处。

Chow检验需要指定可能发生结构变化的断点位置，且原假设是模型回归系数都没有结构变化。BP检验不需要事先指定可能结构变化的断点，但要给出要检验的断点个数。BP检验可以设定无结构变化的回归系数。无论是Chow检验还是BP检验，只能检验回归系数的断点，不能检验GARCH模型断点，且只能检验断点，没有针对断点的模型修正功能。

7.2.6 参数约束及检验

实际应用中常需要根据情况对模型参数施加必要的限制。语句restrict对模型中参数施加线性等式约束。语法格式为

restrict eq1,eq2,…,eqn;

eq$i$是由参数形成的线性方程。参数的表示方法为

- 回归系数：由对应的解释变量名表示，常数项用intercept表示。
- 误差项自回归系数$\varphi_i, i=1,2,\dots,m$：用自动变量_A_i表示，$i=1,2,\dots,m$，m为选项nlag=m指定的滞后阶数。
- GARCH模型参数：ARCH项模型参数α_i用自动变量_AH_i表示，$i=0,1,2,\dots,m$，常数项ω用_AH_0表示；GARCH项模型参数γ_j用自动变量_GH_j表示，$j=1,2,\dots,n$。m和n为选项GRACH=(q=m,p=n)设定的GARCH模型阶数。
- EGARCH模型参数：ARCH项和GARCH项参数用_AH_i和_GH_j表示，杠杆效应参数θ用自动变量_THETA_表示。
- GARCH-M模型参数：出现在回归模型中的GARCH项参数用自动变量_DELTA_表示。
- GARCH模型中外生变量系数：用_HET_i表示语句hetero中设定的GARCH模型中外生变量的系数。
- 其它GARCH模型中参数表示可参考SAS帮助。

程序7-17重新估计程序7-16中的模型，将模型常数项约束为0，GARCH模型中ARCH项和GARCH项系数之和约束为0.99。

程序7-17

```
proc autoreg;
model r_m=r_m_1 r_m_2 r_m_3/garch=(p=1,q=1);
restrict intercept=0,_ah_1+_gh_1=;
run;
```

采用test语句可以检验模型参数是否满足给某种约束。语句格式为

test eq1,eq2,…,eqn/option;

eq$i$为参数约束方程，参数表达方法与restrict语句相同。option设定检验方法，可以取F、Wald、LM、LR和ALL，对应F检验、Wald检验、LM（拉格朗日乘子）检验、LR（极大似然比）检验和所有检验。程序7-18对自回归系数和GARCH模型参数约束进行LR检验。

程序7-18

```
proc autoreg data=test;
```

```
model r_m=r_m_1 r_m_2 r_m_3/garch=(p=1,q=1);
test intercept=0,_ah_1+_gh_1=0.99/LR;
run;
```

检验结果显示为

检验结果				
检验	类型	统计量	Pr ＞ 卡方	标签
Test 1	L.R.	1.11	0.5729	intercept=0,_ah_1+_gh_1=0.99

不能拒绝原假设，可以认为参数约束成立。

注意，同一test语句中多个约束方程是联合检验，只检验一次。不同test语句的约束方程是各自检验。restrict语句和test语句中只能设定线性等式约束，不能设定非线性不等式约束。

7.2.7 模型检测

模型检测（model checking）通过检验回归残差的独立性，考察模型滞后阶数设定是否充足（adequacy）。如果残差不是独立序列，表明模型没有充分捕捉数据中的相依性（dependency），需要进一步优化。model语句提供了回归残差独立性的三种检验方法，均为非参数方法，检验的原假设都是残差序列独立，同时提供回归残差为不相关序列的Durbin-Watson检验和Durbin-h检验。

BDS检验

BDS由Brock、Dechert和Scheinkman（1987）给出，用于检验序列的独立性，检验统计量$S_{BDS}(T,m,r)$定义为

$$
\begin{aligned}
S_{BDS}(T,m,r) &= \frac{\sqrt{T-m+1}\left(c_{m,m,T}(r)-c_{1,m,T}^{m}(r)\right)}{\sigma_{m,T}(r)} \\
c_{m,n,T} &= \frac{2}{(T-n+1)(T-n)}\sum_{s=n}^{T}\sum_{t=s+1}^{T}\prod_{j=0}^{m-1} I_r\left(z_{s-j},z_{t-j}\right) \\
I_r(z_s,z_t) &= \begin{cases} 1, & if\ \ |z_s-z_t|<r \\ 0, & otherwise \end{cases}
\end{aligned}
\tag{27}
$$

$\sigma_{m,T}(r)$为$c_{m,m,T}(r)-c_{1,m,T}^{m}(r)$的标准差。m称为嵌入维度（embedding dimension），可理解为存在相依性的最大滞后期，r称为半径（radius），可理解为z_t和z_s是否相依的标准，BDS检验将r值设定为$r=D\cdot\sigma$，σ为检验变量的标准误。检验的原假设为独立，备择假设为不独立。检验统计量的性质和检验效果的讨论可参考Kanzler（1999）。

model语句中添加选项BDS=(options)对指定序列独立性进行检验。语法格式为

```
model y=x1 x2 … xk/BDS=(M=m,D=d,z=options);
```

■ z=options确定要检验的序列，options取值为

Y：对模型因变量进行独立性检验；

RO：对OLS残差进行独立性检验。所谓OLS残差是指采用尤勒-沃尔克方法对（4）进行OLS估计得出估计残差$\hat{u}$。

R：对最终模型估计残差进行独立性检验。最终模型估计残差是指尤勒-沃尔克方法中第二步对模型的广义最小二乘估计或者迭代尤勒-沃尔克方法收敛后最后一次模型估计残差。

RM：对结构化残差进行独立性检验。结构化残差定义见（9）式。

SR：对标准化残差进行独立性检验。标准化残差是指最终模型残差除以标准误，标准误由GARCH模型估计的方差开方得出。

默认值为Y。

■ M=m:设定内嵌维度为m，默认值为20。

■ D=d：设定半径值r表达式中的倍数D的值为d，默认值为1.5。

程序7-19对sashelp.stocks的IBM的变量close建立的AR(3)+GARCH(1,1)模型的标准化残差进行BDS独立性检验。

程序7-19

```
data test  (where=(stock='IBM'));
set sashelp.stocks;
lclose1=lag(close);
lclose2=lag2(close);
lclose3=lag3(close);
lnvol=log(volume);
proc autoreg;
model close=lclose1-lclose3/GARCH=(p=1,q=1) BDS=(m=15,z=sr);
hetero lnvol;
run;
```

结果“BDS独立性检验”显示为

BDS 独立性检验			
距离	嵌入维	BDS	Pr > |BDS|
1.5000	2	-1.3692	0.1710
	3	-0.1521	0.8791
	4	0.2574	0.7969
	5	0.0107	0.9914
	6	0.3557	0.7220

“距离”（即D值）为默认值1.5，表中列出嵌入维度m=2到15对应的检验结果（仅列出前5个）。检验结果不能拒绝原假设，可以认为标准化残差独立，模型设定合适。

游程检验

游程检验（run test）是常见的一种独立性检验，属于符号检验，统计量的构造分为三步。

第一步，将原序列$\{z_t\}$转化为符号序列s，$\{s = sign(z_t - m_z)\}$，m_z为$\{z_t\}$的样本均值，$sign(x)$为符号函数，$x \geq 0$取值+，$x < 0$取值-。

第二步，计算序列s中的游程（run）个数R。一个游程是指连续为+或者连续为-的子序列片段。

第三步，计算+游程的个数R^+和-游程个数R^-。

游程检验统计量S_{runs}为

$$S_{runs} = \frac{(R-\mu)}{\sigma}, \mu = \frac{2R^+R^-}{T} + 1, \sigma = \frac{(\mu-1)(\mu-2)}{T-1}$$

model语句中添加选项BDS=(options)对指定序列进行独立性检验。语法格式为

```
model y=x1 x2 … xk/RUNS(z=options);
```

options的取值和功能与BDS=选项相同。

市场有效性可以归结为资产交易价格对数p_t是否鞅序列，即$E(p_t|F_{t-1}) = p_{t-1}$或者收益序列$r_t = p_t - p_{t-1}$是否鞅差序列，即$E(r_t|F_{t-1}) = 0$。独立序列一定是鞅差序列，通过检验序列是否独立可判断市场是否（弱）有效。程序7-20对sashelp.stocks中Intel收盘价close计算的收益r进行独立性游程检验。

程序7-20

```
data test(where=(stock='Intel'));
```

```
set sashelp.stocks;
r=dif(close);
proc autoreg;
model r=/runs=(z=y);
run;
```

“运行独立性检验”[①]一栏显示为

运行独立性检验	
RUNS	Pr > |RUNS|
0.6570	0.5112

表明不能拒绝r为独立序列的原假设，市场弱有效。model语句中既没有回归自变量，也没有nlag=选项设定的自回归项，模型参数只有常数项。

秩形式的von Neumann比检验

检验统计量S_{VNR}定义为

$$S_{VNR}=\frac{\sqrt{T}}{2}\left(\frac{\sum_{t=1}^{T-1}(R_{t+1}-R_t)^2}{T(T-1)^2/12}-2\right)$$

R_t为第t个观测在序列中的秩（Rank）。

model语句添加选项VNRRANK =(options)对指定序列进行独立检验。语法格式为

```
model y=x1 x2 … xk/VNRRANK=(pvalue=dis|sim,z=options);
```

- pvalue=dis：按渐近正态分布计算p-值。
- pvalue=sim：用模拟方法计算p-值。

样本量在11-100之间时，模拟方法计算的p-值较为准确；样本量小于10时，检验统计量具有精确分布函数，样本量大于100时，采用正态分布效果较好。默认为pvalue=dis。

- z=options中的选项的取值和功能与BDS=选项相同。

程序7-21对股票Intel的收盘价close的AR(3)+GARCH(1,1)模型的结构残差和标准化残差进行独立性VNRRNAK检验。

程序7-21

```
data test (where=(stock='Intel'));
set sashelp.stocks;
lclose1=lag(close);
lclose2=lag2(close);
lclose3=lag3(close);
lnvol=log(volume);
proc autoreg data=test;
model close=lclose:/GARCH=(p=1,q=1) VNRRANK=(z=rm);
hetero lnvol;
model close=lclose:/GARCH=(p=1,q=1) VNRRANK=(z=sr);
hetero lnvol;
run;
```

对结构化残差（rm）和标准化残差（sr）的“秩版本的von Neumann 比独立性检验”结果为

秩版本的 von Neumann 比独立性检验	
RVN	Pr > |RVN|
1.0358	0.3003

秩版本的 von Neumann 比独立性检验	
RVN	Pr > |RVN|
0.2832	0.7770

① “运行”是对“run”的汉化。应该汉化为“游程”。

可以认为结构化残差还是标准化残差是独立序列，模型设定合理。

Durbin-Watson检验和Durbin-h检验

Durbin-Watson检验用于模型误差项u_t是否存在自相关的检验，是时间序列回归模型中常见的一种检验。model语句执行后在模型汇总信息“普通最小二乘法估计”一栏中列出Durbin-Watson检验统计量值。

Durbin-h检验用于回归自变量包含因变量滞后变量时，回归模型误差项的序列相关检验。语句格式为

```
model y=x1 x2 … xk/LAGDV DW=n DWPROB;
```

选项LAGDV表明自变量中包含因变量滞后变量（LAGged Dependent Variables），DWPROB要求输出检验的p-值。执行后的输出结果包括滞后1阶到滞后n阶的DW值，以及Durbin-h检验的统计量值和p-值。

程序7-22对变量程序7-21数据集test中close建立AR(3)+GARCH(1,1)模型，进行Durbin-h检验，显示检验的概率值。

程序7-22

```
proc autoreg data=test;
model close=lclose:/GARCH=(p=1,q=1) LAGDV DW=2 DWPROB;
hetero lnvol;
run;
```

结果“其它统计量”和“Durbin-Watson统计量量”显示为

其他统计量			
统计量	值	概率	标签
Durbin t	-0.1645	0.4348	Pr > t

Durbin t即为Durbin-h统计量值。Durbin-h检验表明误差项存在一阶序列相关。由于解释变量中包含因变量滞后变量，D-W检验失效。

7.2.8 预测

proc autoreg既可以对因变量进行预测，也可以对因变量条件方差进行预测。既可以进行样本内预测（拟合），也可以进行样本外预测。样本内预测（拟合）通过输出数据集得到，样本外预测需要用DATA步编程实现。这里对样本内预测进行介绍，

因变量预测

因变量预测分为结构成分预测与完整模型预测，预测值及其置信区间计算公式参见（12）式和（13）式。从预测公式看出，完整模型预测$\tilde{y}_t$不仅包含解释变量对因变量的预测，即结构成分预测$\hat{y}_t = \hat{c} + \sum_{i=1}^{k} \hat{\beta}_i x_{it}$，也包含误差项按其自回归模型的预测$u_{t|t-1}$，考虑了因变量自相关性对预测的贡献，即$\tilde{y}_t = \hat{y}_t + u_{t|t-1}$。如果$u_t \sim AR(m)$，则

$$u_{t|t-1} \equiv E(u_t|F_{t-1}) = -\sum_{l=1}^{m} \hat{\varphi}_i \hat{u}_{t-l}, \hat{u}_{t-l} = y_{t-l} - \tilde{y}_{t-l}, l = 1,2,\ldots,m$$

模型有解释变量时，因变量预测需要解释变量的值，因此往往仅限于样本内预测。模型没有解释变量时，因变量预测按等价的自回归模型

$$y_t = c(1 + \sum_{i=1}^{m} \varphi_i) - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \cdots - \varphi_m y_{t-m} + \epsilon_t$$

进行。

条件方差预测

条件方差预测是按照GARCH公式（以GARCH（1,1）为例）计算样本内的条件方差h_t的估计值$\hat{h}_t$

$$\hat{h}_t = \hat{\omega} + \hat{\alpha}_1 \hat{\epsilon}_{t-1}^2 + \hat{\gamma}_1 \hat{h}_{t-1} + \hat{\delta}_1 w_{1t} + \hat{\delta}_2 w_{2t}$$

$\hat{\omega},\hat{\alpha}_1,\hat{\gamma}_1,\hat{\delta}_1,\hat{\delta}_2$为模型参数估计值，$\hat{\epsilon}_t$为残差，$w_{1t},w_{2t}$为外生变量。

输出数据集-output语句

output语句将模型预测值和残差等输出为SAS数据集。语句格式如下

OUTPUT OUT=SAS-data-set <options>;

选项options用来选择要输出的统计量以及统计量在输出数据集中的变量名，选项形式为“keyword=v”，keyword为统计量关键词，v是用户给出的变量名。

- HT=vname:输出GARCH=选项设定的GARCH模型计算出的误差项条件方差值$\hat{\sigma}_t^2 \triangleq \hat{h}_t$。
- P=vname:输出因变量的预测值，实际上是样本内拟合值。
- LCL=vname:输出因变量预测值置信区间下限。
- UCL=vname:输出因变量预测值置信区间上限。
- PM=vname:输出因变量的结构预测值。
- LCLM=vname:输出因变量的结构预测值置信区间下限。
- UCLM=vname:输出因变量的结构预测值置信区间上限。
- R=vname:输出整体模型残差。
- RM=vname:输出结构残差。

除了统计量关键词指定的输出内容外，输出数据集中还包含输入变量的所有变量。没有统计量关键词选项时，输出数据集与输入数据集相同。

程序7-23对程序数据集sashelp.stock中Microsoft收盘价close建立AR(2)+GARCH(1,1)模型，GARCH中有外生变量lvol。

程序7-23

```
data test (where=(stock='Microsoft'));
set sashelp.stocks;
lclose1=lag(close);
lclose2=lag2(close);
lclose3=lag3(close);
lnvol=log(volume);
proc autoreg data=test;
model close=lclose:/nlag=2 GARCH=(p=1,q=1) DWPROB;
hetero lnvol;
output out=s p=close_hat ht=gh;
run;
```

output语句及其选择out=s设定输出数据集为work.s，统计量关键词p=close_hath和ht=gh将因变量预测值和GARCH模型计算的条件方差输出到数据集work.s，分别以close_hat和gh为变量名。

习题：

1. SAS数据集tongrt是股票同仁堂（600085）2016年1月4日至2020年12月31日的日交易数据，变量clp_trt为日收盘价。用proc arima中的identify对变量clp_trt进行单位根检验，采用ADF检验方法，检验模型采用5阶滞后。
2. SAS数据集sh_bank是招商银行（60036）、农业银行（601288）、工商银行（601398）等多家上市银行2016年1月4日至2020年12月31日的股票日交易数据，stkcd、tdate、clp和vol分别表示股票代码、交易日期、收盘价和交易量。
   （1）采用where语句对工商银行（601398）的日收盘价进行Phillips-Perron单位根检验；当检验模型采用不同滞后阶数时，检验结果是否发生变化？
   （2）首先根据日收盘价计算日对数收益率，然后采用where数据集选项对招商银行的日对数收益率进行ADF单位根检验。当检验模型采用不同滞后阶数时，检验结果是否发生变化？
3. 用ESACF确定数据集tongrt中变量clp_trt的ARMA模型的阶数。确定结果是否唯一？你

更愿意选择哪个结果建立模型？为什么？如果要建立AR模型，如何采用ESACF确定阶数？确定的阶数是否比ARMA模型中AR阶数高？能够从理论上进行分析？

4.对数据集tongrt中的变量clp_trt建立ARMA模型。如果clp_trt有单位根而其一阶差分Δclp_trt没有单位根，则对Δclp_trt建立模型。

（1）采用MINIC方法确定模型阶数，采用极大似然方法估计模型回归系数。

（2）从数据结果给出的图形，判断残差是否服从正态分布。如果残差不服从正态分布，估计结果可信吗？哪些方面不可信？为什么？

（3）estamat语句中的两个选项outmodel=和outstat=，其功能有什么区别？用这两个选项将clp_trt的ARMA模型估计结果保存为SAS数据集。

（4）如果建立的是Δclp_trt的ARMA模型，用forecast语句进行10步预测，并将预测结果保存为SAS数据集fcast。这样的预测的对Δclp_trt的预测还是对原始变量clp_trt的预测？

5.过程proc autoreg和proc arima的功能有什么区别？如何用proc autoreg过程实现proc arma的功能？

6.SAS数据集sh_yiyao为同仁堂（600085）、太极集团（600129）、复兴医药（600196）等多家医药上市公司2016年1月4日至2020年12月31日的股票日交易数据，stkcd、tdate、clp和vol分别表示股票代码、交易日期、收盘价和交易量。对复兴医药的日收益建立模型。

（1）根据日收盘价计算日收益率变量；

（2）用data步及其where数据集选项生成只包含复兴医药(600196)数据的数据集fxyy；

（3）对日收盘价和日收益率数据分别进行KPSS单位根检验。KPSS检验和其它单位根检验的最大区别是什么？

（4）用proc arima过程确定日收益变量的AR模型阶数；

（5）用proc autoreg过程估计收益率变量的AR模型。

（6）对交易量vol取对数生成变量lgvol。用proc autoreg过程建立收益率变量的AR模型和GARCH（1,1）模型，并在GARCH（1,1）模型中添加外生变量lgvol，并将lgvol在GARCH模型中的系数限制为非负。

7. SAS数据集shindex和bys分别是上证综合指数（83_000001）和白云山（600332）2016年1月4日至2020年12月31日的日收盘信息，clp为收盘指数和日收盘价。为求出白云山股票的贝塔系数，进行如下回归

$$r_{b,t} = c + \phi_1 r_{b,t-1} + \phi_2 r_{b,t-2} + \beta_1 r_{shind,t} + \beta_2 r_{shind,t-2} + \epsilon_t,$$

其中误差项条件方差服从EGARCH（1,1）。

（1）用data步rename命令将shindex和bys中变量clp分别改名为clp_shindex和clp_bys；

（2）用proc sort对两个数据集按tdate排序；

（3）用merge和by语句匹配并接两个数据集，生成新数据集shid_bys；

（4）生成上证指数日收益率变量r_{shind}和白云山日收益变量r_b；

（5）生成r_b的1阶和2节滞后变量，生成r_{shind}1阶和2节滞后变量；

（6）用model语句设定模型，分别用选项nlag=和不使用该选项两种方式；

（7）将EGARCH模型中表示杠杆效应的参数约束为0；

（8）对杠杆效应参数是否为0进行参数约束检验；

（9）对标准化残差进行独立性游程检验，以判断模型设定是否充分。

（10）对模型进行Bai-Perron变点检验，变点个数设定为3。对检验结果进行说明。