**40. 时间序列分析Ⅲ—GARCH模型**

**（一）GRACH模型**

即自回归条件异方差模型，是金融市场中广泛应用的一种特殊非线性模型。

1982年，R.Engle在研究英国通货膨胀率序列规律时提出ARCH模型，其核心思想是残差项的条件方差依赖于它的前期值的大小。

1986年，Bollerslev在ARCH模型基础上对方差的表现形式进行了线性扩展，并形成了更为广泛的GARCH模型。

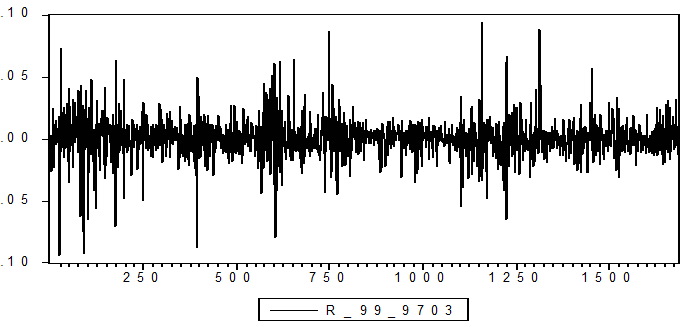
**一、金融时间序列的异方差性特征**

金融时间序列，无恒定均值（非平稳性），呈现出阶段性的相对平稳的同时，往往伴随着出现剧烈的波动性；具有明显的异方差（方差随时间变化而变化）特征：

尖峰厚尾：金融资产收益呈现厚尾和在均值处呈现过度波峰；

波动丛聚性：金融市场波动往往呈现簇状倾向，即波动的当期水平往往与它最近的前些时期水平存在正相关关系。

杠杆效应：指价格大幅度下降后往往会出现同样幅度价格上升的倾向。



因此，传统线性结构模型（以及时间序列模型）并不能很好地解释金融时间序列数据。

**二、ARCH(p)模型**

考虑k变量的回归模型



若残差项的均值为0，对*yt*取基于*t-*1时刻信息的期望：



该模型中，*yt*的无条件方差是固定的。但考虑*yt*的条件方差：



其中，表示基于t-1时刻信息集合Yt-1的yt的条件方差，若残差项存在自回归结构，则*yt*的条件方差不固定。

假设在前*p*期所有信息的条件下，残差项平方服从AR(*p*)模型：

 (\*)

其中为0均值、方差的白噪声序列。则残差项服从条件正态分布：



残差项的条件方差：



由两部分组成：

（1）常数项；

（2）ARCH项——变动信息，前*p*期的残差平方和

注：未知参数和利用极大似然估计法估计。

方差非负性要求都非负。为了使协方差平稳，需进一步要求方程



的根都位于单位圆外。若都非负，上式等价于.

注：若扰动项的条件方差不存在自相关，则有，此时，即残差的条件方差同方差性情形。

**三、GARCH(*p*,*q*)模型**

ARCH(*p*)模型在实际应用中，为了得到较好的拟合效果，往往需要很大的阶数*p*，从而增加了待估参数个数、引发多重共线性、非限制估计违背非负性要求。

1986年，Bollerslev将ARCH(*p*)模型推广为广义自回归条件异方差模型GARCH(p, q)：残差的条件方差表示为



由三项组成，

（1）常数项；

（2）ARCH项；

（3）GARCH项——前q期预测方差.

注：未知参数用极大似然法估计，通常残差的假设分布有正态分布、t分布、广义误差分布；该模型也要求非负；若要求是平稳过程，需要限制. 实际上，GARCH(p, q)模型是将残差平方用ARMA(q,p)模型描述。

**四、ARCH检验**

检验模型的残差是否具有ARCH效应有两种方法：

1. ARCH LM检验——拉格朗日乘数检验

检验原假设H0：残差序列直到p阶都不存在ARCH效应；需进行如下回归：



检验回归有两个统计量：

（1）F统计量——检验回归系数是否显著为0.

（2）T×R2统计量——LM统计量，其中T为观察值个数，R2为回归拟合优度，该统计量渐近服从分布。

2. 残差平方相关图

残差平方相关图显示残差平方序列，直到任意指定的滞后阶数的自相关函数（AC）和偏自相关函数（PAC），并计算相应滞后阶数的QLB统计量。若不存在ARCH效应，则任意滞后阶数的自相关函数（AC）和偏自相关函数（PAC）都近似为0.

**五、GARCH-M模型**

一般风险越大，预期收益越大。在回归模型中加入一项“利用条件方差表示的预期风险”：





称为GARCH-M模型。

另外，还有非对称冲击模型：TARCH、EGARCH、PARCH等（略）。

**（二）PROC AUTOREG过程**

SAS中的AUTOREG过程，是用于估计和预测误差项自相关或异方差的时间序列数据的线性回归模型。

自回归误差模型被用来校正自相关系数和广义自回归条件异方差模型GARCH，并且其变体如广义的ARCH（GARCH）、方差无穷的GARCH（IGARCH）、指数的GARCH（EGARCH）和依均值的GARCH（GARCH-M）被用于异方差的建模和校正。

自回归过程autoreg可以拟合任意阶的自回归误差模型，并且可以拟合子集自回归模型。

为了诊断自相关性，过程产生广义Durbin-Watson（DW）统计量和其边缘概率。普通回归分析假定误差方差对于所有观察是相同的，但当误差方差不相同时，数据被称为异方差，此时普通最小二乘法估计不是有效的，同时也影响预测值置信区间的精确性。Autoreg过程能检验异方差，并且提供GARCH模型族来估计和校正数据易变性。对于带有自相关扰动和随时间变化的条件异方差模型，过程输出条件均值和条件方差的预测值。

基本语法：

proc autoreg data=数据集 <可选项> ;

model 因变量=独立回归变量列表 </选项列表>;

output out=数据集 <选项列表>;

<by 变量;>

说明：

（1）proc autoreg语句可选项

 outest=数据集——把估计参数输出到指定数据集；

 covout=数据集——把估计参数的协方差阵输出到指定数据集；

（2）model语句

 center——通过减去均值中心化因变量并且取消模型的均值参数；

 noint——取消模型的均值参数；

 nlag＝数值 | (数值列表)——指定自回归误差的阶或者自回归误差的时间间隔的子集。例如，nlag=3与nlag=(1 2 3)作用相同，但与nlag=(1 3)等不同；

 garch=(q=数值, p=数值,type=选择值, mean, noint, tr)——指定广义条件异方差GARCH模型的类型。例如，定义GARCH(2,1)回归模型：

model y=x1 x2 / garch=(q=2, p=1);

 注意：SAS系统的自回归参数符号q和p与前文所述公式中的符号p和q正好相反。定义GARCH-M(1,1)回归模型：

model y=x1 x2 / garch=(q=2,p=1,mean);

 type=选择值——指定GARCH模型的类型：默认为noineq表示无约束GARCH模型； nonneg表示非负约束GARCH模型；stn表示约束GARCH模型系数的和小于1； integ表示IGARCH模型； exp表示EGARCH模型；

noint——取消条件异方差模型中的均值参数；

tr——GARCH模型的估计使用信赖区域方法，缺省值为对偶拟牛顿法；

archtest——要求用portmantea Q检验统计量和Engle的拉格朗日乘子LM检验是否存在条件异方差情况，即是否有ARCH效应；

coef——输出前几条观察的变换系数；

corrb——输出参数估计的估计相关系数；

covb——输出参数估计的估计协方差；

 dw=n——输出直到n阶的DW统计量，默认n=1;

 dwprob——输出DW统计量的p值，当误差自由度大于300时dwprob选项被忽略；

 ginv——输出Yule-Walker解的自协方差的Toeplitz矩阵的逆；

 itprint——输出每步迭代的目标函数和参数估计；

 lagdetp——输出DW t统计量，它用于检验存在时滞因变量时残差的自相关性；

 lagdep=回归变量——输出DW h统计量，它用于检验一阶自相关性；

 partial——输出偏自相关；

 backstep——去掉非显著自回归参数，参数按最小显著性的次序被去掉；

 slstay=数值——指定被backstep选项使用的显著水平，默认为0.05；

 converge=数值——指定在迭代自回归参数估计时参数的变化量的最大绝对值小于此数值，那么认为收敛，默认为0.001;

 maxiter=数值——指定允许迭代的最大次数，默认为50;

 method=ml | ols | yw | ityw——指定估计的方法，分别为：最大似然估计、无条件最小二乘法、Yule-Walker估计、迭代Yule-Walker估计；

 nomiss——使用没有缺失值的第一个连贯时间序列数据集，进行模型拟合估计。否则，跳过数据集开始的任何缺失值，使用独立回归变量和因变量都不带缺失值的所有数据。请特别注意，为了保持时间序列中正确的时间间隔，必须要增加时间刻度值，这样就会产生因变量缺失值的观察。当因变量缺失时，过程可以产生预测值。如果缺失值很多，则应使用ML估计。

（3）output语句

 out=数据集——指定包含预测值和变换值的输出数据集；

 alphacli=数值——设置时间序列预测值置信区间的显著水平，默认为0.05;

 alphaclm=数值——设置模型结构部分预测值置信区间的显著水平，缺省值为0.05;

 cev=变量——把条件误差方差写入到输出数据集的指定变量中，仅GARCH模型被估计时才使用；

 cpev＝变量——把条件预测误差方差写入到输出数据集的指定变量中，仅GARCH模型被估计时才使用；

 constant＝变量——把被变换的均值写入到输出数据集的指定变量中；

 lcl＝变量——把预测值的置信下限写入到输出数据集的指定变量中；

 ucl＝变量——把预测值的置信上限写入到输出数据集的指定变量中；

 lclm＝变量——把模型结构部分预测值的置信下限写入到输出数据集的指定变量中；

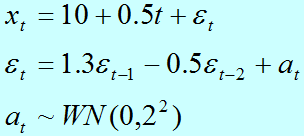
 uclm＝变量——把模型结构部分预测值的置信上限写入到输出数据集的指定变量中；

 p＝变量——把预测值写入到输出数据集的指定变量中；

 rm＝变量——把来自模型结构部分预测的残差写入到输出数据集的指定变量中；

 transform＝变量——把被变换的变量写入到输出数据集的指定变量中。

**例1**对模拟方法生成的时间趋势加二阶自回归误差模型的时间序列数据，用PROC AUTOREG过程进行分析和建模，以便于比较和判断各种求解模型和运算结果的好坏。  
 模型：



**（一）按照模拟模型生成数据集**

代码：

**data** randar;

e1=**0**;

e11=**0**;

do t=-**10** to **36**;

e=**1.3**\*e1-**0.5**\*e11+**2**\*rannor(**12346**);

x=**10**+**0.5**\*t+e;

e11=e1;

e1=e;

if t>**0** then

output;

end;

**run**;

**proc** **print** data=randar;

**run**;

运行结果：

| **Obs** | **e1** | **e11** | **t** | **e** | **x** |
| --- | --- | --- | --- | --- | --- |
| **1** | -2.73816 | -4.99186 | 1 | -2.73816 | 7.7618 |
| **2** | 0.03674 | -2.73816 | 2 | 0.03674 | 11.0367 |
| **3** | 0.66356 | 0.03674 | 3 | 0.66356 | 12.1636 |
| **4** | -0.86233 | 0.66356 | 4 | -0.86233 | 11.1377 |
| **5** | -4.30165 | -0.86233 | 5 | -4.30165 | 8.1984 |
| **6** | -3.88957 | -4.30165 | 6 | -3.88957 | 9.1104 |
| **7** | -3.56635 | -3.88957 | 7 | -3.56635 | 9.9337 |
| **8** | -3.68908 | -3.56635 | 8 | -3.68908 | 10.3109 |
| **9** | -2.50263 | -3.68908 | 9 | -2.50263 | 11.9974 |
| **10** | -1.80957 | -2.50263 | 10 | -1.80957 | 13.1904 |
| **11** | -0.14763 | -1.80957 | 11 | -0.14763 | 15.3524 |
| **12** | 1.78537 | -0.14763 | 12 | 1.78537 | 17.7854 |
| **13** | 4.09011 | 1.78537 | 13 | 4.09011 | 20.5901 |
| **14** | 3.54577 | 4.09011 | 14 | 3.54577 | 20.5458 |
| **15** | -0.92995 | 3.54577 | 15 | -0.92995 | 16.5701 |
| **16** | -3.39550 | -0.92995 | 16 | -3.39550 | 14.6045 |
| **17** | -4.57207 | -3.39550 | 17 | -4.57207 | 13.9279 |
| **18** | -4.02602 | -4.57207 | 18 | -4.02602 | 14.9740 |
| **19** | -3.75333 | -4.02602 | 19 | -3.75333 | 15.7467 |
| **20** | -1.73776 | -3.75333 | 20 | -1.73776 | 18.2622 |
| **21** | -1.42859 | -1.73776 | 21 | -1.42859 | 19.0714 |
| **22** | -0.98120 | -1.42859 | 22 | -0.98120 | 20.0188 |
| **23** | -1.48741 | -0.98120 | 23 | -1.48741 | 20.0126 |
| **24** | -4.41515 | -1.48741 | 24 | -4.41515 | 17.5848 |
| **25** | -4.14229 | -4.41515 | 25 | -4.14229 | 18.3577 |
| **26** | -5.57633 | -4.14229 | 26 | -5.57633 | 17.4237 |
| **27** | -5.55633 | -5.57633 | 27 | -5.55633 | 17.9437 |
| **28** | -4.17604 | -5.55633 | 28 | -4.17604 | 19.8240 |
| **29** | -2.45303 | -4.17604 | 29 | -2.45303 | 22.0470 |
| **30** | -0.32189 | -2.45303 | 30 | -0.32189 | 24.6781 |
| **31** | -0.24246 | -0.32189 | 31 | -0.24246 | 25.2575 |
| **32** | 1.61987 | -0.24246 | 32 | 1.61987 | 27.6199 |
| **33** | 2.26259 | 1.61987 | 33 | 2.26259 | 28.7626 |
| **34** | 0.48540 | 2.26259 | 34 | 0.48540 | 27.4854 |
| **35** | -1.11570 | 0.48540 | 35 | -1.11570 | 26.3843 |
| **36** | -2.95901 | -1.11570 | 36 | -2.95901 | 25.0410 |

变量e对应εt, e1对应εt-1, e2对应εt-2.

表达式2\*rannor(12346)，将生成独立同分布均值为0，标准差为2的正态分布随机数，对应于公式中均值为0，方差为22的白噪声误差序列。

DO循环从t=－10开始而不是直接从t=1开始的原因，是让模拟生成的二阶自回归误差序列 有一段时间（t=－10到0）进行初始化，以便到达稳定的随机序列值。

**（二）普通最小二乘法回归模型**

代码：

**proc** **autoreg** data=randar PLOTS(ONLY) = FITPLOT;

model x=t;

**run**;

运行结果及说明：

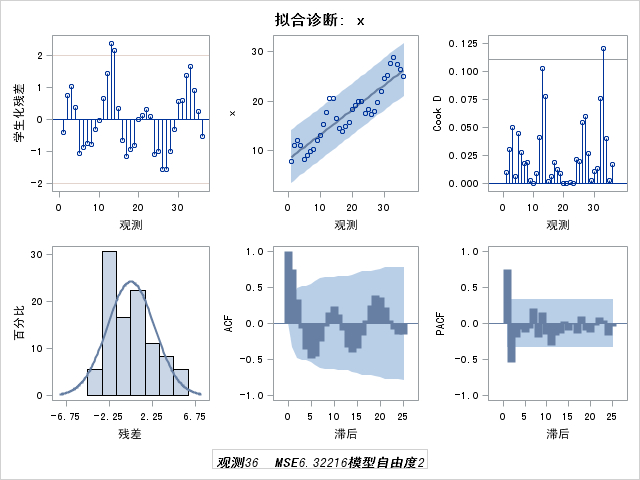
AUTOREG 过程

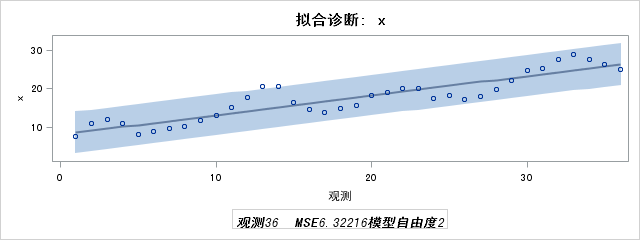
|  |  |
| --- | --- |
| **因变量** | x |

| **普通最小二乘法估计** | | | |
| --- | --- | --- | --- |
| **SSE** | 214.953429 | **DFE** | 34 |
| **MSE** | 6.32216 | **均方根误差** | 2.51439 |
| **SBC** | 173.659101 | **AIC** | 170.492063 |
| **MAE** | 2.01903356 | **AICC** | 170.855699 |
| **MAPE** | 12.5270666 | **HQC** | 171.597444 |
| **Durbin-Watson** | 0.4752 | **回归 R 方** | 0.8200 |
|  |  | **总 R 方** | 0.8200 |

| **参数估计值** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **变量** | **自由度** | **估计值** | **标准误差** | **t 值** | **近似 Pr > |t|** |
| **Intercept** | 1 | 8.2308 | 0.8559 | 9.62 | <.0001 |
| **t** | 1 | 0.5021 | 0.0403 | 12.45 | <.0001 |

AUTOREG 过程





普通最小二乘回归基于统计假设：误差相互对立。然而，时间序列数据，普通回归后的残差常常是相关的。

这将导致：第一，对于参数的显著性和置信限的统计检验将不正确；第二，回归系数的估计不象考虑到自相关性时的估计一样有效；第三，由于回归残差不独立，它们包含可用来改进预测值的信息。

由于这些原因，所以对时间序列数据不使用普通回归proc reg过程而使用带自回归误差的回归pro autoreg过程。Model语句中指定回归模型，没有可选项，是要求利用普通最小二乘法做x对t的回归。为便于对比，绘制了散点图和线性回归趋势线。

回归R2是对回归模型的R2，总R2是包括自回归误差在内的整体模型的R2. 现在还无自回归误差模型，故两个R2相等。

估计模型为：

OLS：



该模型较合理地接近真实值，但是误差方差估计6.32216远大于真实值4。误差方差估计值远大于真实值（通过对模型的残差作自相关性检验来判断和识别），说明模型还有信息没有提取。

**（三）检验模型的自相关系数**

在实际问题中，需要检验自相关性是否存在，以及存在几阶自相关。Durbin-Watson检验是广泛使用的自相关性的检验方法。选项dw=4和dwprob是要求过程进行1到4阶的OLS残差中自相关性Durbin-Watson检验，并要求输出Durbin-Watson统计量的边缘显著水平p值。

注意：对于季节性时间序列数据，自相关性检验应该至少检验与季节性阶一样大的阶。例如，对于月度数据至少应取dw=12。

代码：

**proc** **autoreg** data=randar;

model x=t /dw=**4** dwprob;

**run**;

运行结果及说明：

| **Durbin-Watson 统计量** | | | |
| --- | --- | --- | --- |
| **顺序** | **DW** | **Pr < DW** | **Pr > DW** |
| **1** | 0.4752 | <.0001 | 1.0000 |
| **2** | 1.2935 | 0.0137 | 0.9863 |
| **3** | 2.0694 | 0.6545 | 0.3455 |
| **4** | 2.5544 | 0.9818 | 0.0182 |

|  |  |
| --- | --- |
| **Note:** | **Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.** |

| **参数估计值** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **变量** | **自由度** | **估计值** | **标准误差** | **t 值** | **近似Pr > |t|** |
| **Intercept** | 1 | 8.2308 | 0.8559 | 9.62 | <.0001 |
| **t** | 1 | 0.5021 | 0.0403 | 12.45 | <.0001 |

一阶Durbin-Watson统计量为0.4752，其p值为<0.0001，极其显著，强烈拒绝一阶自相关系数为0的原假设。因此，自相关性的校正是必须的。

要注意的是，利用Durbin-Watson检验可决定是否需要做自相关性校正。但广义的Durbin-Watson检验不应该用于确定自回归的阶数。因为高阶的检验是在无低阶自相关性的原假设下进行的。例如，若普通的Durbin-Watson检验表明无一阶自相关性，那么可以用二阶检验去检验二阶自相关性。一旦检验出某阶有自相关性存在，那么更高阶的检验将不适用。这里由于1阶自相关性检验是显著的，所以2、3、4阶的检验是被忽略的。

**（四）自回归误差模型**

代码：

**proc** **autoreg** data=randar;

model x=t /nlag=**2** method=ml;

output out=pout p=xhat pm=trendhat;

**run**;

**proc** **gplot** data=pout;

plot x\*t=**1** xhat\*t=**2** trendhat\*t=**3** /overlay;

symbol1 v=star i=none c=red h=**2.5**;

symbol2 v=plus i=join c=blue h=**2.5**;

symbol3 v=none i=join c=green w=**2**;

title1 'Auto-Regression';

title2 'nlag=2 method=ml';

**run**;

运行结果及说明：

nlag=2选项，指定误差为1阶、2阶自回归模型，其另一种格式，如nlag=(1 4 5)，表示自回归误差模型为



选项method=ml，指定回归参数的估计采用精确最大似然估计，默认为Yule-Walker估计。

用output语句输出预测值到pout数据集中，预测值有两种类型：第一类是部分模型预测值：仅通过模型的结构部分得到，即由部分得到，这是响应变量*xt*在时刻t的无条件均值估计，用选项pm=trendhat将该预测值数据集pout的变量trendhat中；

第二类，是整体模型预测值，既包含模型的结构部分也包含自回归误差过程的预测值，即由整体模型得到，用选项p=xhat将该预测值输出到数据集pout的变量xhat中。

| **普通最小二乘法估计** | | | |
| --- | --- | --- | --- |
| **SSE** | 214.953429 | **DFE** | 34 |
| **MSE** | 6.32216 | **均方根误差** | 2.51439 |
| **SBC** | 173.659101 | **AIC** | 170.492063 |
| **MAE** | 2.01903356 | **AICC** | 170.855699 |
| **MAPE** | 12.5270666 | **HQC** | 171.597444 |
| **Durbin-Watson** | 0.4752 | **回归 R 方** | 0.8200 |
|  |  | **总 R 方** | 0.8200 |

| **参数估计值** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **变量** | **自由度** | **估计值** | **标准误差** | **t 值** | **近似 Pr > |t|** |
| **Intercept** | 1 | 8.2308 | 0.8559 | 9.62 | <.0001 |
| **t** | 1 | 0.5021 | 0.0403 | 12.45 | <.0001 |

| **自相关估计值** | | | |
| --- | --- | --- | --- |
| **滞后** | **协方差** | **相关** | **-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1** |
| **0** | 5.9709 | 1.000000 | |                    |\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*| |
| **1** | 4.5169 | 0.756485 | |                    |\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*     | |
| **2** | 2.0241 | 0.338995 | |                    |\*\*\*\*\*\*\*             | |

|  |  |
| --- | --- |
| **初步 MSE** | 1.7943 |

| **自回归参数的估计值** | | | |
| --- | --- | --- | --- |
| **滞后** | **系数** | **标准误差** | **t 值** |
| **1** | -1.169057 | 0.148172 | -7.89 |
| **2** | 0.545379 | 0.148172 | 3.68 |

|  |
| --- |
| 算法收敛。 |

| **最大似然估计** | | | |
| --- | --- | --- | --- |
| **SSE** | 54.7493022 | **DFE** | 32 |
| **MSE** | 1.71092 | **均方根误差** | 1.30802 |
| **SBC** | 133.476508 | **AIC** | 127.142432 |
| **MAE** | 0.98307236 | **AICC** | 128.432755 |
| **MAPE** | 6.45517689 | **HQC** | 129.353194 |
| **对数似然** | -59.571216 | **回归 R 方** | 0.7280 |
| **Durbin-Watson** | 2.2761 | **总 R 方** | 0.9542 |
|  |  | **观测** | 36 |

| **参数估计值** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **变量** | **自由度** | **估计值** | **标准误差** | **t 值** | **近似 Pr > |t|** |
| **Intercept** | 1 | 7.8833 | 1.1693 | 6.74 | <.0001 |
| **t** | 1 | 0.5096 | 0.0551 | 9.25 | <.0001 |
| **AR1** | 1 | -1.2464 | 0.1385 | -9.00 | <.0001 |
| **AR2** | 1 | 0.6283 | 0.1366 | 4.60 | <.0001 |

| **指定所用的自回归参数** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **变量** | **自由度** | **估计值** | **标准误差** | **t 值** | **近似 Pr > |t|** |
| **Intercept** | 1 | 7.8833 | 1.1678 | 6.75 | <.0001 |
| **t** | 1 | 0.5096 | 0.0551 | 9.26 | <.0001 |

第一部分输出普通最小二乘法回归模型的回归结果（与前文一致），包括诊断统计量和参数估计表，模型公式见前文。

第二部分输出自回归误差模型的输出结果，首先是Preliminary MSE = 1.7943为最大似然估计迭代计算的初始Yule-Walker估计值。然后给出了在原假设时间间隔为1和2的自相关系数为0情况下，1阶和2阶自相关系数检验统计量、标准差和t值，t=-7.890和3.681（绝对值都很大），拒绝1阶和2阶自相关系数为0的原假设。

自回归误差模型的回归结果也包括回归统计量和参数估计表。参数估计表显示了回归系数的ML估计，比OLS输出的参数估计表多二个附加行，被标记为AR(1)和AR(2)，分别为自回归误差的一阶系数和二阶系数的估计值。估计模型为：

 (\*)

基于自回归参数的估计值，重新计算了参数的标准误差和t值，如重新计算截距7.8833的标准差为1.1678而不是1.1693.

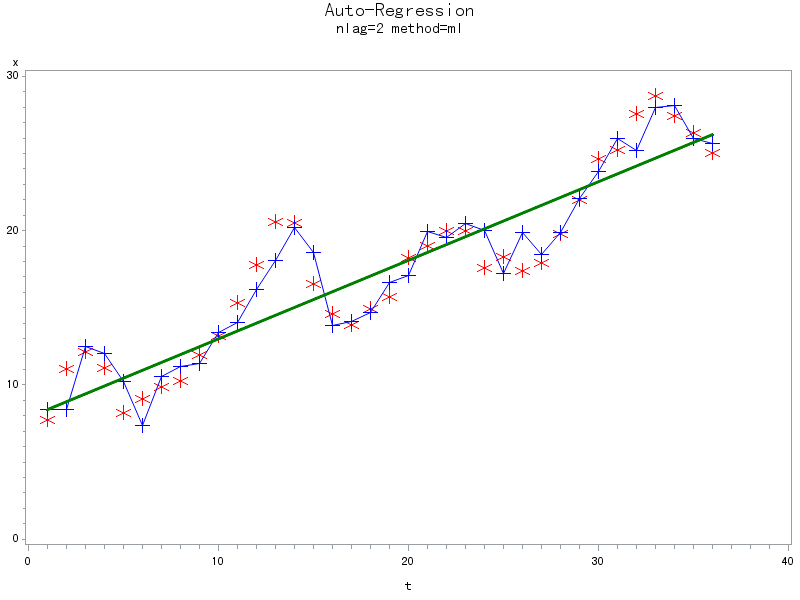
比较自回归误差模型和普通最小二乘法回归模型，整体的R2=0.9542>0.8200. 这反映了由于考虑了残差作用而改进了拟合模型，将会使预测值更准确。

显然，自回归误差模型部分的R2总是小于普通最小二乘法回归模型部分R2（0.7280<0.8200），这是因为在部分模型时，普通最小二乘法回归已经达到最优，它的部分模型R2与整体模型R2是相等的。自回归误差模型的信息准则SBC和AIC值都分别比OLS模型的要小，如133.476508<173.659101，127.142432<170.492063，同样说明估计模型(40.83)式比估计模型(40.82)式要好。

对于自回归误差模型的MSE为1.71092，此值远小于真实值4，而对于OLS模型的MSE为6.32216，此值远大于真实值4，从MSE值的大小对比中，也可以得出自回归误差模型较好。

注意：在小样本情况下，自回归误差模型倾向于低估σ2，而OLS模型倾向于高估σ2。

最后，注意到自回归误差模型的DW统计量为2.2761，接近2，残差序列不相关，而OLS模型的DW统计量为0.4752，接近0偏离2，残差序列正相关，同样可以得出自回归误差模型已经完全提取了内在规律的信息，只剩下纯随机波动。



红\*为原始序列散点图；蓝色线+号为整体模型预测值；绿色线为部分模型预测值。注意，该绿色趋势线模型为：

AR(2)：

不同于前文的OLS模型趋势线。

**（五）预测自回归误差模型**

代码：

**data** new;

x=**.** ;

do t=**37** to **46**;

output;

end;

**run**;

**data** randar37;

merge randar new;

by t;

**run**;

**proc** **autoreg** data=randar37;

model x=t /nlag=**2** method=ml;

output out=poutp p=xhat pm=trendhat lcl=lcl ucl=ucl;

**run**;

**proc** **gplot** data=poutp;

plot x\*t=**1** xhat\*t=**2** trendhat\*t=**3** lcl\*t=**3** ucl\*t=**3**/overlay href=**36.5**;

symbol1 v=star i=none c=red h=**2.5**;

symbol2 v=plus i=join c=blue h=**2.5** l=**1**;

symbol3 v=none i=join c=green w=**2** ;

title1 'Auto-Regression:AR(2)';

title2 'predict estimation';

**run**;

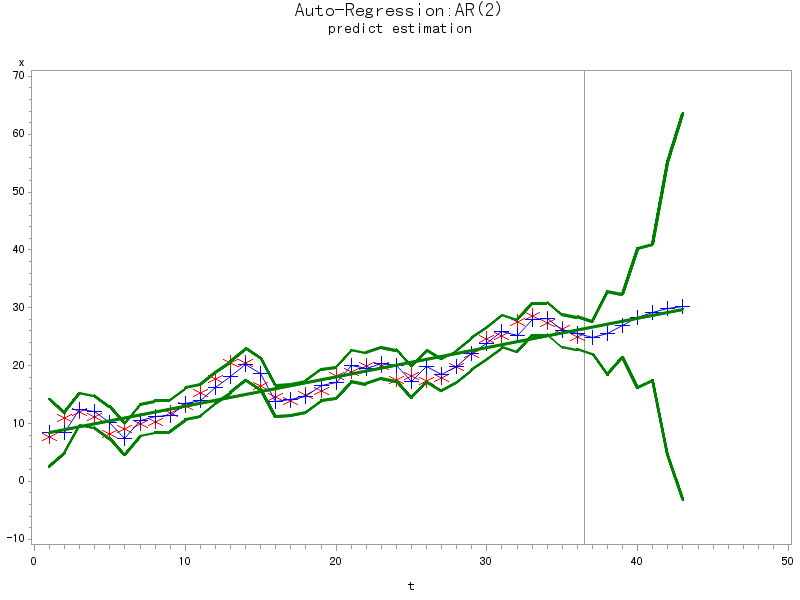
运行结果及说明：

先产生10个等间隔的将来时间点，对应的因变量*x*值先设为缺失值。再与原序列合并，作为数据集randar37。

proc autoreg过程，对数据集randar37进行自回归误差模型拟合，拟合自回归误差模型为公式(\*)，拟合的预测趋势线是 (\*)式中的部分模型公式（即不包括自回归误差）。

通过output语句将预测值（xhat）、预测趋势（trendhat）、以及95%置信上限（ucl）和置信上限（lcl）输出到poutp数据集中。

proc gplot过程，绘制原始值、预测值、趋势值和置信限的散布图。在时间t=36.5处画一条垂直参考线（选项href=36.5）表示将来时间间隔上的预测值从这里开始。请注意，预测考虑了对于趋势的偏离但对于较长的预测收敛回到趋势直线上，但同时置信区间却是变大。



**（六）自回归误差模型阶数的确定——逐步自回归**

若模型的自相关系数检验确定需要做自相关性校正，那么必须选取要用的自回归误差模型的阶数。通常是通过逐步自回归来寻找这个阶数。

逐步自回归方法最初拟合一个长阶自回归模型，然后逐步去掉自回归参数直到所剩下的自回归参数都有显著t检验为止。

SAS实现：在proc autoreg过程的model语句中使用backstep选项，并用nlag=选项指定一个长阶。在指定长阶时，一个好的想法是指定大于任何潜在季节性的阶值，例如，对于月度数据，取nlag=13；对于季度数据，取nlag=5。

代码：

**proc** **autoreg** data=randar;

model x=t /nlag=**5** method=ml backstep;

title 'Auto-Regression:Backstep';

**run**;

运行结果：

| **自相关估计值** | | | |
| --- | --- | --- | --- |
| **滞后** | **协方差** | **相关** | **-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1** |
| **0** | 5.9709 | 1.000000 | |                    |\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*| |
| **1** | 4.5169 | 0.756485 | |                    |\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*     | |
| **2** | 2.0241 | 0.338995 | |                    |\*\*\*\*\*\*\*             | |
| **3** | -0.4402 | -0.073725 | |                   \*|                    | |
| **4** | -2.1175 | -0.354632 | |             \*\*\*\*\*\*\*|                    | |
| **5** | -2.8534 | -0.477887 | |          \*\*\*\*\*\*\*\*\*\*|                    | |

| **向后消除自回归项** | | | |
| --- | --- | --- | --- |
| **滞后** | **估计值** | **t 值** | **Pr > |t|** |
| **4** | -0.052908 | -0.20 | 0.8442 |
| **3** | 0.115986 | 0.57 | 0.5698 |
| **5** | 0.131734 | 1.21 | 0.2340 |

|  |  |
| --- | --- |
| **初步 MSE** | 1.7943 |

| **自回归参数的估计值** | | | |
| --- | --- | --- | --- |
| **滞后** | **系数** | **标准误差** | **t 值** |
| **1** | -1.169057 | 0.148172 | -7.89 |
| **2** | 0.545379 | 0.148172 | 3.68 |

|  |
| --- |
| 算法收敛。 |

结果显示了5个时间间隔的自相关系数的估计。自回归项向后消除报告显示了在时间间隔4、3、5时自回归不显著（p值0.8442、0.5698、0.2340都大于0.05），即不能拒绝在这些时间间隔上自相关系数为0的原假设，因此被消除。而在时间间隔1、2时自回归显著，因此被保留下来。故应该采用拟合二阶自回归误差模型。

默认显著水平是0.05，可以用“slstay=选项”来修改它。