

# SPSS 时间序列分析

## 3.3.1 时间序列概述

### 1. 基本概念

- (1)一般概念：系统中**某一变量**的观测值**按时间顺序**（时间间隔相同）排列成一个数值序列，展示研究对象在一定时期内的变动过程，从中寻找和分析事物的**变化特征、发展趋势和规律**。它是系统中某一变量受其它各种因素影响的总结果。
- (2)研究实质：通过处理预测目标**本身**的时间序列数据，获得事物随时间过程的演变特性与规律，进而预测事物的未来发展。它不研究事物之间相互依存的因果关系。
- (3)假设基础：惯性原则。即在一定条件下，被预测事物的过去变化趋势会延续到未来。暗示着历史数据存在着某些信息，利用它们可以解释与预测时间序列的现在和未来。  
近大远小原理（时间越近的数据影响力越大）和无季节性、无趋势性、线性、常数方差等。
- (4)研究意义：许多经济、金融、商业等方面的数据都是时间序列数据。  
时间序列的预测和评估技术相对完善，其预测情景相对明确。  
尤其关注预测目标可用数据的数量和质量，即时间序列的长度和预测的频率。

### 2. 变动特点

- (1)趋势性：某个变量随着时间进展或自变量变化，呈现一种比较缓慢而长期的持续上升、下降、停留的同性质变动趋向，但变动幅度可能不等。
- (2)周期性：某因素由于外部影响随着自然季节的交替出现高峰与低谷的规律。
- (3)随机性：个别为随机变动，整体呈统计规律。
- (4)综合性：实际变化情况一般是几种变动的叠加或组合。预测时一般设法过滤除去不规则变动，突出反映趋势性和周期性变动。

### 3. 特征识别

认识时间序列所具有的变动特征，以便在系统预测时选择采用不同的方法。

- (1)随机性：均匀分布、无规则分布，可能符合某统计分布。(用因变量的散点图和直方图及其包含的正态分布检验随机性，大多数服从正态分布。)
- (2)平稳性：样本序列的自相关函数在某一固定水平线附近摆动，即方差和数学期望稳定为常数。  
样本序列的自相关函数只是时间间隔的函数，与时间起点无关。其具有对称性，能反映平稳序列的周期性变化。

特征识别利用自相关函数 ACF:  $\rho_k = \gamma_k / \gamma_0$

其中  $\gamma_k$  是  $y_t$  的  $k$  阶自协方差，且  $\rho_0=1$ 、 $-1 < \rho_k < 1$ 。

平稳过程的自相关系数和偏自相关系数都会以某种方式衰减趋近于 0，前者测度当前序列与先前序列之间简单和常规的相关程度，后者是在控制其它先前序列的影响后，测度当前序列与某一先前序列之间的相关程度。

实际上, 预测模型大都难以满足这些条件, 现实的经济、金融、商业等序列都是非稳定的, 但通过数据处理可以变换为平稳的。

#### 4. 预测类型

- (1)点预测: 确定唯一的最好预测数值, 其给出了时间序列未来发展趋势的一个简单、直接的结果。但常产生一个非零的预测误差, 其不确定程度为点预测值的置信区间。
- (2)区间预测: 未来预测值的一个区间, 即期望序列的实际值以某一概率落入该区间范围内。区间的长度传递了预测不确定性的程度, 区间的中点为点预测值。
- (3)密度预测: 序列未来预测值的一个完整的概率分布。根据密度预测, 可建立任意置信水平的区间预测, 但需要额外的假设和涉及复杂的计算方法。

#### 5. 基本步骤

- (1)分析数据序列的变化特征。
- (2)选择模型形式和参数检验。
- (3)利用模型进行趋势预测。
- (4)评估预测结果并修正模型。

### 3.3.2 随机时间序列

系统中某一因素变量的时间序列数据没有确定的变化形式, 也不能用时间的确定函数描述, 但可以用概率统计方法寻求比较合适的随机模型近似反映其变化规律。(自变量不直接含有时间变量, 但隐含时间因素)

#### 1. 自回归 AR(p)模型

(R: 模型的名称 P: 模型的参数) (自己影响自己, 但可能存在误差, 误差即没有考虑到的因素)

- (1)模型形式 ( $\varepsilon_t$  越小越好, 但不能为 0:  $\varepsilon$  为 0 表示只受以前 Y 的历史的影响不受其他因素影响)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$

式中假设:  $y_t$  的变化主要与时间序列的历史数据有关, 与其它因素无关;

$\varepsilon_t$  不同时刻互不相关,  $\varepsilon_t$  与  $y_t$  历史序列不相关。

式中符号:  $p$  模型的阶次, 滞后的时间周期, 通过实验和参数确定;  
 $y_t$  当前预测值, 与自身过去观测值  $y_{t-1}$ 、...、 $y_{t-p}$  是同一序列不同时刻的随机变量, 相互间有线性关系, 也反映时间滞后关系;

$y_{t-1}$ 、 $y_{t-2}$ 、...、 $y_{t-p}$  同一平稳序列过去  $p$  个时期的观测值;

$\varphi_1$ 、 $\varphi_2$ 、...、 $\varphi_p$  自回归系数, 通过计算得出的权数, 表达  $y_t$  依赖于过去的程度, 且这种依赖关系恒定不变;

$\varepsilon_t$  随机干扰误差项, 是 0 均值、常方差  $\sigma^2$ 、独立的白噪声序列, 通过估计指定的模型获得。

#### (2)识别条件

当  $k > p$  时, 有  $\varphi_k = 0$  或  $\varphi_k$  服从渐近正态分布  $N(0, 1/n)$  且  $(|\varphi_k| > 2/n^{1/2})$  的个数  $\leq 4.5\%$ , 即平稳时间序列的偏相关系数  $\varphi_k$  为  $p$  步截尾, 自相关系数  $r_k$  逐步衰减而不截尾, 则序列是 AR(p)模型。

实际中，一般 AR 过程的 ACF 函数呈单边递减或阻尼振荡，所以用 PACF 函数判别(从 p 阶开始的所有偏自相关系数均为 0)。

### (3) 平稳条件

一阶： $|\varphi_1| < 1$ 。二阶： $\varphi_1 + \varphi_2 < 1$ 、 $\varphi_1 - \varphi_2 < 1$ 、 $|\varphi_2| < 1$ 。 $\varphi$  越大，自回归过程的波动影响越持久。

### (4) 模型意义

仅通过时间序列变量的自身历史观测值来反映有关因素对预测目标的影响和作用，不受模型变量相互独立的假设条件约束，所构成的模型可以消除普通回归预测方法中由于自变量选择、多重共线性等造成的困难。

## 2. 移动平均 MA(q) 模型

### (1) 模型形式

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p}$$

### (2) 模型含义

用过去各个时期的随机干扰或预测误差的线性组合来表达当前预测值。

AR(p) 的假设条件不满足时可以考虑用此形式。

总满足平稳条件，因其中参数  $\theta$  取值对时间序列的影响没有 AR 模型中参数  $p$  的影响强烈，即这里较大的随机变化不会改变时间序列的方向。

### (3) 识别条件

当  $k > q$  时，有自相关系数  $r_k = 0$  或自相关系数  $r_k$  服从  $N(0, 1/n(1+2\sum r_i^2)^{1/2})$  且  $(|r_k| > 2/n^{1/2}(1+2\sum r_i^2)^{1/2})$  的个数  $\leq 4.5\%$ ，即平稳时间序列的自相关系数  $r_k$  为  $q$  步截尾，偏相关系数  $\varphi_k$  逐步衰减而不截尾，则序列是 MA(q) 模型。

实际中，一般 MA 过程的 PACF 函数呈单边递减或阻尼振荡，所以用 ACF 函数判别(从 q 阶开始的所有自相关系数均为 0)。

### (4) 可逆条件

一阶： $|\theta_1| < 1$ 。二阶： $|\theta_2| < 1$ 、 $\theta_1 + \theta_2 < 1$ 。

当满足可逆条件时，MA(q) 模型可以转换为 AR(p) 模型

## 3. 自回归移动平均 ARMA(p,q) 模型

### (1) 模型形式

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

式中符号： $p$  和  $q$  是模型的自回归阶数和移动平均阶数；

$\varphi$  和  $\theta$  是不为零的待定系数； $\varepsilon_t$  独立的误差项；

$y_t$  是平稳、正态、零均值的时间序列。

### (2) 模型含义

使用两个多项式的比率近似一个较长的 AR 多项式，即其中  $p+q$  个数比 AR(p) 模型中阶数  $p$  小。前二种模型分别是该种模型的特例。

一个 ARMA 过程可能是 AR 与 MA 过程、几个 AR 过程、AR 与 ARMA 过程的迭加，也可能是测度误差较大的 AR 过程。

### (3) 识别条件

平稳时间序列的偏相关系数  $\varphi_k$  和自相关系数  $r_k$  均不截尾，但较快收敛到 0，则该时间序列可能是 ARMA(p,q) 模型。实际问题中，

多数要用此模型。因此建模解模的主要工作是求解  $p$ 、 $q$  和  $\varphi$ 、 $\theta$  的值，检验  $\varepsilon_t$  和  $y_t$  的值。

#### (4) 模型阶数

**AIC 准则：**最小信息准则，同时给出 ARMA 模型阶数和参数的最佳估计，适用于样本数据较少的问题。目的是判断预测目标的发展过程与哪一随机过程最为接近。因为只有当样本量足够大时，样本的自相关函数才非常接近母体的自相关函数。具体运用时，在规定范围内使模型阶数从低到高，分别计算 AIC 值，最后确定使其值最小的阶数是模型的合适阶数。

模型参数最大似然估计时  $AIC=(n-d)\log\sigma^2+2(p+q+2)$

模型参数最小二乘估计时  $AIC=n\log\sigma^2+(p+q+1)\log n$

式中： $n$  为样本数， $\sigma^2$  为拟合残差平方和， $d$ 、 $p$ 、 $q$  为参数。

其中： $p$ 、 $q$  范围上线是  $n$  较小时取  $n$  的比例， $n$  较大时取  $\log n$  的倍数。

实际应用中  $p$ 、 $q$  一般不超过 2。

### 4. 自回归综合移动平均 ARIMA( $p,d,q$ )模型

#### (1)模型识别

平稳时间序列的偏相关系数  $\varphi_k$  和自相关系数  $r_k$  均不截尾，且缓慢衰减收敛，则该时间序列可能是 ARIMA( $p,d,q$ )模型。

#### (2)模型含义

模型形式类似 ARMA ( $p,q$ ) 模型，但数据必须经过特殊处理。特别当线性时间序列非平稳时，不能直接利用 ARMA ( $p,q$ ) 模型，但可以利用有限阶差分使非平稳时间序列平稳化，实际应用中  $d$  一般不超过 2。

若时间序列存在周期性波动，则可按时间周期进行差分，目的是将随机误差有长久影响的时间序列变成仅有暂时影响的时间序列。

即差分处理后新序列符合 ARMA( $p,q$ ) 模型，原序列符合 ARIMA( $p,d,q$ )模型。

### 3.3.3 建模解模过程

#### 1. 数据检验

检验时间序列样本的平稳性、正态性、周期性、零均值，进行必要的数据处理变换。

##### (1)作直方图：检验正态性、零均值。

按图形 Graphs—直方图 Histogram 的顺序打开如图 3.15 所示的对话框。

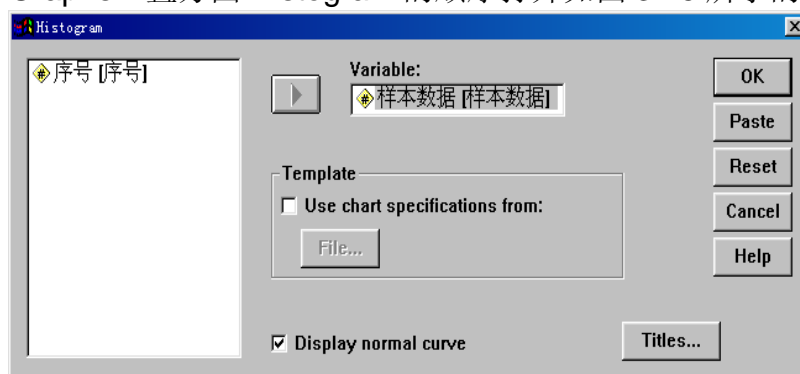
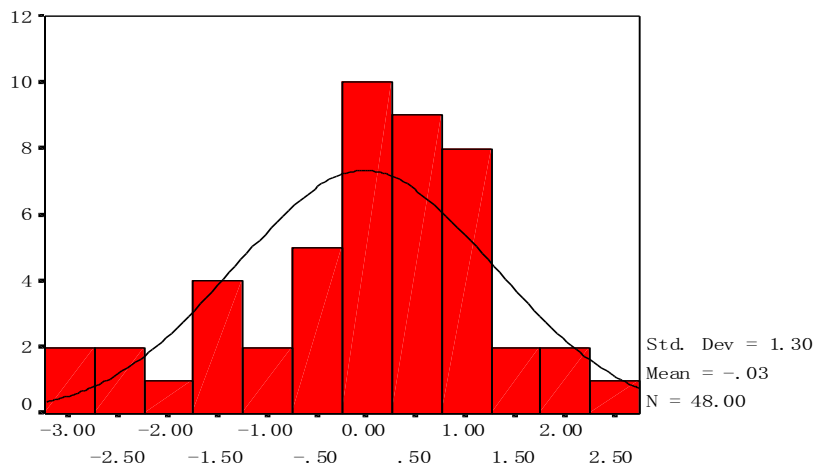


图 3.15

将样本数据送入变量 Variable 框，选中显示正态曲线 Display normal curve 项，点击 OK 运行，输出带正态曲线的直方图，如图 3.16 所示。



样本数据

图 3.16

从图中看出：标准差不为 1、均值近似为 0，可能需要进行数据变换。  
(2)作相关图：检验平稳性、周期性。

按图形 Graphs—时间序列 Time Series—自相关 Autocorrelations 的顺序打开如图 3.17 所示的对话框。

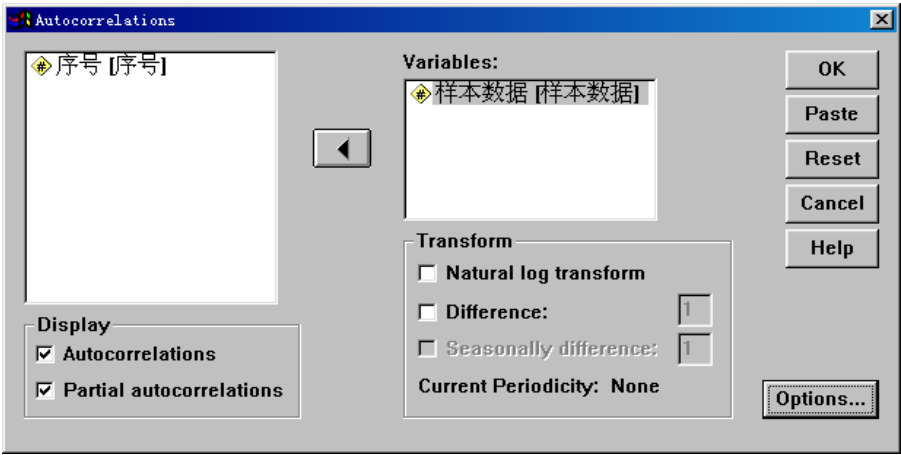


图 3.17

将样本数据送入变量 Variable 框，选中自相关 Autocorrelations 和偏自相关 Partial Autocorrelations 项，暂不选数据转换 Transform 项，点击设置项 Options，出现如图 3.18 所示对话框。

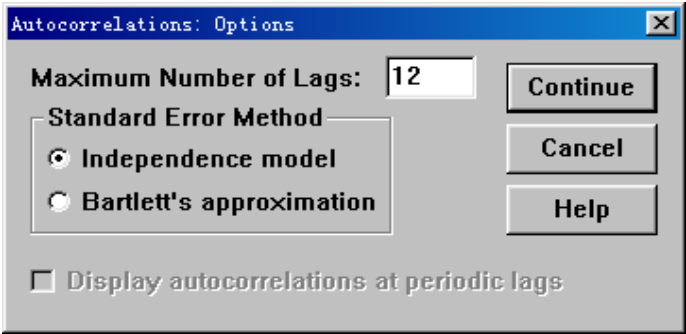


图 3.18

因为一般要求时间序列样本数据  $n > 50$ ，滞后周期  $k < n/4$ ，所以此处控制最大滞后数值 Maximum Number of Lags 设定为 12。点击继续 Continue 返回自相关主对话框后，点击 OK 运行系统，输出自相关图如图 3.19 所示。

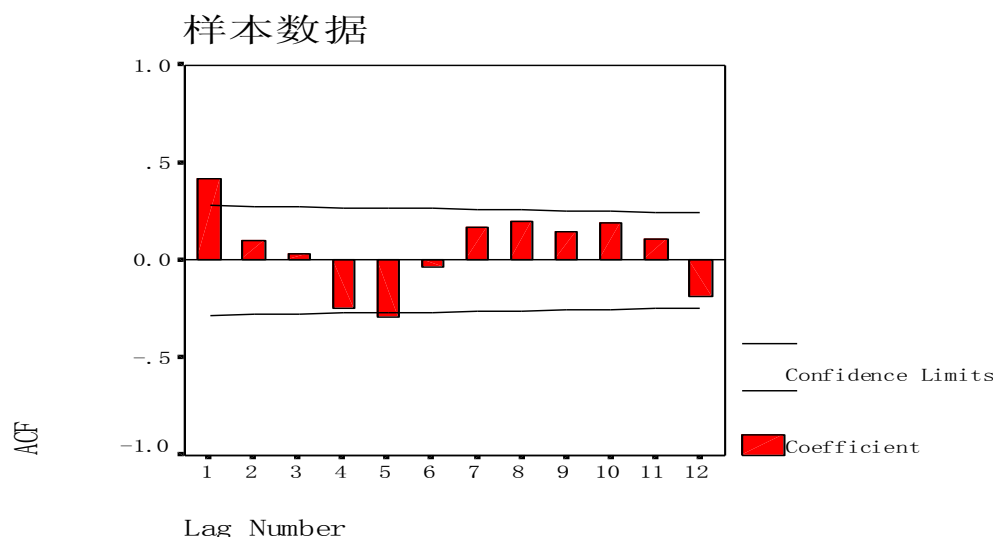


图 3.19

从图中看出；样本序列数据的自相关系数在某一固定水平线附近摆动，且按周期性逐渐衰减，所以该时间序列基本是平稳的。

### (3)数据变换：

若时间序列的正态性或平稳性不够好，则需进行数据变换。常用有差分变换(利用 transform—Create Time Series)和对数变换(利用 Transform—Compute)进行。一般需反复变换、比较，直到数据序列的正态性、平稳性等达到相对最佳。

## 2. 模型识别

分析时间序列样本，判别模型的形式类型，确定  $p$ 、 $d$ 、 $q$  的阶数。

### (1)判别模型形式和阶数

#### ①相关图法：

运行自相关图后，出现自相关图（图 3.19）和偏自相关图（图 3.20）。

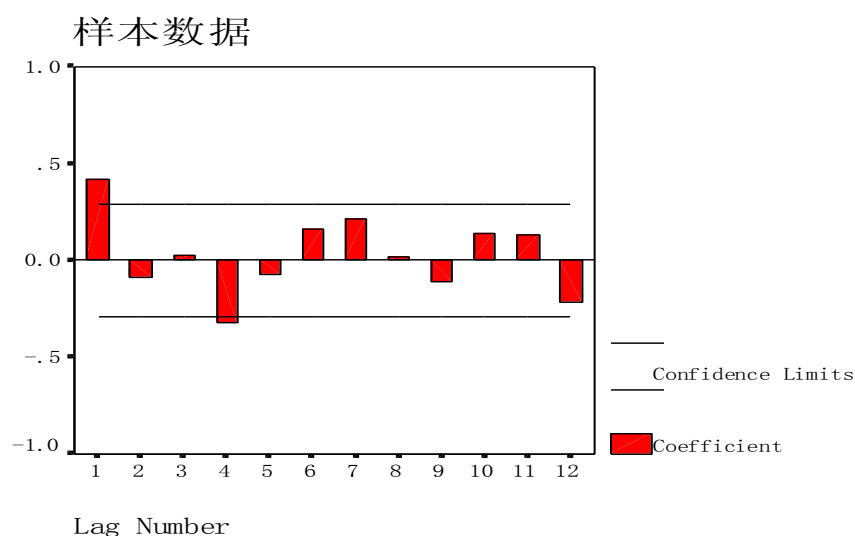


图 3.20

从图中看出：自相关系数和偏相关系数具有相似的衰减特点：衰减快，相邻二个值的相关系数约为 0.42，滞后二个周期的值的相关系数接近 0.1，滞后三个周期的值的相关系数接近 0.03。所以，基本可以确定该时间序列为 ARMA (p,q) 模型形式，但还不能确定是 ARMA (1,1) 或是 ARMA (2,2) 模型。但若前四个自相关系数分别为 0.40、0.16、0.064、0.0256，则可以考虑用 AR(1) 模型。

另外，值得说明的是：只是 ARMA 模型需要检验时间序列的平稳性，若该序列的偏自相关函数具有显著性，则可以直接选择使用 AR 模型。

实际上，具体应用自相关图进行模型选择时，在观察 ACF 与 PACF 函数中，应注意的关键问题是：函数值衰减的是否快；是否所有 ACF 之和为-0.5，即进行了过度差分；是否 ACF 与 PACF 的某些滞后项显著和容易解释的峰值等。但是，仅依赖 ACF 图形进行时间序列的模型识别是比较困难的。

## ②参数估计：

从(m,m-1)开始试验，一般到  $m=p+q=1/n$ 。实际应用中，往往从(1,1)、.....、(2,2)，逐个计算比较它们的 AIC 值（或 SBC 值），取其值最小的确定为模型。

## (2)建立时间序列新变量

无论是哪种模型形式，时间序列总是受自身历史数据序列变化的影响，因此需将历史数据序列作为一个新的时间序列变量。

按数据转换 transform—建立时间序列 Create Time Series 的顺序展开对话框，图 3.21。

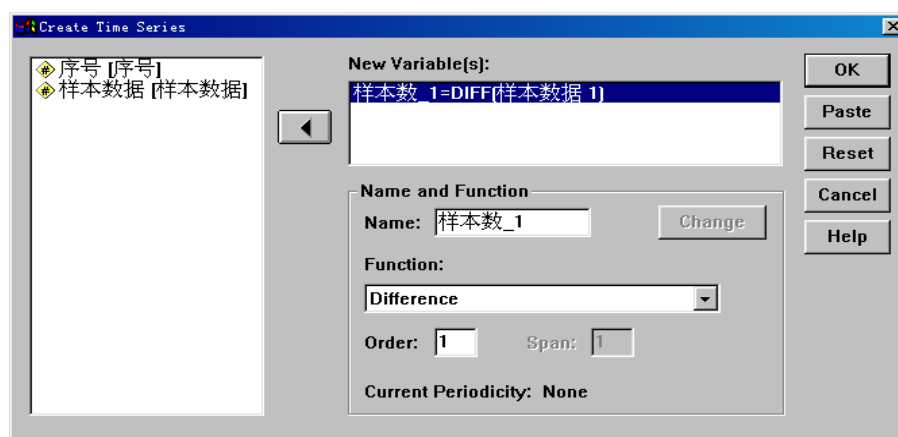


图 3.21

## ①在功能 Function 下拉框中选择变量转换的函数，其中：

非季节差分 Differences: 计算时间序列连续值之间的非季节性差异。

季节性差分 Seasonal Differences: 计算时间序列跨距间隔恒定值之间的季节性差异，跨距根据定义的周期确定。

领先移动平均 Prior moving average: 计算先前的时间序列数值的平均值。

中心移动平均 Centered moving average: 计算围绕和包括当前值的时间序列数值的平均值。

中位数 Running medians: 计算围绕和包括当前值的时间序列的中位数。

累积和 Cumulative sum: 计算直到包括当前值的时间序列数值的累计总数。

滞后顺序 Lag: 根据指定的滞后顺序，计算在前观测量的值。

领先顺序 Lead: 根据指定的领先顺序，计算连续观测量的值。

平滑 Smoothing: 以混合数据平滑为基础，计算连续观测量的值。

以上各项主要用在生成差分变量、滞后变量、平移变量，并且还要关注差分、滞后、平移的次数，以便在建立模型、进行参数估计时，使方程达到一致。

②在顺序 **Order** 框中填入在前或在后的时间序列数值间隔的数目。

在新变量 **New Variable** 框中接受左边框移来的源变量。

在名称 **Name** 框中定义新变量的名称，但必单击改变 **Change** 方能成立。

③单击 **OK** 运行系统，在原数据库中出现新变量列。

另外，若需产生周期性时间序列的日期型变量，则按数据 **Data**—定义日期 **Define Dates** 的顺序展开如图 3.22 所示对话框。

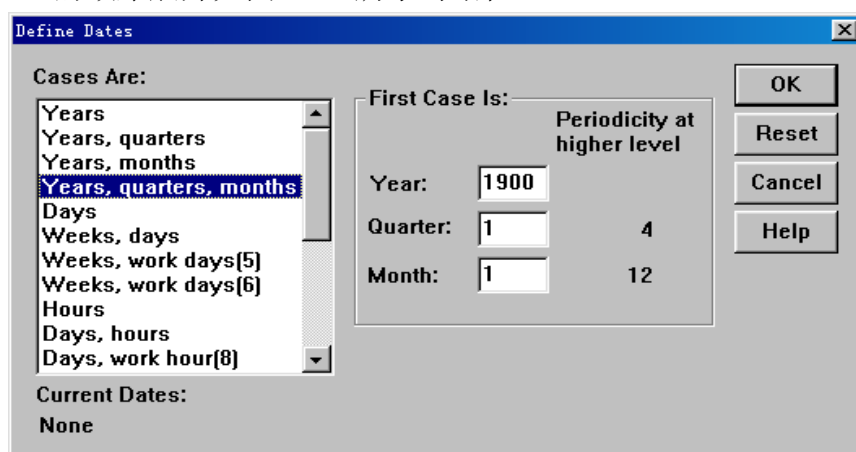


图 3.22

在样本 **Cases Are** 栏中选择定义日期变量的时间间隔，在起始日期 **First Case Is** 栏中设定日期变量第一个观测量的值，单击 **OK** 完成定义。

### 3. 参数估计

采用最大似然估计或最小二乘估计等方法估计  $\phi$ 、 $\theta$  参数值，并进行显著性检验。

按分析 **Analyze**—时间序列 **Time series**—ARIMA 模型的顺序展开如图 3.23 对话框。

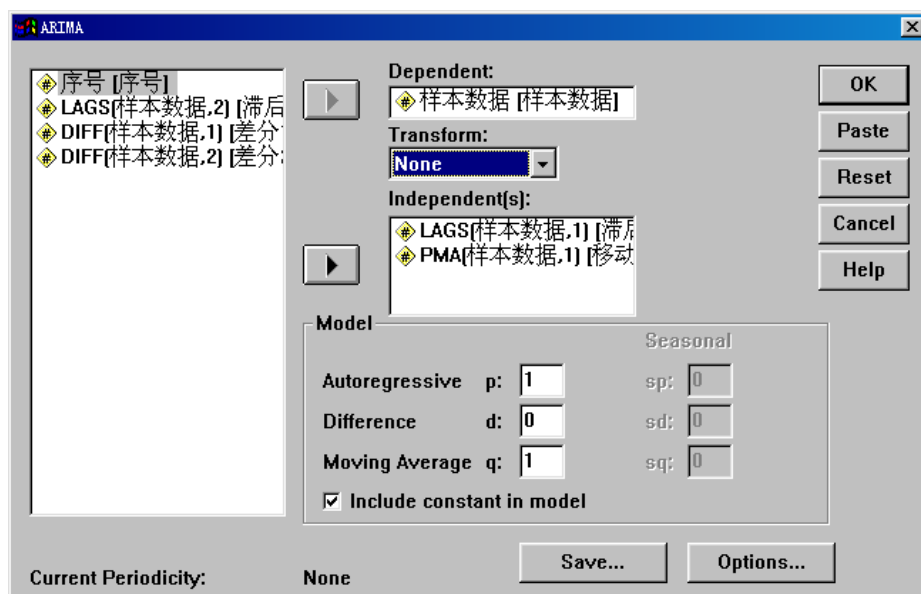


图 3.23

在图 3.23 中：

选择原时间序列变量进入因变量框；  
根据模型识别结果和建立的新时间变量，选择一个或多个变量进入自变量框；暂时不进行因变量的数据转换；  
与自变量的选择对应，根据模型识别结果或实验的思路设定  $p$ 、 $(d)$ 、 $q$  的值；选择模型中包含常数项；  
分别单击保存和设置按钮，展开如图 3.24 和 3.25 对话框。

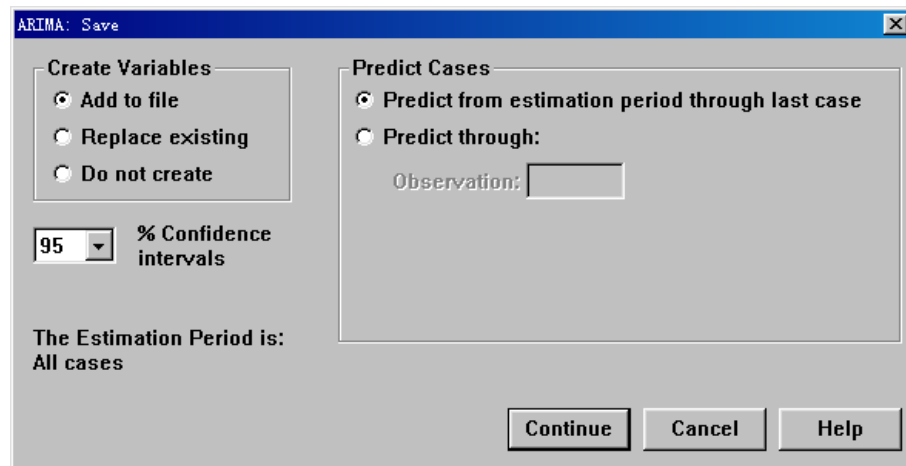


图 3.24

图 3.24 中：

在建立变量 **Create Variable** 栏选择新建变量结果暂存原数据文件 **Add to file** 项，也可选择用新建变量代替原数据文件中计算结果 **Replace existing** 项；  
在设定置信区间百分比 **%Confidence Intervals** 下拉框选择 **95**；  
在预测样本 **Predict Cases** 栏选择根据时期给出预测结果的方法。

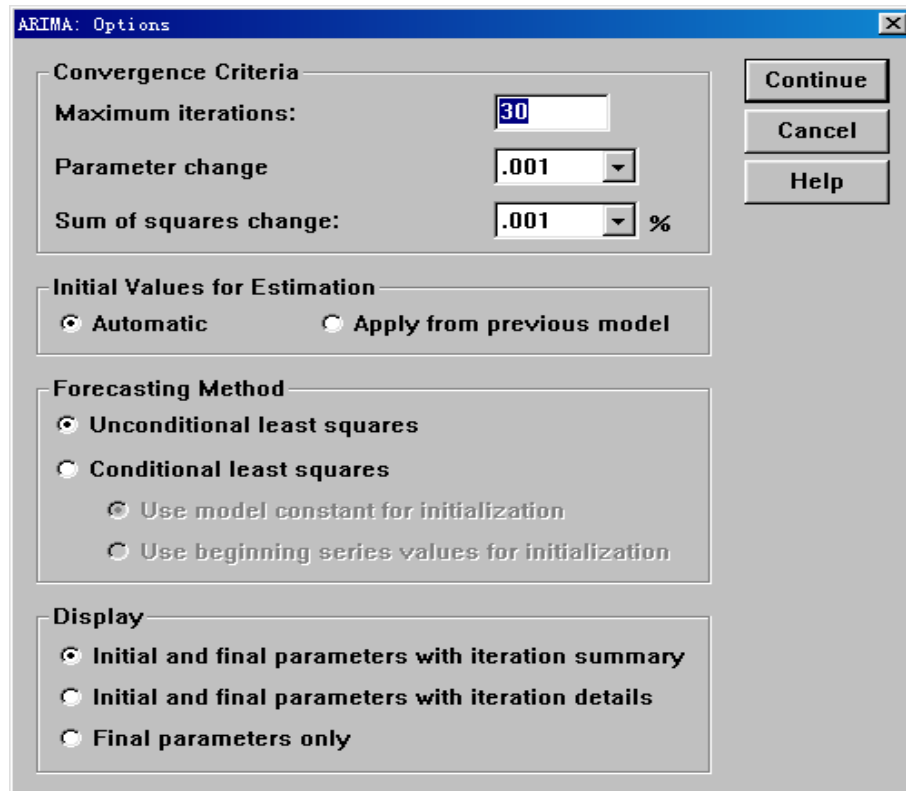


图 3.25

图 3.25 中:

在收敛标准 **Convergence Criteria** 栏选择迭代次数 **Maximum iterations**、参数变化精度 **Parameter change**、平方和变化精度 **Sum of squares change**，当运算达到其中一个参数的设定，则迭代终止；在估计初始值 **Initial Values for Estimation** 栏选择由过程自动选择 **Automatic** 或由先前模型提供 **Apply from previous model**，一般默认前者；在预测方法 **Forecasting Method** 栏选择无条件 **Unconditional** 或有条件最小二乘法 **Conditional least squares**；在输出控制 **Display** 栏选择最初和最终参数的迭代摘要 **Initial and final parameters with iteration summary** 或详细资料 **details**、或只显示最终参数 **Final parameters only**。

单击 **OK**，系统立即执行，输出信息如下：

MODEL: MOD\_1

Split group number: 1 Series length: 48

No missing data.

Melard's algorithm will be used for estimation.

Conclusion of estimation phase.

Estimation terminated at iteration number 7 because:Sum of squares decreased by less than .001 percent.

FINAL PARAMETERS:

Number of residuals 48

Standard error	1.1996949
Log likelihood	-75.463915
AIC	156.92783
SBC	162.54143

#### Analysis of Variance:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	45	65.099923	1.4392678

#### Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	.02318739	.31945836	.0725835	.94245925
MA1	-.44871554	.28829314	-1.5564558	.12660552
CONSTANT	-.02421308	.25505018	-.0949346	.92478827

The following new variables are being created:

Name	Label
FIT_1	Fit for 样本数据 from ARIMA, MOD_1 CON
ERR_1	Error for 样本数据 from ARIMA, MOD_1 CON
LCL_1	95% LCL for 样本数据 from ARIMA, MOD_1 CON
UCL_1	95% UCL for 样本数据 from ARIMA, MOD_1 CON
SEP_1	SE of fit for 样本数据 from ARIMA, MOD_1 CON

各个输出统计量的意义:

**常数项:** 认为是取值恒为 1 的常数变量, 其系数就是自变量为 0 时因变量的最优预测值, 也称为预测基准值。

**系数:** 反映自变量对因变量影响的权重。

**标准误:** 表明样本数据的可靠性。在(残差)参数近似服从正态分布条件下, 系数加减两倍的标准误差近似等于总体参数 95%的置信区间。其值越小, 置信区间越窄; 并且其对于系数的相对值越小, 估计结果越精确。

**t 统计量:** 估计系数与标准误差的比值, 检验变量的不相关性。一般给定 5%显著水平, 则拒绝原假设的 0 值位于 95%的置信区间外, 其绝对值必大于 2。

**t 概率值:** 其值越小, 则拒绝原假设不相关性的证据越充分。其值接近 0.05 与 t 统计量接近 2 相对应。

**均值:** 度量变量的集中度, 传递随机变量的位置信息。

**标准差:** 度量变量的离散度, 传递随机变量的规模信息。

**平方和:** 残差平方和是许多统计量的组成部分, 孤立考察无太大价值。

**准则:** 信息准则 AIC 和 SBC 用于模型的选择, 越小越好, 但受自由度约束较为严重。

**R<sup>2</sup> 校正:** 是模型中自变量对因变量变动的解释比例, 度量方程预测因变量的成功程度, 其是回归标准误差与因变量标准差比较的结果。另一个比较方法是回归标准误差不超过因变量

均值的 10%则为好的模型。

DW 统计：用于检验随机误差项是否存在序列相关。

LN 似然：用于模型比较和假设检验，越大越好。

残差图：

#### 4. 模型检验

检验新建模型的合理性。若检验不通过，则调整(p,q)值，重新估计参数和检验，反复进行直到接受为止。但模型识别、参数估计、检验修正三个过程之间相互作用、相互影响，有时需要交叉进行、反复实验，才能最终确定模型形式。

##### (1)相关图检验残差白噪声：

因为白噪声过程是序列无关的，所以白噪声过程的自相关函数和偏自相关函数在自相关图中均为等于 0 的水平直线。

##### (2)散点图检验残差独立性：

以误差值为纵坐标、以预测值为横坐标，观察散点分布的均匀性、随机性。

理想预测模型的预测误差一定是不可预测的、无规律的、序列无关的。

相应的 DW 统计量仅适用检验一阶序列。

##### (3)直方图检验残差零均值：

零均值仅检验残差序列无关，若正态分布则检验独立性。

##### (4)概率图检验残差自相关：以显著性水平 0.05 计算 $\chi^2()$ 概率值，。

##### (5)均方差检验预测的效果：以预测误差的均方差最小为标准，注意预测误差仅与预测周期有关，而与起始时刻无关。

#### 5. 模型预测

预测系统研究对象的未来某时刻状态。列出预测模型，计算预测值。