第十三章 时间序列分析

资料来源 ：网络资料

第一节 时间序列的组成部分

一、时间序列数据

29.1pt">1. 数据类型：截面数据与时间序列数据

人们对统计数据往往可以根据其特点从两个方面来切入，以简化分析过程。一个是研究所谓横截面(cross section)数据，也就是对大体上同时，或者和时间无关的不同对象的观测值组成的数据。

另一个称为时间序列(time series)，也就是由对象在不同时间的观测值形成的数据。

前面讨论的模型多是和横截面数据有关。这里将讨论时间序列的分析。我们将不讨论更加复杂的包含这两方面的数据。

2. 时间序列和回归

时间序列分析也是一种回归。

回归分析的目的是建立因变量和自变量之间关系的模型；并且可以用自变量来对因变量进行预测。通常线性回归分析因变量的观测值假定是互相独立并且有同样分布。

而时间序列的最大特点是观测值并不独立。时间序列的一个目的是用变量过去的观测值来预测同一变量的未来值。也就是说，时间序列的因变量为变量未来的可能值，而用来预测的自变量中就包含该变量的一系列历史观测值。

当然时间序列的自变量也可能包含随着时间度量的独立变量。

例如教材中的tssales.sav数据，就是一个时间序列的数据例子。这是某企业从1990年1月到2002年12月的销售数据(tssales.sav)。我们希望能够从这个数据找出一些规律，并且建立可以对未来的销售额

课程资料

进行预测的时间序列模型。

利用点图则可以得到对该数据更加直观的印象：

从这个点图可以看出。总的趋势是增长的，但增长并不是单调上升的；有涨有落。大体上看，这种升降不是杂乱无章的，和季节或月份的周期有关系。当然，除了增长的趋势和季节影响之外，还有些无规律的随机因素的作用。

二、时间序列的组成部分

从该例可以看出，该时间序列可以有三部分组成：趋势(trend)、季节(seasonal)成分和无法用趋势和季节模式解释的随机干扰(disturbance）。

例中数据的销售就就可以用这三个成分叠加而成的模型来描述。

一般的时间序列还可能有循环或波动(Cyclic, or fluctuations)成分；循环模式和有规律的季节模式不同，周期长短不一定固定。比如经济危机周期，金融危机周期等等。

一个时间序列可能有趋势、季节、循环这三个成分中的某些或全部再加上随机成分。因此，

如果要想对一个时间序列本身进行较深入的研究，把序列的这些成分分解出来、或者把它们过虑掉则会有很大的帮助。

如果要进行预测，则最好把模型中的与这些成分有关的参数估计出来。

就例中的时间序列的分解，通过SPSS软件，可以很轻而易举地得到该序列的趋势、季节和误差成分。

下图表示了去掉季节成分，只有趋势和误差成分的序列。

下图用两条曲线分别描绘了趋势成分和季节成分。

下图用两条曲线分别描绘了趋势成分和误差成分。

第二节 指数平滑

一、指数平滑的基本原理

如果我们不仅仅满足于分解现有的时间序列，而且想要对未来进行预测，就需要建立模型。首先，这里介绍比较简单的指数平滑(exponential smoothing)。

指数平滑只能用于纯粹时间序列的情况，而不能用于含有独立变量时间序列的因果关系的研究。

指数平滑的原理为：当利用过去观测值的加权平均来预测未来的观测值时（这个过程称为平滑），离得越近的观测值要给以更多的权。

而“ 指数 ”意味着：按照已有观测值“ 老 ”的程度，其上的权数按指数速度递减。

二、指数平滑的基本方法

以简单的没有趋势和没有季节成分的纯粹时间序列为例，指数平滑在数学上这实际上是一个几何级数。这时，如果用Yt表示在t时间的平滑后的数据（或预测值），而用X1, X2, …, Xt表示原始的时间序列。那么指数平滑模型为：

自然，这种在简单情况下导出的公式（如上面的公式）无法应对具有各种成分的复杂情况。

根据数据，可以得到这些模型参数的估计以及对未来的预测。在和我们例子有关的指数平滑模型中，需要估计12个季节指标和三个参数（包含前面公式权重中的，和趋势有关的，以及和季节指标有关的）。

在简单的选项之后，SPSS通过指数平滑产生了对2003年一年的预测。下图为原始的时间序列和预测的时间序列（光滑后的），其中包括对2003年12个月的预测。图下面为误差。

第三节 Box—Jenkins方法：ARIMA模型

一、ARIMA模型介绍

1. ARIMA模型结构

如果要对比较复杂的纯粹时间序列进行细致的分析，指数平滑往往是无法满足要求的。

而若想对有独立变量的时间序列进行预测，指数平滑更是无能为力。

于是需要更加强有力的模型。这就是下面要介绍的Box-Jenkins ARIMA模型。

数学上，指数平滑仅仅是ARIMA模型的特例。

比指数平滑要有用和精细得多的模型是Box-Jenkins引入的ARIMA模型。或称为整合自回归移动平均模型(ARIMA 为Autoregressive Integrated Moving Average一些关键字母的缩写)。该模型的基础是自回归和移动平均模型或ARMA(Autoregressive and Moving Average) 模型。

它由两个特殊模型发展而成，一个特例是自回归模型或AR (Autoregressive) 模型。假定时间序列用X1, X2, …, Xt表示，则一个纯粹的AR (p)模型意味着变量的一个观测值由其以前的p个观测值的线性组合加上随机误差项at（该误差为独立无关的）而得：

这看上去象自己对自己回归一样，所以称为自回归模型；它牵涉到过去p个观测值（相关的观测值间隔最多为p个）。

ARMA模型的另一个特例为移动平均模型或MA (Moving Average) 模型，一个纯粹的MA (q)模型意味着变量的一个观测值由目前的和先前的q个随机误差的线性的组合：

显然ARMA(p,0)模型就是AR (p)模型，而ARMA(0,q)模型就是MA(q)模型。这个一般模型有p+q个参数要估计，看起来很繁琐，但利用计算机软件则是常规运算；并不复杂。

2. ARIMA模型的平稳性和可逆性

要想ARMA(p,q)模型有意义则要求时间序列满足平稳性(stationarity)和可逆性(invertibility)的条件，

这意味着序列均值不随着时间增加或减少，序列的方差不随时间变化，另外序列本身相关的模式不改变等。

一个实际的时间序列是否满足这些条件是无法在数学上验证的，

这没有关系，但可以从下面要介绍的时间序列的自相关函数和偏相关函数图中可以识别出来。

一般人们所关注的的有趋势和季节/循环成分的时间序列都不是平稳的。这时就需要对时间序列进行差分(difference)来消除这些使序列不平稳的成分，而使其变成平稳的时间序列，并估计ARMA模型，估计之后再转变该模型，使之适应于差分之前的序列（这个过程和差分相反，所以称为整合的(integrated)ARMA模型），得到的模型于是称为ARIMA模型。

3. ARIMA模型：差分

差分是什么意思呢？差分可以是每一个观测值减去其前面的一个观测值，即Xt-Xt-1。这样，如果时间序列有一个斜率不变的趋势，经过这样的差分之后，该趋势就会被消除了。

当然差分也可以是每一个观测值减去其前面任意间隔的一个观测值；比如存在周期固定为s的季节成分，

那么相隔s的差分 为Xt-Xt-s就可以把这种以s为周期的季节成分消除。

对于复杂情况，可能要进行多次差分，才能够使得变换后的时间序列平稳。

二、ARIMA模型的识别和估计

上面引进了一些必要的术语和概念。下面就如何识别模型进行说明。

要想拟合ARIMA模型，必须先把它利用差分变成ARMA(p,q)模型，并确定是否平稳，然后确定参数p,q。

现在利用一个例子来说明如何识别一个AR(p)模型和参数p。

由此MA(q)及ARMA(p,q)模型模型可用类似的方法来识别。

根据ARMA(p,q)模型的定义,它的参数p,q和自相关函数(acf，autocorrelations function)及偏自相关函数(pacf，partial autocorrelations function)有关。

自相关函数描述观测值和前面的观测值的相关系数；

而偏自相关函数为在给定中间观测值的条件下观测值和前面某间隔的观测值的相关系数。

这里当然不打算讨论这两个概念的细节。引进这两个概念主要是为了能够了解如何通过研究关于这两个函数的acf和pacf图来识别模型。

三、用ARIMA模型拟合带有独立变量的时间序列。

用ARIMA模型拟合带有独立变量的时间序列

从各种角度来看拟合带独立变量平方的ARIMA(2,1,2)(0,1,1)7模型给出更好的结果。

虽然从上面的acf和pacf图看不出（一般也不应该看出）独立变量对序列的自相关性的影响，但是根据另外的一些判别准则，独立变量的影响是显著的，而且加入独立变量使得模型更加有效。

用ARIMA模型拟合带有独立变量的时间序列

要注意，一些独立变量的效果也可能是满足某些时间序列模型的，也可能会和季节、趋势等效应混杂起来不易分辩。这时，模型选择可能就比较困难。也可能不同模型会有类似的效果。

一个时间序列在各种相关的因素影响下的模型选择并不是一件简单明了的事情。实际上没有任何统计模型是绝对正确的，它们的区别在于，在某种意义上，一些模型的某些性质可能要优于另外一些。

思考题：

1. 举例说明时间序列的各组成部分。
2. 请用简洁的语言说明指数平滑的基本思路。
3. 时间序列分析与一般的简单回归分析有何不同？
4. 请简要说明ARIMA模型的基本思想。

数据挖掘就业前景

推荐站点：免费下载神经网络、遗传算法、人工智能源程序、源代码 ,我的关联规则综述 ,向大伙推荐一个相当不错的网站,请weka 高人解惑,java搜索引擎: lucene学习笔记 1 ,需要多少日志才能算是web usage mining

---