

时序分析(9) -- 时序平滑（上）

前面的文章中，我们着重讲述了金融时序分析关于观测值和波动性分析和预测的常见技术。现在我们需要探讨一下时序分析的基本技术：平滑。时序平滑实际上在时序分析中扮演着非常重要的角色。它不但可以提高时序信号的信噪比，还可以帮助我们进行预测。

我们将主要讨论如下的平滑技术：

- 简单移动平均(Simple Moving Average)
- 累积移动平均(Culmulative Moving Average)
- 加权移动平均(Weighted Moving Average)
- 指数移动平均(Exponential Moving Average)

这里的移动平均和前面文章中所讲过的MA模型并不一样，读者很快就会发现其区别。

指数移动平均是现代时序分析技术中较为重要的方法，它有多种形式和模型，我们将会一一介绍。

- 简单移动平均(SMA)

SMA其实就是时序的不加权的均值而已，

$$y_t = \frac{1}{t} \sum_{i=1}^t y_i$$

实际操作中，往往是依据一个固定的窗口大小，当有新的观测值加入时，会移除最旧的那个值。这种方式称为移动窗口。

- 累积移动平均(CMA)

CMA就是用新值来更新所有历史平均值，

$$CMA_n = \frac{y_1 + y_2 + \dots + y_n}{n}$$

$$CMA_{n+1} = \frac{y_{n+1} + n * CMA_n}{n+1}$$

和SMA相比，CMA就是所有历史的平均值，窗口就是所有已知时序。

- 加权移动平均(WMA)

WMA就是按照算术级数递减权重加权平均，

$$WMA_n = \frac{my_n + (m-1)y_{n-1} + \dots + 2y_{n-m+2} + y_{n-m+1}}{m + (m-1) + \dots + 2 + 1}$$

m 为窗口大小。

- 指数移动平均(EWMA)

在详细探讨指数权重移动平均以前，我们简单回顾一下时序预测技术的发展进程。

最原始的想法就是假如我们对时序没有任何先验知识，只指导当前时刻的观测值，我们如何预测下一时刻的观测值呢？很简单，就以当前观测值作为预测值：

$$\hat{y}_{T+h|T} = y_T,$$

这种称为naive method的方法就是认为只有最近的观测值提供了关于未来的信息，而之前的历史值不提供任何信息，所以它把所有的权重都放在最后一个已知的观测值上。

另外一个思路就是认为所有的历史数据都对预测有帮助，但我们并不知道该如何分配权重，所以就给所有历史数据分配同样的权重，得到

$$\hat{y}_{T+h|T} = \frac{1}{T} \sum_{t=1}^T y_t,$$

显而易见，这两种思路都比较极端，是否有一种方法可以平衡这两种方法呢？一个比较自然的思路就是给予比较近的观测值比较高的权重，权重按照指数衰减，这就是简单指数平滑背后的思路。

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \cdots \quad (9.1)$$

这里 $0 \leq \alpha \leq 1$ 是平滑参数，它控制了衰减的速度。

我们把上式稍加变换就可以得到

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1} \quad (9.2)$$

可以理解为对 $t + 1$ 时刻的预测就等于最近的时序观测值和最近的时序预测值的加权平均。

鉴于指数平滑在时序分析上的重要性，我们需要从多个维度上理解，下面我们来介绍其几种表现形式：

1. Component Form

可以认为，时序的预测就是得到时序的水平(level)，以 ℓ 表示，有

$$\hat{y}_{t+1|t} = \ell_t \quad (9.3)$$

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} \quad (9.4)$$

我们可以这样理解上面第二个公式，对时序新的水平的预测等于对上一个水平的预测和新的数据的加权平均。从概念上来说，也可以用Bayesian方法来解释。在实际操作中，一般都会给出一个移动窗口来进行估算，这种方式的另外一个名字就是指数权重移动平均(EWMA)

2. Error Correction Form

简单变换一下(9.4)式，可以得到

$$\begin{aligned} \ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned} \quad (9.5)$$

这里的 e_t 是样本内误差项，上式可以理解为对新的水平的预测等于上一次的水平预测加上最近一次的预测误差修正。

下一篇我们将讲述指数平滑技术的其他变种

未完待续.....