金融时间序列及其特征（一）：资产收益率

- 大家好！这篇文章是金融后院新专栏——金融计量专栏的开篇，既然是开篇，我们就首先来了解一些这个栏目是什么，学习这个栏目我们需要什么知识，最后通过这个栏目的学习，我们能学习到什么东西。
- 按照MBA智库百科的定义：金融计量学通常就是指对金融市场的计量分析，这里的“计量分析”不仅包括对金融市场各种交易变量(如价格、交易量、波动率等)进行相应的统计分析和计量建模，还包含研究金融市场中大量的可行性方案和基于随机分析框架下实证金融的主要成果。
- 如果你对量化投资、金融建模非常感兴趣，那么金融计量学就是一个很好的学习切入点。值得一提的是，小编本人的金融计量也是处于入门阶段。因此这个专栏不仅仅是向大家分享金融计量的知识，同时也是自己不断学习的历程。
- 最后，欢迎对金融计量、量化投资感兴趣的同学在这个专栏投稿，欢迎前来交流。
- 接下来我们进入正题部分。

首先我们需要知道，金融计量学主要是研究金融时间序列。那针对于金融时间序列，我们到底要研究什么?

- 对于股价的时间序列，我们想研究什么？无非两个，一是股价，二是股价的波动。我们再想一下，不同股票的股价的尺度是不一样的，有的股票1块钱1股，有的股票100块钱1股。那如何才能把这些不同尺度的股价纳入到一个统一的模型中呢？显然不行。怎么办？收益率。在投资中，直观的感受是我们关心投资品的价格变动，而实际上我们关心的是投资的收益率。因此，多数金融研究针对的是资产收益率而非资产价格。Campbell，Lo&MacKinlay（1997）给出了使用收益率的两个主要理由：
  - 对于普通投资者来说，资产的收益率完全体现了该资产的投资机会，且与其投资规模无关；
  - 收益率序列比价格序列更容易处理，因为前者具有更好的统计性质。但是关于资产收益率的定义又有很多种，我们将会在下文提及。
- 另外一个就是波动率，英文是volatility。一个序列的波动率或者说波动度该用什么衡量呢?

1. 条件标准差。使用历史价格的标准差衡量波动度，这也是最常用的一种方法；
2. 使用期权的价格倒退期权标的资产价格的条件标准差σ_t，这种波动率被称为隐含波动率（Implied volatility），往往反映着交易者的“恐慌心理”（援引宋豪漳老师）。芝加哥交易所的VIX波动率指数就是隐含波动率指数。
3. 通过高频数据计算出的波动率。例如我们考虑万科股票的日收益率，因为一个交易日只有一个观测值，所以日波动率不能从收益率中观测出来。如果可以得到一天之内的股票数据，如10分钟的收益率数据，那么我们就可以计算每个十分钟内的所有股票成交价的标准差然后把它们全部加起来，就可以作为该日万科股票收益率的波动率。但是要注意的是，股票的波动率包括日内波动率和隔夜波动率，而交易只在白天发生。隔夜波动率代表不同交易日之间的变化。高频交易数据计算出来的日内收益率只包含有限的隔夜波动率的信息。因此，通过这种方法计算出的波动率的估计的准确性值得仔细研究。

其次学习金融计量我们需要具备什么样的基础?

- 扎实的微积分知识；

- 扎实的概率论、数理统计的知识;
- 基础的计量经济学知识（扎实更好）;
- 一些时间序列的基础知识（没有的话也没问题，有更好）;
- 良好的模型感知力;
- 等等。

关于软件使用

- 主要是使用R、Python以及一些Stata的客串，不过并不需要很好的这些语言的基础也可以很容易入门。
- 这些软件的学习，金融后院也会同步提供，这也的多多依靠澍兄的勤劳写作啦。

通过这个栏目的学习，我们能学习到什么?

- ARMA、ARCH、GARCH模型及其衍生模型的构建;
- 非线性模型极其应用;
- 高频数据分析及市场微结构;
- 连续时间模型及其应用;
- 极值理论、分位数估计、风险值（VaR）;
- 多元时间序列分析;
- 主成分分析与因子模型;
- 多元波动率模型及其应用;
- 状态空间模型和卡尔曼滤波;
- 马尔可夫蒙特卡洛方法及其应用;
- 金融交易策略的实现;
- 金融数据整理;
- 金融数据可视化;
- 等等......我也吹不出来了

下面开始我们今天的正文

资产收益率的衡量

- 首先我们规范一下我们的符号表达:
  设P_t是资产在t时刻的价格，假定资产不支付红利。

单期简单收益率

- 若从第$t-1$天到t天（一个周期）持有某种资产，则简单毛收益率为:

$$1 + R_t = \frac{P_t}{P_{t-1}}$$

  或

$$P_t = P_{t-1}(1 + R_t)$$

- 对应的单期简单净收益率或称简单收益率为:

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}$$

多期简单收益率

- 若从第$t-k$天到第t天这k个周期内持有某种资产，则k期简单毛收益率为：

$$1 + R_t[k] = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}}$$
$$= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})$$
$$= \prod_{j=0}^{k-1}(1 + R_{t-j})$$

- 这样，k期简单毛收益率就是其所包含的着k个单期简单毛收益率的乘积，称为复合收益率，k期简单净收益率是

$$R_t[k] = \frac{P_t - P_{t-k}}{P_{t-k}}$$

- 实际中我们经常需要将收益率年化，如果所持有资产的期限是k年，则平均的年化收益率为：

$$\text{年化的}R_t[k] = [\prod_{j=0}^{k-1}(1 + R_{t-j})]^{1/k} - 1$$

- 通过简单的取对数等价变化，我们可以把上式变成下面这个样子：

$$\text{年化的}R_t[k] = exp[\frac{1}{k}\sum_{j=0}^{k-1} ln(1 + R_{t-j})] - 1$$

- 之所以这样变换的一个重要目的是，计算算术平均值比计算几何平均值要容易的多，并且由于单期收益率一般很小，我们可以用一阶泰勒展开式来近似年化的收益率，得到：

$$\text{年化的}R_t[k] \approx \frac{1}{k}\sum_{j=0}^{k-1} R_{t-j}$$

连续复合收益率

- 1块钱如果按照10%的年利率每年计息一次，则一年之后会变成1.1块钱，如果半年计息一次，则每期的利率为5%，一年之后会变成

$$(1 + \frac{10\%}{2})^2 = 1.1025$$

如果以利率r每年计息m次则一年之后会变成

$$(1 + \frac{r}{m})^m$$

，那么当$m \to +\infty$时，上式为：

$$\lim_{m\to+\infty}(1 + \frac{r}{m})^m$$

$$= \lim_{m \to +\infty} (1 + \frac{r}{m})^{\frac{m}{r} \times r}$$

$$= e^r$$

- 一般地，连续复合的资产净值A为：

$$A = Cexp(r \times n)$$

  其中r为年利率，C为初始资本，n是年数。这也就是连续复利。
- 连续贴现则正好相反：

$$C = Aexp(-r \times n)$$

  叫作n年后价值为A的资产现值，这里我们假定连续复合的年利率为r。
- 下面我们就可以引入连续复合收益率的概念了：
- 资产地简单毛收益率地自然对数称为连续复合收益率或对数收益率(log-return)：

$$r_t = ln(1 + R_t) = ln \frac{P_t}{P_{t-1}} = p_t - p_{t-1}$$

  其中$p_t = lnP_t$，这个式子给我们一个很好地启示是，当我们拿到的是价格序列地时候，我们只需要把序列取对数然后进行差分就可以得到对数收益率序列了。
- 与简单净收益率R_t相比，r_t具有很多良好的价值，首先对于多期收益率，我们有：

$$r_t[k] = ln(1 + R_t[k]) = ln[(1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})]$$

$$= ln(1 + R_t) + ln(1 + R_{t-1}) + \cdots + ln(1 + R_{t-k+1})$$

$$= r_t + r_{t-1} + \cdots + r_{t-k+1}$$

  这样，连续复合多期收益率就是它所包含的连续复合单期收益率之和。其次是对数收益率具有良好的统计性质。

我们可以具体来想一想对一个序列取对数的原因，大概有如下几点：

1. 便于计算；
2. 序列大致呈现指数增长；
3. 为了是数据更趋紧正态分布（为了符合古典线性回归模型的假定）；
4. 为了平稳数据，但是根据小编自己的建模经验，取对数只能平稳序列值均大于1的序列，而对于小于1的序列，不仅不能平稳，反而会增加波动。

资产的组合收益率

- 若一个资产组合由N个资产组成，则该资产组合的简单收益率是它所包含的各个资产的简单收益率的加权平均：

$$R_{p,t} = \sum_{i=1}^{N} \omega_i R_{it}$$

  其中R_{it}是资产i的简单收益率。
- 然而，资产组合的连续复合收益率就没有上述方便的性质了，如果简单收益率R_{it}的绝对值都很小，

则我们有：

$$r_{p,t} \approx \sum_{i=1}^{N} \omega_i r_{it}$$

，其中$r_{p,t}$是该组合在t时刻的连续复合收益率。这种近似经常被用来研究资产组合的收益率。

有红利支付的简单净收益率和连续复合收益率

- 设t时刻分红D_t，则简单净收益率为：

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1$$

- 连续复合收益率为：

$$r_t = ln(P_t + D_t) - ln(P_{t-1})$$

超额收益率

- 一个资产在t时刻的超额收益率是该资产的收益率与某个参考资产的收益率之差。这个参考资产通常是无风险的，例如短期国债的收益率。简单超额收益率为：

$$Z_t = R_t - R_{0t}$$

，对数超额收益率为：

$$z_t = r_t - r_{0t}$$

- 其中R_{0t}和r_{0t}分别是该参考资产的简单收益率和对数收益率。在金融文献中，超额收益率被认为是某个套利组合的赢利。在这个投资组合中，对某资产持多寸头寸而对参考资产持空头头寸，且初始投资净值为0。

今天的推文到这里就算结束啦，文中的公式打的我头晕脑胀。最后我再坚持打几个来作为本文的总结。

总结

- 简单收益率R_t和r_t的关系是：

$$r_t = ln(1 + R_t)$$

$$R_t = e^{r_t} - 1$$

- 如果收益率R_t和r_t是百分比，则：

$$r_t = 100ln(1 + R_t/100)$$

$$R_t = 100(e^{r_t/100} - 1)$$

- 收益率的时间累加使得：

$$1 + R_t[k] = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})$$

$$r_t[k] = r_t + r_{t-1} + \cdots + r_{t-k+1}$$

- 如果连续复合年利率为r，则资产的现值和未来价值之间的关系为：

$$A = Cexp(r \times n)$$

$$C = Aexp(-r \times n)$$