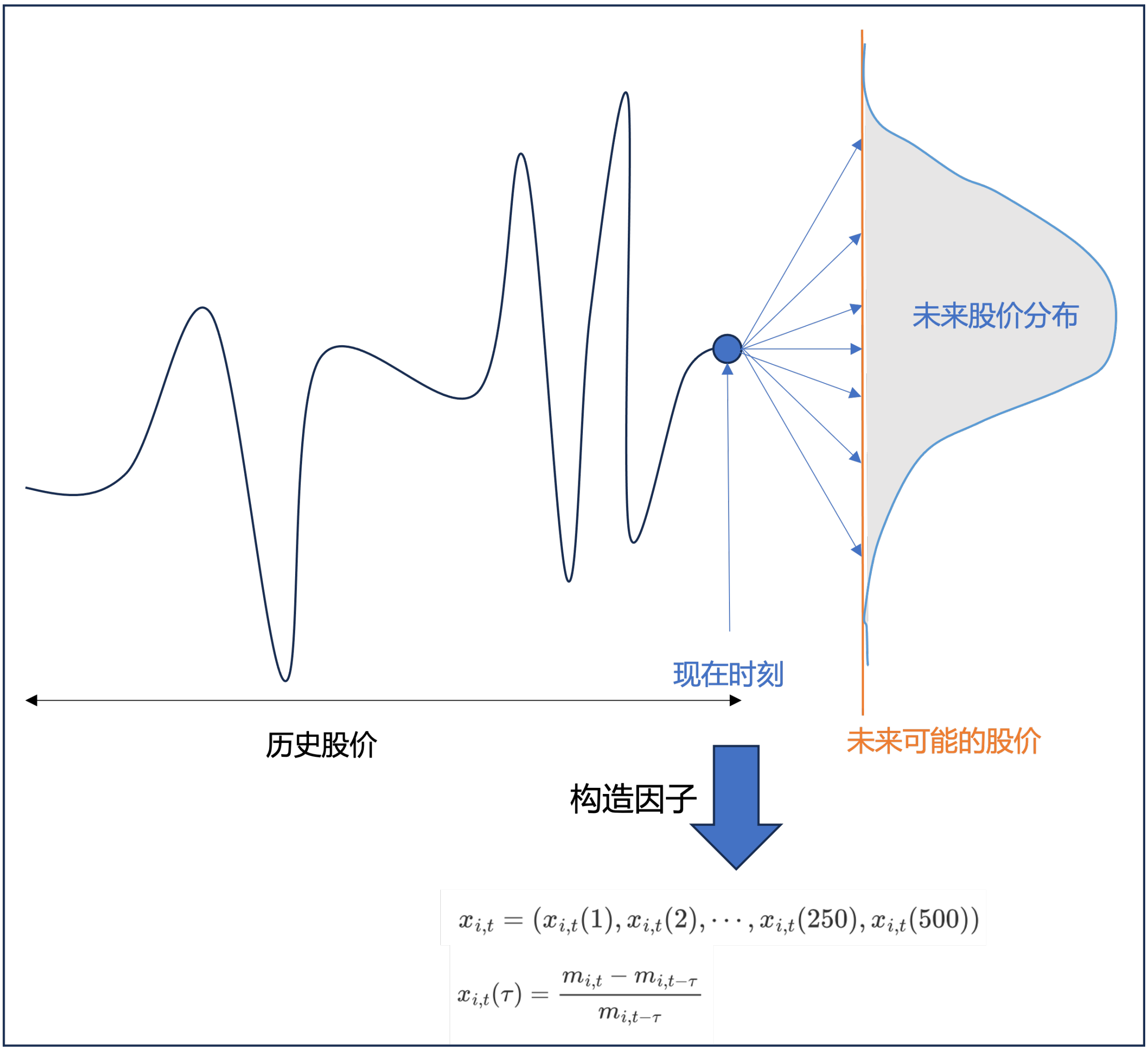
量化投资量化交易技术研发记录

寻找股价变化率期限结构的统计相关性

这个部分的目的本质上是希望通过近一段时间的股价曲线来获取接下来某个时间长度内股价变化率的分布，例如一周股价变化率的分布。为了能够充分利用不同的股票进行参数估计，即我们先假设这种统计相关性与单个股票无关，是一个全市场的属性；我们用当下股价与不同的历史股价的的变化率来构造当下的因子，用以描述历史股价曲线（如下图）。



基本定义

假设第i家公司t时刻的股价为$m_{i,t}$，因子$x_{i,t}(\tau)$表示最近τ时间段的变化率：

$$x_{i,t}(\tau) = \frac{m_{i,t} - m_{i,t-\tau}}{m_{i,t-\tau}} \tag{1}$$

我们选取不同的τ，从一天，两天，到2年，由此构造t时刻的因子：

$$x_{i,t} = (x_{i,t}(1), x_{i,t}(2), \cdots, x_{i,t}(250), x_{i,t}(500)) \tag{2}$$

因子$x_{i,t}$需要预测的变量为：

$$\hat{y}_{i,t}(s)=\frac{\hat{m}_{i,t+s}-m_{i,t}}{m_{i,t}} \tag{3}$$

我们可以先针对单一期限s进行建模，例如令s等于一周，即5个交易日。这里我们使用$\hat{m}_{i,t+s}$表示$t+s$时刻的股价是未知的。

我们希望构建$\hat{y}_{i,t}(s)$和$x_{i,t}$之间的关系：

$$\hat{y}_{i,t}(s)=f\left(x_{i,t};\beta(s)\right)+\epsilon_{i,t}(s) \tag{4}$$

其中，$f(x)$是关于x的非随机函数，$\beta(s)$是决定函数f形状的参数向量：

$$\beta(s)=\left(\beta_1(s),\beta_2(s),\cdots,\beta_{250}(s),\beta_{500}(s)\right) \tag{5}$$

$\beta_\tau(s)$对应于$x_{i,t}(\tau)$。

$\epsilon_{i,t}(s)$是误差项，我们假设其满足正态分布：$\epsilon_{i,t}(s)\sim N(0,\sigma_i(s))$。如果我们假设不同股票的误差项的标准差没有区别，则$\sigma_i(s)=\sigma(s)$。$\hat{y}_{i,t}(s)$的期望值为：

$$E[\hat{y}_{i,t}(s)]=f\left(x_{i,t};\beta(s)\right) \tag{6}$$

方差为：

$$V[\hat{y}_{i,t}(s)]=E\left[\left(\hat{y}_{i,t}(s)-E[\hat{y}_{i,t}(s)]\right)^2\right]=E[\epsilon_{i,t}^2(s)]=\sigma_i^2(s) \tag{7}$$

上式可以用来估计误差项的标准差$\sigma_i(s)$。

如果$f(x)$是线性函数，则我们有如下：

$$\begin{aligned}\hat{y}_{i,t}(s)&=\beta_1(s)x_{i,t}(1)+\beta_12(s)x_{i,t}(2)+\cdots+\beta_{500}(s)x_{i,t}(500)+\epsilon_{i,t}(s) && (8)\\ &=\sum_{j\in\mathcal{T}}\beta_j(s)x_{i,t}(j)+\epsilon_{i,t}(s) && (9)\end{aligned}$$

其中，$\mathcal{T}=(1,2,\cdots,250,500)$，表示因子所使用的期限的集合。

样本准备

因子$x_{i,t}$和因变量$\hat{y}_{i,t}(s)$的准备比较直接，均为不同期限的变化率。$\hat{y}_{i,t}(s)$的样本$y_{i,t}(s)$际上为未来s时刻的因子：

$$y_{i,t}(s)=\frac{m_{i,t+s}-m_{i,t}}{m_{i,t}}=x_{i,t+s}(s) \tag{10}$$

由于因子$x_{i,t}(\tau)$涉及不同的期限，因此其长度不一，我们只选取所有因子都能计算的部分，即至少上式2年后的时间开始算起。

参数估计

因子$\beta(s)$的估计取决于模型$f(x)$的选择。在模型确定后，我们需要估计误差项$\epsilon_{i,t}(s)$的标准差$\sigma_i(s)$。假设不同公司有相同的标准差，则利用等式(7)，我们有如下估计方式：

$$\sigma^2(s)=\frac{1}{n}\sum_{i,t}\left|y_{i,t}(s)-f(x_{i,t};\beta(s))\right|^2 \tag{11}$$

每一个股票的$\sigma_i(s)$则为：

$$\sigma_i^2(s) = \frac{1}{n}\sum_t \left|y_{i,t}(s) - f(x_{i,t};\beta(s))\right|^2 \tag{12}$$

预测

当我们有参数$\beta(s)$和$\sigma_i(s)$后，我们可以估计未来到s时刻的股价增长率的分布：

$$\hat{r}_{i,t}(s) = f(x_{i,t};\beta(s)) + \sigma_i(s)\hat{Z} \tag{13}$$

其中，$\hat{Z}$满足标准正态分布。因此，收益率$\hat{r}_{i,t}(s)$满足正态分布：$N\left(f(x_{i,t};\beta(s)), \sigma_i(s)\right)$。某个百分位$q$的收益率为：$f(x_{i,t};\beta(s)) + \sigma_i(s)\Phi^{-1}(q)$，$\Phi(x)$为标准正态分布的累积分布函数。