# 金融机器学习算法

第五讲

金融机器学习的分数差分处理

# 本讲主要内容

- 进行分数差分处理的动机
- F-差分原理
- 固定窗口 F-差分
- 平稳性与记忆性平衡

# 进行分数差分的动机

- 金融序列数据通常信噪比 (signal-to-noise ratio) 很低
- 金融预测同时依赖于信号的时序记忆性 (memory) 和平稳性 (stationary)
- 平稳性对于预测的意义：监督学习算法要求 $y$ 标签的变量是平稳的，否则无法将未知新观测对应到历史已知观测
- 整数差分造成了过度差分，为保证平稳性，但进一步降低了时序记忆性，信噪比更低
- 缺乏时序记忆性时，使用再复杂的统计技术都无法完成预测，只能贡献错误发现
- 分数差分能够同时兼顾记忆性和平稳性

# F-差分原理

## 整数差分的原理

F-差分 (Fractional Differentiation) 是整数差分的自然推广。

- 滞后算子 $\mathcal{L}$ 作用与序列 $X_t$，$\mathcal{L}X_t = X_{t-1}$，$\mathcal{L}^k X_t = X_{t-k}$
- $(1-\mathcal{L})^2 X_t = X_t - 2X_{t-1} + X_{t-2}$
- $(1-\mathcal{L})^{-1} = 1 + \mathcal{L} + \mathcal{L}^2 + \mathcal{L}^3 + \cdots$
- $X_t = X_{t-1} + u_t$，$X_t - \mathcal{L}X_t = u_t$，$X_t = (1-\mathcal{L})^{-1} u_t$
- 【思考】$X_t = \rho X_{t-1} + u_t$, $|\rho| < 1$ 如何表示为 $u_t$ 滞后项的和?

# F-差分原理

## 整数差分的原理

- $d$ 为任意整数，$(1+\mathcal{L})^d = \sum_{k=0}^{d} C_d^k \mathcal{L}^k 1^{d-k} = \sum_{k=0}^{\infty} C_d^k \mathcal{L}^k 1^{d-k} = \sum_{k=0}^{\infty} C_d^k \mathcal{L}^k$
- 将差分算子式展开

$$
\begin{aligned}
(1-\mathcal{L})^d = \sum_k^{\infty} C_d^k (-\mathcal{L})^k &= \sum_k^{\infty} \frac{\prod_{i=0}^{k-1}(d-i)}{k!}(-\mathcal{L})^k \\
&= \sum_k^{\infty} (-\mathcal{L})^k \prod_{i=0}^{k-1} \frac{d-i}{k-i} \\
&= 1 - d\mathcal{L} + \frac{d(d-1)}{2!}\mathcal{L}^2 - \frac{d(d-1)(d-2)}{3!}\mathcal{L}^3 + \cdots
\end{aligned}
$$

# F-差分原理

- 将整数 $d$ 拓展为任意实数，差分算子同样表示为

$$(1-\mathcal{L})^d = 1 - d\mathcal{L} + \frac{d(d-1)}{2!}\mathcal{L}^2 - \frac{d(d-1)(d-2)}{3!}\mathcal{L}^3 + \cdots$$

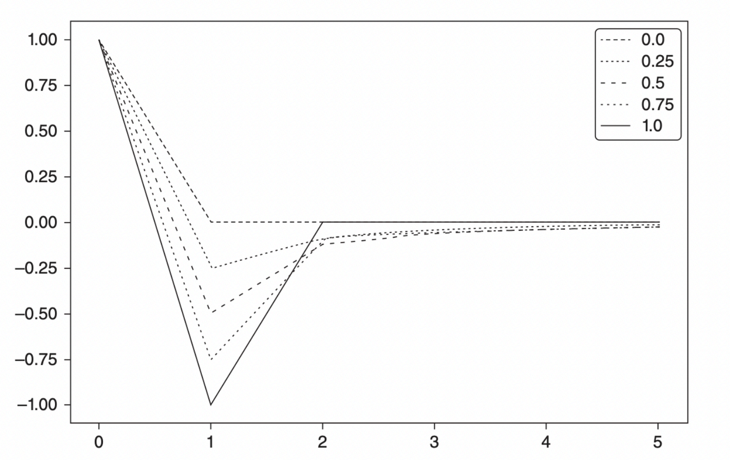
- 因此，选择 $d$ 为非整数，F-差分将保留序列的记忆性

$$\widetilde{X}_t = \sum_{k=0}^{\infty} \omega_k X_{t-k}$$

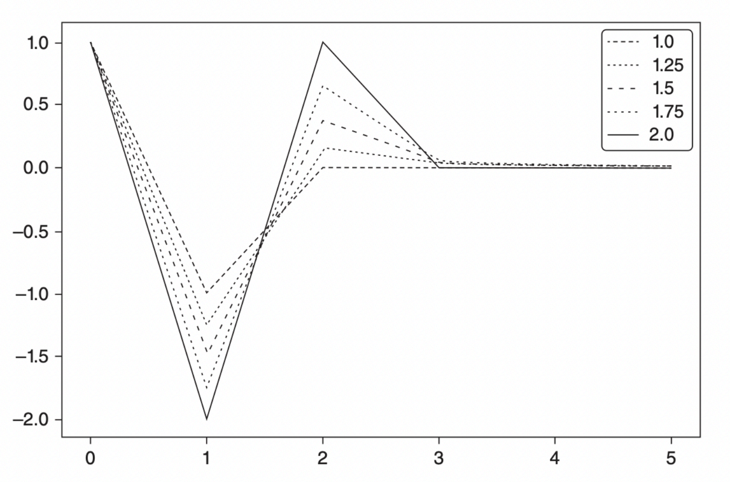
其中，$\omega = \left\{1, -d, \frac{d(d-1)}{2!}\mathcal{L}^2, -\frac{d(d-1)(d-2)}{3!}\mathcal{L}^3, \cdots, , (-1)^k \prod_{i=0}^{k-1} \frac{d-i}{k!} \cdots \right\}$

- $\omega_k = -\omega_{k-1} \frac{d-k+1}{k}$

# F-差分原理: 示例 $d \in [0, 1]$



# F-差分原理: 示例 $d \in [1, 2]$



# 固定窗口 F-差分

- 固定窗口 F-差分 (Fixed Window Fractional Difference, FFD)：序列 $X_t$ 长度有限，$\omega_k$ 值有限，选取一个固定的回溯窗口 $[t-l^*, t]$ 中的数据来计算 F-差分
- 相当于选取了一个阈值 $\tau$，其中 $|\omega_{l^*}| \geq \tau$，但 $|\omega_{l^*+1}| \leq \tau$，
- 计算方法

$$\widetilde{X}_t = \sum_{k=0}^{l^*} \tilde{\omega}_k X_{t-k}, \qquad t = T - l^* + 1, \cdots, T$$

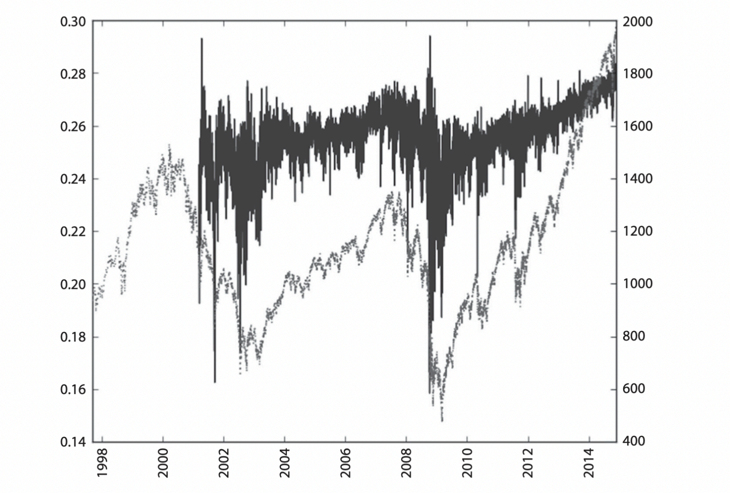
其中

$$\tilde{\omega}_k = \begin{cases} \omega_k, & k \leq l^* \\ 0, & k > l^* \end{cases}$$

# 固定窗口 F-差分的结果

- 获得平稳的时间序列 $\widetilde{X}_t$
- 分布可能不再是正态分布
- 分布可能具有一定的偏度
- 分布可能具有一定的峰度
- 选取合适的 $d$ 值可以得到平稳序列

# 固定窗口 F-差分实例：E-mini 标普 500 期货价格



# 平稳性与记忆性平衡

- 考虑一个原始序列 $\{X_t\}_{t=1,\ldots,T}$，在运用 FFD 时，可优化选取一个最小的 $d^*$ 使得 F-差分得到的序列 $\{X_t\}_{t=1,\ldots,T}$ 通过 ADF 检验
- ADF 检验：$H_0$ 原始序列存在单位根非平稳，$H_1$ 原始序列平稳
- 可选取显著性水平 $\alpha = 0.05$ （$95\%$ 的置信度）来进行 ADF 检验

# 平稳性与记忆性平衡

例子：E-mini 标普 500 期货价格的记忆性保留