利用信息协变量和伪发现率构建基金组合

韩雪菲

指导老师：邓璎函

2022 年 5 月 12 日

目录

Introduction

- Hsu et al. (2020) 认为
  有一部分基金出现正 alpha 是出于运气
  使用传统的的 FDR^+ 框架可以识别真正有能力的熟练型基金（skilled fund）-即由于优秀的管理能力及交易策略而具有真正正向 alpha 的基金。
  将 FDR^+ 与信息协变量结合提出 $fFDR^+$ 并设定一个目标 $fFDR^+$，基金组合策略可以获得相当不错的表现

信息协变量的计算

基金池为同花顺开放式基金中的 647 个股票型基金，信息协变量数据为 2017.03-2020.03 的月度收益。

- R-square:Carhart 四因子模型的 R-square
- Fund Size:

$$FundSize_{i,t} = ln\frac{TNA_{i,t}}{IndustrySize_t} - ln\frac{TNA_{i,0}}{IndustrySize_0}$$

- Return Gap：基金的回报差距等于基金本期实际披露的回报与基金根据其上一次披露信息的理论回报之间的差异。

信息协变量的计算

- Active Weight：活跃权重定义为股票价值权重与基金分配给其投资组合中股票的实际权重的绝对差值之和（本文未讨论）。
- Fund Flow：

$$FundFlow_t = \frac{TNA_t - (1 + r_t)\,TNA_{t-1}}{(1 + r_t)\,TNA_{t-1}}$$

- Idiosyncratic volatility (Sigma)：Carhart 四因子模型的 alpha 的波动率 (Sigma)
- Sharpe ratio, Beta，Treynor ratio

根据 Carhart 模型的 alpha 分布

我们使用 Carhart (1997) 的四因子模型来计算基金月度业绩：

$$r_{i,t} = \alpha_i + b_i r_{m,t} + s_i r_{smb,t} + h_i r_{hml,t} + m_i r_{mom,t} + \varepsilon_{i,t}, i = 1, ..., m,$$

根据估计的 alpha 我们定义:

- 非熟练基金/零 alpha 基金（unskilled fund 或者 zero alpha fund）：基金经理的选股技能不足产生大于 0 的 alpha。
- 熟练基金（skilled fund）：拥有足够选股能力的基金，可以提供超额 alpha（alpha $>$ 0）。

FDR^+

FDR 就是统计上进行假设检验时候错误地拒绝假设的个数占所有被拒绝的原假设个数的比例
在选定的基金组合（alpha 为正而且 p 值显著的基金组合）中的 false discovery rate（伪发现率）即错误的选择了 alpha 为 0 的基金的概率，

$$FDR^+ = \mathbb{E}(\frac{V^+}{max\{R^+, 1\}})$$

R^+ 是实证中熟练基金个数，V^+ 为错误地选择了零阿尔法或不熟练的基金个数

传统 FDR^+ 策略

当我们得到每个基金的 alpha，我们对每个基金进行假设性检验:

$$H_0 : \alpha = 0, \qquad H_1 : \alpha \neq 0$$

由此得到 p 值，我们认为 p 值是均匀分布的，那么对于某个基金，零 alpha 基金的占比：$\pi_0(\lambda)$ 可以由下面公式获得：

$$\pi_0(\lambda) = \frac{\#(p_i|p_i > \lambda, i = 1, ...m)}{(1-\lambda)m}$$

这里的 λ 等于 λ^* 最优，此时 λ^* 是一个足够大的 p 值的阈值由优化求解使 $\pi_0(\lambda)$ 的均方误差 (MSE) 最小获得.

传统 FDR^+ 策略

在给定的一个显著性的阈值 γ，我们认为当一个基金 p 值 $\leq \gamma$ 时，他有大于 0 的估计 alpha ，也就是认为他是熟练基金，所以此基金组合的 FDR^+ 为

$$FDR^+ = \frac{\pi_0 \gamma / 2}{R^+ / m}$$

R^+ 是 alpha 大于 0 且 p 值 $\leq \gamma$ 的基金个数, 在复现过程中，在 10% FDR^+ 目标下，我们选中 444 个基金的组合。

利用信息协变量将 FDR^{+} 升级 $fFDR^{+}$

将信息协变量纳入考量后，在已知给定的信息协变量，单个基金的假设性检验如下：

$$H_0 : \alpha = 0, \qquad H_1 : \alpha \neq 0$$

我们定义 h 为原假设的状态，即如果假设 $\alpha = 0$ 为真，则 h = 0，否则为 h = 1。此外，P 是检验 p 值的随机变量表示，Z 代表某个信息协变量，所以单个基金的信息集为 T = (P, Z)，我们假设 (h|Z = z) 服从 Bernoulli$(1 - \pi_0(\lambda))$，以 Z = z 为条件，该基金具有零 alpha 的概率为 $\pi_0(\lambda)$.
以之前同样的方式求解 λ 的最优值（我们没有对单个基金求 λ 而是在多重检验时，将基金依据相同的信息协变量分为小组，对每组以最小化平均积分平方误差（MISE）求最优化的 λ.

利用信息协变量将 FDR^+ 升级 $fFDR^+$

接着我们对实证中已选中的熟练基金组合（alpha 大于 0 且 p 值 $\leq \gamma$ 的 m 个基金）进行评估，进行 multiple hypothesis testing 多重假设检验

$$H_{0,i}: \alpha_i = 0, \qquad H_{1,i}: \alpha_i \neq 0, \qquad i = 1, \ldots m$$

我们假设所有基金都是独立的，并且每个基金的信息集 T =(P,Z) 都具有相同的分布。最后，我们用 f(p,z) 表示 (P,Z) 的联合密度函数。我们可得：在观察到某个 T = (p, z) 时 null hypothesis 零假设为真的后验概率为

$$\mathbb{P}(h = 0 | T = (p, z)) = \frac{\pi_0(z)}{f(p, z)} \doteq r(p, z)$$

利用信息协变量将 FDR^+ 升级 $fFDR^+$

$$\mathbb{P}(h=0|T=(p,z))=\frac{\pi_0(z)}{f(p,z)}\doteq r(p,z)$$

也就是一个基金被因为运气被选入在组合中的概率。
这里的 $\pi_0(z)$ 是将上述每组的 $\pi_0(\lambda)$ 使用平滑样条法 smoothing spline method 所得；
f(p,z) 是 probit 转换后的 local likelihood kernel density estimation (KDE) 方法求得
在正 alpha 基金的基金池中我们得到 $fFDR^+$：

$$fFDR^+(\Gamma)=\mathbb{P}(h=0|T\in\Gamma)=\int_\Gamma r(p,z)dpdz$$

Γ 是所有的信息集组合。

利用信息协变量将 FDR^+ 升级 $fFDR^+$

我们定义在给定的目标 τ 下的 q 值为：

$$q(p,z)=\inf_{\{\Gamma_\tau|(p,z)\in\Gamma_\tau\}} fFDR^+(\Gamma_\tau)$$

实践中在给定的 (p_i, z_i) 下的 q 值计算公式为：

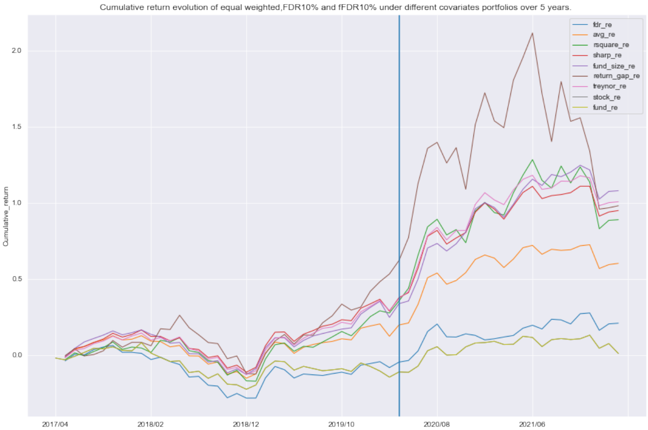
$$q(p_i,z_i)=\frac{1}{S_i}\sum_{k\in S_i} r(p,z) \qquad S_i=\{j|r(p_j,z_j)\leq r(p_i,z_i)\}$$

p_i 是第 i 次假设检验的 p-value，$z_i=r_i/m$，这样我们将信息协变量 Z_i 的观测值转换为满足假设 Z_i 服从 Uniform(0, 1) 的形式.r_i 是观测值 Z_i 在的样本中的排名。
这样在给定目标 $\tau\in[0,1]$，当且仅当其 q 值 $\leq\tau$ 我们才会拒绝零假设，这样可以保证才可以保证 FDR 控制在 τ。

构建投资组合

通过 $fFDR^+$ 来控制因为运气选入的非熟练资金的比例，目标 $fFDR^+=10\%$
利用 2017-2020 年数据得到基金的信息协变量的观测值，估计的 alpha 和假设检验的 p 值。按照 $fFDR^+$ 目标 10% 和滚动的 3 年样本构建投资组合。从 2021 年开始，在 t 年末，根据过去 3 年（t-2 到 t）的历史信息，选择一组基金在 t+1 年进行投资。

利用不同信息协变量策略的回报在样本内和样本外的时间序列



利用不同信息协变量策略的回报的表现

	FDR+	avg	Rsquare	Sharp	Fund_size	Return_gap	Treynor
Annual return	5,2%	10,8%	15,0%	14,9%	16,3%	17,6%	15,5%
Sigma	4,8%	4,5%	5,8%	4,6%	4,7%	7,9%	4,6%
Sharp	8,9%	19,7%	21,3%	26,5%	28,8%	18,5%	27,8%
alpha_Carhart	0,1%	0,3%	0,7%	0,6%	0,7%	0,8%	0,7%
alpha_FF	-0,1%	0,8%	1,1%	1,2%	1,7%	2,0%	1,5%

五个信息协变量下 $fFDR^+$ 策略均有正向 alpha，而且可以产生 15% 及以上的年化收益。目前来看，相比于等权重策略和传统 FDR^+，加入信息协变量的 $fFDR^+$ 表现更为优秀。

总结与展望

- 本文利用开放型基金市场股票型基金，复现了 Hsu et al. (2020) 多假设检验框架的 $fFDR^+$ 策略，来选择表现出色的基金组合。在估计伪发现率时创新性的结合了信息协变量。在不同的信息协变量和给定的 $fFDR^+$ 目标下，构建的投资组合可以有较好收益。
- 风险提示:
  本文的研究是基于对历史数据的统计和分析，策略的历史收益率不代表未来收益率。若市场环境发生变化，策略的最终表现可能发生改变。

参考文献

Hsu, Po-Hsuan and Kyriakou, Ioannis and Ma, Tren and Sermpinis, Georgios, Informative Covariates, False Discoveries and Mutual Fund Performance (November 25, 2020).

Thank you!