基于多因子模型，指数化和 Smart Beta 模型的因子类算法因子开发计划

从 Fama French 92 之后，各类因子类算法层出不穷，业界比较成功的应用主要有：基于单因子回归，多元回归 cross section analysis 的线性模型；基于因子的指数化指数；基于因子显著后 Beta 类的算法。随着市场的有效性不断提高，alpha 将是投入产出比越来越低的算法。我的预期是在尽量严谨的计算主流因子的情况下，在多计算更多的类似“因子评分加权”的方法，找到长期有效显著的因子组合预计扩充因子库至千个级别。

因为我不能知道明天市场发生什么，由于个人水平所限，甚至也不能一直通过个人努力保持算法的先进性甚至有效性。我最多可以做到的是避免犯下愚蠢的错误。此外，市场上总有一类资产亦或者是因子能在超额回报上胜于其他因子。那么这些因子的转换，或许可以称之为“轮动”。

1. 主流因子

下图为有代表性的因子类别，并不代表所有的因子。

Factor	Description
	Low Voaltility
Low Voaltility	*Seeks to measure the performance of the fixed number of leaset volatile stocks in Benchmark*
Reduced Volatility	*Seeks to measure the performance of the Benchmark, with an overweight of stocks with relatively low volatility and an underweight of stocks with relatively high volatility*
Minimum Volatility	*Seeks to measure low volatility stocks with in stocks of Benchmark using an optimizer and stock and sector constrains*
Low Beta	*Seeks to measure the 70% of stocks with the loweset beta with benchmark index for a given region/sector with a 5% to minimize turnover*
	High Beta
High Beta	*Seeks to measure the performance of the benchmark index that are most sensitive to changes in the market returns*
High Beta Bullish	*The same one, but the benchmark is designed for investoring initiation a bullish strategy or making a directional bet on current markets.*
High Beta Bearish	*The same one, but the benchmark is designed for investoring initiation a berish strategy or making a directional bet on current markets.*
High Beta Speicified	*The same one, but the benchmark is sector/listed country specified.*
	Growth and Value
Growth and Value	*Seeks to measure stocks in the benchmark index using three factors: sales growth, the ratio of earning change to price and momentum.* Seeks to measure stocks in the benchmark index using three factors: the ratios of book value, earnings, and sales to price.
	Value
Intrinsic Value	*Seeks to measure the performance of the respective markets, as weighted by intrinsic stock value rather than by traditional market capitalization*
Enhanced Value	*Seeks to measure the performance of the top quintile of stocks in the respective region based on their value scores, calculated using three fundamental measures: book value-to-price, earnings-to-price, and sales-to-price by different sector and listed countries*
Low Valuation	*Seeks to measure stocks from the benchmark index, overweighting stocks with relatively cheap valuations and underweighting stocks with relatively expensive valuations*
	Dividends
Aristocrats	*Seeks to measure the performance of the highest-dividend-yielding constituents of the benchmark index that have followed a policy of increasing or stable dividends for at a considerable number consecutive years.*
Select	*The same one, but the benchmark is sector/listed country specified.*
High Yields	*Seeks to measure the performance of a fixed number of highest-yielding constituents of the benchmark index that meet size, liquidity, and profitability criteria*
	Momentum
Momentum	*Seeks to measure the performance of top quintile of securities in the respective markets that exhibit persistence in their relative performance.*
	Quality
Quality	*Seeks to measure the top quintile of high-quality, large- and mid-cap stocks in the respective markets, as determined by their quality score. This score is calculated based on return on equity, accruals ratio and financial leverage ratio.*
	Capital Expenditures
Capital Expenditures	*Seeks to measure a fixed number constituents of the benchmark index that have exhibited the strongest capital discipline, in the form of efficient capital expenditures, over the near term.*
	Multi Factor
Quality, Value & Momentum	*Seek to measure the performance of 70 stocks within the S&P Europe 350 that are characterized as having the top combination of quality, value, and momentum as determined by a multifactor score.*
GIVI	*Seeks to measure the 70% of stocks with the lowest volatility from the volatility dashboard, as measured by regional market stock beta.*
Low Volatility High Yield	*Seeks to measure the performance of the fixed number least-volatile, high-dividend-yielding stocks in the benchmark*
Quality, Value & Growth	*Seek to measure the performance of stocks in the benchmark index that have the highest combination of quality, value, and growth.*
Factor Rotator	*Seek to rotate its investment strategy across four distinct strategies based on the most recent economic data from the Economic Index, with a target volatility of fixed level (e.g. 6%).*

大多数因子类算法的市场参与者想到的因子都不大相同，能持续找到有效显著的因子是这个行业大多数人做的事情。而且理财魔方作为专业的投资机构，我个人的建议不应该是只考虑到这个算法的年化收益，而且也应该考虑到投资者的风险偏好，流动性要求，计划投资时长等等问题。那么如此一来，这个工程中绝不仅仅需要考虑风险，（最大）回撤，因子“轮动”周期时长等等。

1.1. 因子分类

大体金融时间序列的因子可以通过以下三种方式产生：

- 数值分析：基于价/量的数值分析。
- 三大报表：收益/偿债能力/流动性
- 其他数据

其中有很多优秀的公司已经花了很大精力在第三类别中，想要找到“不同的”有效的因子构造方法，比如天气，温度，地理位置等等。非结构化数据结构化也有一些人在尝试。由于本人水平有限，本次工程计划主要基于前两点。

1.2. 因子暴露

因子在暴露后的风险应衡量为是：

$$\sigma = \text{Exposure} \times \text{Volatility} \times \text{Correlation}$$

比如典型的 Barra 模型，就是上述公式的优秀实践者。在单元回归中，很多问题不需要考虑。但是在 cross validation 中以上的每一个矩阵在相乘之前都需要做诸多处理，比如 correlation 矩阵的自相关问题等等。

2. 算法处理：

我本人不太喜欢也并没有资格说“细节在魔鬼”之类的鸡汤，但是如果算法本身处理以及打磨的不够精细，那么未免最后成为垃圾，至少也落在同行业之后。

2.1. 残差

残差的处理是大多数计量经济学的从业者做的事情之一。然而，鄙人看到有很大一部分有一些报告喜欢用“显然”，然后放到 OLS 的架构中去计算，而完全忽略了本身 OLS 的前提：

- Residual ~N(0,1) IID
- Heteroscedasticity
- Autocorrelation

导致结果一定是垃圾，当然随着行业水平的越来越高，很多优秀的团队开始 GMM，Cointegration 等处理的非常好。

2.2. 协方差矩阵

Newey West 的处理非常好的解决了在多回归的时候因子之间协方差矩阵系数的时间序列自相关性。期待我们可以找到更优的算法。

2.3. 因子矩阵

比如矩阵不满秩，不可逆，不正定等，需要不同的切割方法。

2.4. 标的

本着不局限于投资于公募基金作为 MOM 的原则，希望这一套框架可以在完成后以非常低的成本转移到股票标的，那么标的的选择成为了第一步也同样是极其重要的一步。不只是乱停牌，交易日短等情况，那么如何 Fill NA，如何 Rolling 都是需要进一步需要处理的。比如，如果 Fill NA 用 Mean 这个值，那么势必会让因子的标准差下降，那么认为提高了因子的稳定性，如此一来，那

么停牌日多的标的相应的因子暴露结果甚至比一直交易的标的因子暴露结果更加稳定，这个结论显然是荒谬的。

2.5. 其他处理

Volatility Drag； FDR； Tracking Error v.s. Volatility 等等

3. 优化目标

大名鼎鼎的 Black Litterman 褫其华衮就是优化 Sharp Ratio 的模型，在不同的优化目标下，算法表现以及预期风险偏好肯定是不同的。比如最小化风险，最大夏普比，白噪声化因子相关性最低等等。

4. 其他

至于因子“长期有效”和“因子轮动”的 Trade-off 中，一个现实肯能遇到的问题主要就是：假设检验的严格程度适当的情况下，长期有效的因子调整过的组合投资表现在某种程度上一定会损失短期收益以及减少未来某些时间区间内的收益。那么如果仅仅考虑到因子的变化，比如行业以及时常风格等。势必在因子失效半衰期这个近似随机的数值上有所损失。如何管理这些因子以及短期长期的选择将成为此工程完成之后使用的关键。

能力一般水平有限，旨在抛砖引玉。

周渤洋

2019 年 1 月 2 日