本文分别从行业和上市公司两个维度，简要介绍相关基本面量化投资的逻辑要点、因子选择、数据处理、模型搭建等内容。本篇将首先介绍行业基本面量化体系。

### ****一、行业基本面量化体系****

****1、基本面量化在成熟股市崛起的时代背景****

金融市场的量化投资，自20世纪80年代开始飞速发展至今，已经有数十年时间。从欧美的投资实践来看，专业从业人员在早中期围绕数理金融的纯[量化交易](https://so.csdn.net/so/search?q=%E9%87%8F%E5%8C%96%E4%BA%A4%E6%98%93&spm=1001.2101.3001.7020" \t "/Users/alan-hopiy/Documents\\x/_blank)中逐渐发现，数理金融的抽象理论逐渐弱化，而金融市场本身的逻辑则愈加重要。正如投资大师格雷厄姆所言，“市场短期是投票器，长期是称重机”。在股票市场中，纯市场交易的数量模型，其单一性和有效性不足的缺点越发明显。

正是因为这些考虑，对冲基金们把量化和基本面分析有效结合起来，创造出一个跨界的新流派：Quantamental——量化基本面分析法。该量化投资流派既包括了基本面投资者自下而上的个股筛选技巧，同时也包括了借助计算机技术和大数据系统对有效量化投资模型的建立。整体而言，基本面量化作为典型的一种主动数量化投资，在股票投资上的优点非常明显，近两年来也越来越受到国内公私募基金管理人的认可。而在此之中，行业基本面量化与上市公司基本面量化，则是股票投资基本面量化主要的两个部分。

****2、行业基本面量化的定义和优点****

所谓行业基本面量化，顾名思义，主要就是指在股票投资中，以基本面量化的手段，选取未来能够显著跑赢市场的行业。考虑到大多数情况下，股票市场对处于高度景气周期中的行业资产已经进行了相对合理的“估价”，行业基本面量化主要的目标，就在于寻找到高效的先行因子，来预判未来哪些行业在利润等主要基本面指标上，能够显著领先于其他多数行业。

行业资产配置的一个基本逻辑是，各行业在相同的经济环境却可能遵循着不同的经济模式。一方面，经济内在发展本身所具有的行业特点差异所至，有的行业表现好的时候意味着其它行业表现差，比如通常意义上所观察到的上下游行业之间所表现的那样；另一方面则是股票市场实际运行在大量情况下证实，不同行业的股票资产在市场定价中，往往会表现出鲜明的行业轮动特点。

虽然许多面向全球金融市场基金更重视地区之间资产配置的选择，但近年来海外许多实证学术研究已经证实，虽(country allocation)，大量不同国家、不同地区配置的收益，其实更多来源于行业选择，即行业差异对基金投资业绩表现的贡献度，要大于地区差异所带来的投资业绩差异。

****3、行业基本面的备选量化因子库****

在行业基本面的量化因子选择方面，近年来，国内多家券商、公墓基金等主流投资机构均已结合中国A股市场的实际，做了许多实证性研究。

相关研究主要的出发点，当然不是瞄准当前行业的基本面情况，而是为了对未来一段时间行业整体景气度和利润状况进行预判，确定备选的基本面量化因子库。

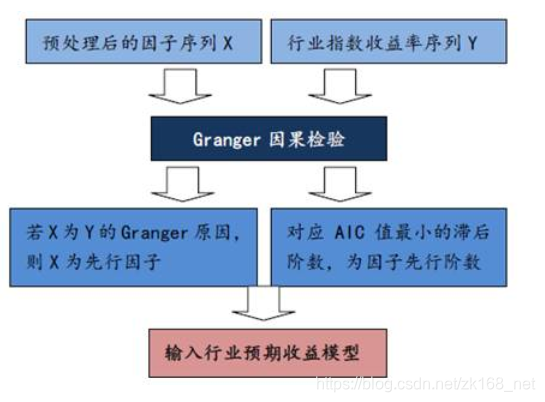
一般而言，在海外市场，许多基本量化投资人也可能主要关注“行业因子”这一类因子对行业的先行性参考。不过考虑到中国经济周期的特殊性，以及近年来愈发明显地对货币环境的依赖性，行业基本面的先行因子，大致可以划分为三大类：宏观因子、货币因子和行业因子（见下图）。



这里需要指出的是，由于上述三大类因子数量庞杂，专业投资人应当根据相关指标在自身因子库的重要性，以及对不同行业景气度的参考性高低等，进行初步选择，避免后续搭建有效因子库和模型时，产生待处理数据过多、模型过度复杂的状况。

****4、有效因子的回测、筛选和模型搭建、结论****

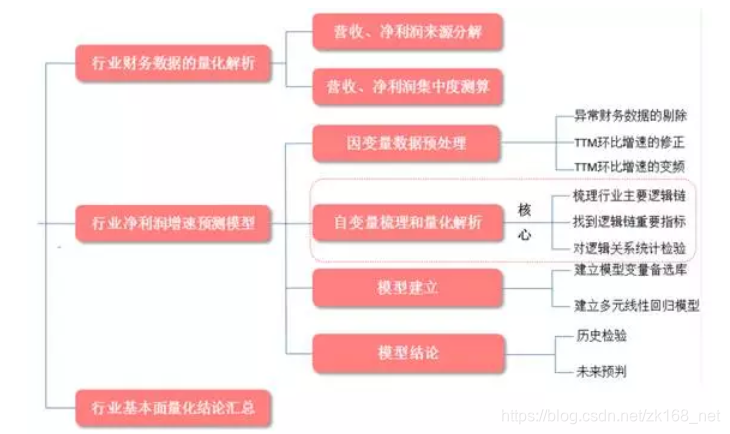
整体来看，在备选因子库确定之后，行业基本面量化有效因子的筛选，应当尽量采取最能够确定备选因子影响力高低的量化方法。一般而言，有两种常用的方法会被专业投资人普遍采用。一是Granger 因果检验，二是对备选因子赋值法。



所谓Granger 因果检验，是由2003 年诺贝尔经济学奖得主克莱夫·格兰杰（Clive W. J.Granger）所开创，用于分析经济变量之间的因果关系。简要来说，Granger 因果检验的原理大致如下图所示，读者在实际操作中对Granger 因果检验的原理、计算公式做进一步详细了解。

所谓因子赋值法，就是根据经验和常识直接对备选因子进行赋值、调整和检验。该方法与Granger 因果检验法大同小异，优点在于可以更快速地找到显著性因子，缺点则是可能会忽略掉某些特殊而在一般投资人经验范围之外的因子。

在进一步的模型搭建上，以行业因子为例，有效因子筛选和模型搭建的流程，可以用下图予以简要概括。



****5、数据处理中的常见问题****

在因子回测和筛选过程的数据处理中，有两大类问题需要予以重点关注。一是对某些数据大量缺失的因子进行排除；二是对某一因子相关异常数据的剔除。

一般而言，某一因子出现过多数据的缺失，势必会令该因子的可信度产生很大疑问。因此，如果数据缺失量如果比较大的话，应该毫不犹豫将该变量剔除。当然，如果数据缺失只是发生在极个别时间点的话，在具体操作中还是可以进行适当插值填补的。整体而言，在行业基本面的回测和预测中，质量过差的数据，都会影响有效因子的筛选和最终模型的搭建。

另一方面，在对具体原始数据的处理中，还应当对一些异常数据进行适当剔除。例如，有些行业市值比较小，这时候很有可能会出现单个上市公司的净利润对整个行业影响较大，这时候如果这个上市公司的业绩异常，并不能很好地用行业的基本面情况去解释。那么就需要把这些异常的上市公司剔除，减少这些公司业绩剧烈波动给行业基本面变化研判带来的失真影响。此外，像因行业新的上市公司上市给行业财务数据带来的扰动、某些特殊行业在单一财年内出现的季节性业绩波动，也都需要在具体的数据处理中进行适当修正。

****6、行业基本面量化体系的案例****

整体而言，尽管过去多年以来A股市场就长期曾经长期表现出很强的高波动性和高投机性，但实际上，从行业基本面量化的角度来看，长期而言市场依然在“价值的兑现”上，有着很好的验证。

仅从某券商机构回测选定的“行业净利润TTM 环比增速”进行的行业配置策略来看，一个单一因子的有效行业基本面量化体系，在2010年年初至2015年三季度，就已经显示出了很好的投资收益率优势。（见下图）



### ****二、上市公司基本面量化体系****

****1、上市公司基本面量化的概念和投资逻辑****

说起依据上市公司基本面的股票投资，就不得不介绍“价值投资”的概念。所谓价值投资，简单来说，就是指通过上市公司基本面的分析，购买价值被低估的股票。价值投资鼻祖格雷厄姆认为，价值投资的本质在于以低于其内在价值（intrinsic value）的价格购买股票，并由此首先提出了安全边际的概念。而投资大师巴菲特则又把价值投资理论向前推进了关键的一大步。他把重点放在了寻找定价合理的杰出公司上，而不是或者不仅仅是以低廉的价格购买普通公司。

而具体到上市公司基本面的量化投资而言，基于上市公司基本面的量化投资，主要的目标就是用量化的方式筛选当前基本面表现优秀、或者基本面正在显著好转的上市公司标的，以达到量化选取的投资组合在收益率上能够显著跑赢市场的目的。这其中的基本面量化体系，主要也正是将价值投资的标准予以数量化表达、并进行筛选。因此，大体可以认为，上市公司基本面量化策略，与价值投资策略，大致上是异曲同工。

从理论上来说，如果价值投资未来在中国股市上愈加有效，那么一套逻辑缜密、因子高效、切合上市公司实际内在价值的上市公司基本面量化模型，也将会在股市的长期运行中显著战胜市场。

****2、上市公司基本面量化的备选因子库****

一般而言，价值投资或者说基本面投资，主要是以股票的内在价值为依据而进行投资。而影响上市公司内在价值的因素，大体划分为宏观、中观、微观等三大方面。如宏观层面的经济状况、宏观层面的金融环境、中观层面的行业前景、行业周期等等。而这些宏观和中观方面，一般也都可以在行业基本面量化体系中予以定量。因此，上市公司基本面量化，则主要聚焦于上市公司微观层面经营业绩的情况，以及相应的股价表现。

根据价值投资的经典研究体系和基本面量化多年以来的投资实践来看，上市公司基本面量化的备选因子库相当庞杂，一般常见的至少有以下数十种。如净资产同比增长率、总资产同比增长率、营业收入同比增长率、扣非净利润同比增长率、每股收益EPS、每股收益EPS增长率、净资产收益率、净资产收益率增长率、流动比率、速动比率、资产负债率、应收账款周转率、存货周转率、毛利率、净资产收益率、总资产净利率、总股本、流通股本、现金营运指数、销售现金比率、市盈率、市净率、市销率、行业市盈率、行业市净率、行业市销率、行业净资产收益率、行业净资产收益率同比增长率、股息率、股息率增长率、总股本、总市值、流通股本、流通市值等等。

整体来说，在与上市公司基本面有关的大量备选因子中，可以进一步将其划分为至少常见的8个大类。具体来说，像净资产同比增长率、总资产同比增长率、营业收入同比增长率、扣非净利润同比增长率，可定义为成长因子（F1）；流动比率、速动比率、资产负债率等财务指标，可以定义为偿债因子（F2）；净资产收益率、总资产净利率、每股收益、每股收益EPS增长率等与公司盈利能力有关的指标，可以定义为盈利因子（F3）；总股本、流通股本、总市值、流通市值等指标可以定义为规模因子（F4）；现金营运指数、销售现金比率、应收账款周转率等可以定义为现金因子（F5）；市盈率、市净率、市销率等指标可以定义为估值因子（F6）；存货周转率、毛利率等可以定义为运营因子（F7）；像股息率、股息率增长率等指标则可以定义为红利因子（F8）。此外，像公司治理因子、行业因子、公司外部政策因子等，也可以成为量化投资者重点关注的备选基本面因子大类。

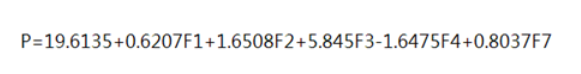
这里需要指出的是，过去多年国内多家券商等金融机构的实证研究显示，成长因子、盈利因子、估值因子、偿债因子等大类因子，已经被普遍证实，其对于上市公司个体股价驱动的贡献率较为显著。也即是说，在基本面量化模型中，上述几个大类因子，能够更好地反映出上市公司基本面与股价涨跌的正相关关系。

****3、模型的构建与回测****

一般而言，上市公司基本面量化体系的模型构建，有两种常用方法：回归法和打分法。

所谓回归法，就是使用股票过去的收益率对基本面因子的大类进行回归，然后得到一个线性回归方程。在该回归方程回测有效和适当修正之后，再把最新的因子值带入回归方程得到一个对未来股票收益的预判，以此为依据进行选股。

例如，中国上市公司研究院对截至2016年5月3日（2015年年报披露结束后首个交易日）的样本数据所做的因子分析显示，将公因子作为自变量与股票价格P做回归分析，经统计检验后，该机构认为，影响股票价格的主要因子有5个，具体量化模型如下：

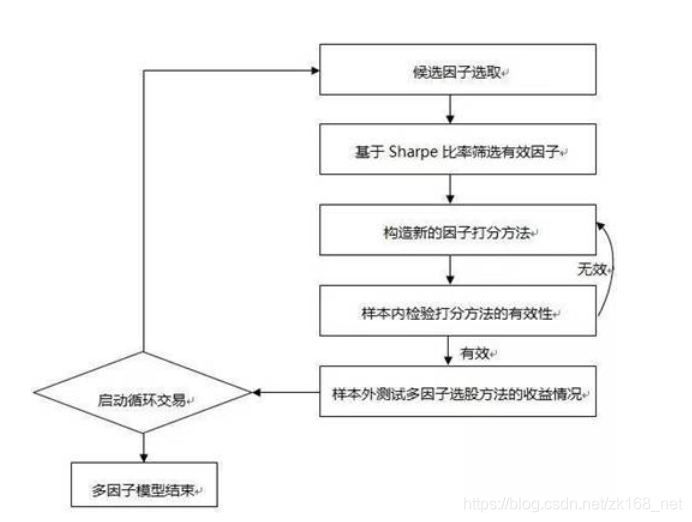


其中，F1为成长因子，F2为偿债因子，F3为盈利因子，F4为规模因子，F7为运营因子；不同类型因子的系数大小，直观反映了其对于股价影响程度的高低。

所谓打分法，就是根据各个因子值进行排序，将其在所有股票因子值的相对位置给予该个股一个分数，再对个股的所有因子值进行加权平均得到该股票的最终得分。最后将所有个股的综合得分进行排序，挑选出分数较高的股票。一般来说，打分法相对较稳健，不容易受到极端值的影响，但因各个因子的权重需要人为设定，因此主观性也较强烈。一般而言，打分法确定上市公司基本面的优劣，可以用以下公式予以简单计算：

IMG_261

此外，在评价打分法模型的有效性时，可以先利用Sharpe比率（或其他常用的投资绩效评价指标）进行有效因子的筛选通过，再构造打分方法，进而在样本股票内检验该打分方法的有效性，确认有效之后再在样本外测试该打分法选股的收益情况。往复统计检验之后，最终确定最有效的因子类型和打分模型（见下图）。



来源：对冲笔记