参考资料整理

1. 券商研报

华泰金工《基金评价及筛选全流程研究框架——股票型与债券型基金多种维度定量与定性评价法》：<https://mp.weixin.qq.com/s/VufbD_IgL6aZrpx81yPw8g>

建投金工《基于净值数据的Campisi型债基归因模型》：<https://mp.weixin.qq.com/s/SFj-4OS7h5uadmZ7TdO1TA>

方正固收《债券组合业绩归因模型——框架、案例及分析》：<https://mp.weixin.qq.com/s/BINFFlZlq15BXEDN2Dow4A>

模型基本和本文相同，但也没有基准选取，直接拿重仓券假设组合做基准算阿尔法。最后简单地做了直方图、一元线性回归、相关系数

兴证金工《债券基金评价体系及基于调整的Campisi模型的业绩归因》：<https://mp.weixin.qq.com/s/enMFGJbMmPqN9XTAkVxjHQ>

模拟组合收入效应期初价格用了净价。根据信用评级和期限选择信用债指数，指数和国债收益率差额作为信用利差，以此推算基金利差效应，而非用基准组合倒推出利差变化。真实基金用半年报和年报披露的利率敏感性数据推算有效久期，财务数据估计收入效应，但利差和择券还是用前五大近似。此外，还提出了用重仓券数量或市值变化测算换手率。主要是债券基金市场现状统计、模型理论分析，实际案例只有单个基金，没有深入研究时间序列上债券基金整体的情况。也没有阿尔法。

海通金工《因子剥离体系下的债券基金久期估测构想》：<https://mp.weixin.qq.com/s/ZDuH2Wv5kRm-JL-xLrRNgg>

1. 论文等

Campisi 2000, Primer on Fixed Income Performance Attribution

CIPM Reading 5, Return Attribution

中债指数编制说明：<http://indices.chinabond.com.cn/cbweb-mn/int/int_yield_zs_doc>

CIPM Reading 5主要介绍权益基金的Brinson业绩归因模型，非常简要地提了一下债券基金，但对两个概念作了区分：

一是return attribution，其试图解释的是基金相对某基准的超额收益，比如Brinson模型，其试图解释的是R-B从配置、个股、交互三个角度的来源，而不是解释R，这种分析按定义必须有一个明确的比较基准。

二是return contribution，其试图解释的是基金各部分对其总收益的贡献，或者叫做absolute return attribution，即这是在解释R，例如某只基金2020年上半年总收益率为30%，股票A贡献了10%，股票B贡献了15%，等等。华泰和建投报告里面对Campisi模型实际上只涉及return contribution的层面，即解释某只债券基金总收益率由票息、国债、利差分别提供了多少，并没有用到基准，分析的并不是相对基准的超额收益。

方正固收的报告则有涉及超额收益归因，与Campisi (2000)基本一致。概括来说，就是对基金和基准分别做return contribution，将二者的总收益R和B各自分别分解到收入、国债、利差、择券效应，再把对应效应对减就可以得到基金在各个效应上获得的alpha，后续详述。

注意如果不与基准比较，只对某基金做return contribution时，是不存在择券效应的。以及在做return attribution时，对基准的return contribution也不存在择券效应，只有待分析基金的return contribution中存在择券效应。

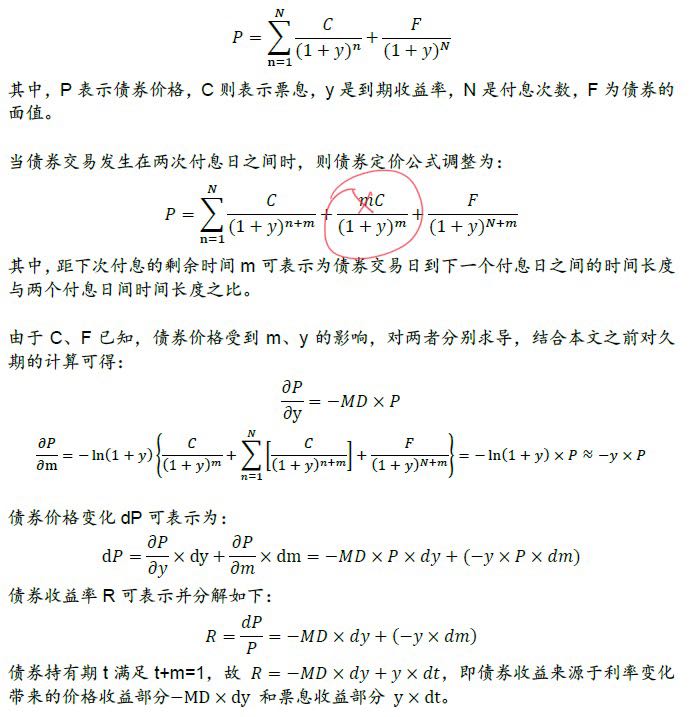
模型原理与推导

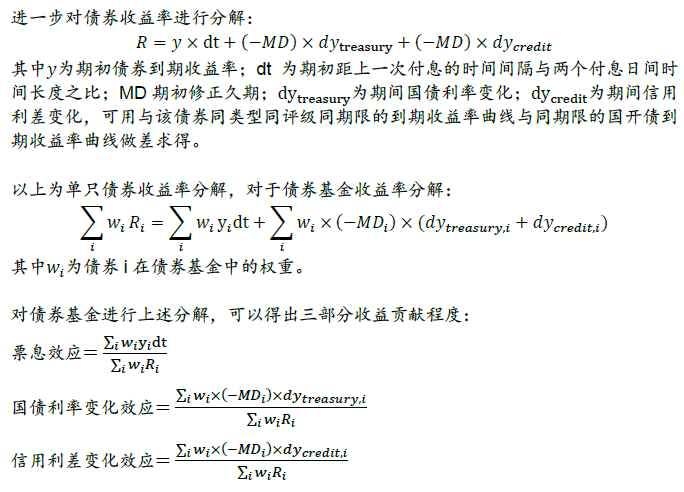
1. 瞬时收益分解

Campisi模型在return contribution层面的原理基于债券价格决定公式的全微分，各种效应可类比期权Greeks来帮助理解。在票息和面值事先确定的情况下，和都是常数，因此可看作YTM和时间的二元函数.

对其进行全微分就是，接下来只不过是基于的公式求解两个偏导数的具体解析式，并通过等价无穷小和久期的定义将其改写成；注意到YTM由无风险利率和风险溢价组成，，把三个加数依次称为国债效应、利差效应、收入效应。具体的推导过程自华泰和建投两篇报告里面都有涉及，且最后跑模型的时候只需要用到最终分解结果，而不太关注推导过程，因此此处不再赘述，可直接参考下图。

这种设定下，我们假设的在两次付息间的某一个瞬间，持有某只债券，全微分的含义实际上是这一瞬间时间流逝和YTM变化带来的价格变化各自对该瞬间债券收益率的影响，因此公式当中的应当使用全价，而不是净价。华泰和建投报告中的两次付息日之间债券定价调整公式第二项分子使用了而不是，含义上更接近净价而不是全价，应该是笔误；使用是无法求得偏导数等一系列图示后续推导结果的，只有使用即全价公式才可以。





债券基金只不过是债券的投资组合，因此组合瞬间收益就是各债券瞬间收益的加权平均，即，去括号之后得到的就是组合层面的三个效应，如上图所示。

1. 一阶近似

相较某只债券在瞬间收益的分解，我们一般更关注持有一段时间后总收益的分解，以用全微分作一阶近似为出发点，有，和分别为该债券在期初的久期和YTM，为与债券期限匹配的国债收益率变化，为与债券类型、期限和评级匹配的信用利差变化，为持有时长；对于投资组合，类似地，在加权平均公式中把换成即可。

但是实际操作起来会有三个问题：一是一阶近似本身就忽略了高阶项，用持有期债券价格变化计算总收益得到左边，结合期初久期、YTM、国债收益率、信用利差等数据算出右边，二者基本是不可能相等的，只是约等，但归因模型要保证各部分综合等于总收益率。二是与票息关系不大，其表达式中并不含有。三是全微分及后续一阶近似基于的假设是在某瞬间的邻域内，当债券持有期跨过了票息支付日，即持有期间有收到票息时，期初和期末债券定价公式的形式除了和以外，期数也发生了变化，之前的推导过程不一定能继续成立。

1和2的作用更多是帮助从数学上理解债券收益构成，从3开始才是真正会使用到的模型。

1. 改进的return contribution模型

（1）模型所需数据输入

**债券基金和其基准各自总体层面的：期初久期、票息、期初全价、总收益率。以及期初和期末的国债收益率曲线。**债券基金的基准一般都是某个债券指数。基金总收益率用复权单位净值即可计算得到，类似地基准总收益率用指数计算。在个券上会有期初全价或净价的差别，但基金和基准直接使用期初单位净值和指数值即可。期初久期和票息相对更难获得，依赖于基金管理人和指数提供者来提供。

当前目标是将某债券基金整体相对于其基准整体的超额收益拆分到收入、国债、利差、择券四个效应上，无需涉及行业或个券，因此先假设可获取到基金和基准整体层面的输入参数。否则，可能需要手动通过持仓或重仓券信息进行加权计算。但考虑到即便我们可以直接获取到整体层面的数据，基金管理人和指数提供者在计算这些数据时仍然也得采用加权平均法，最好假设在研究期或持有期内基金持仓不发生太大的变化。由于基金只在季报、中报、年报中披露持仓信息，现实中最能符合假设的操作方法是研究两次报告披露期间的业绩，假设期间持仓没有发生变化。

同时，Campisi在论文中还提到了计算基金和基准整体的各效应时，除了使用整体层面的输入参数以外，每个持仓债券单独计算后加总理论上可以得到更加精准的结果，但现实中是不可行的，比如债券指数中含有的债券数量可能非常庞大。此外，债券基金报告只会披露前十大重仓债券；而相对股票基金的前十大重仓股，债券基金的前十大重仓债券所占比例多数时候偏低。但Campisi认为，不论用哪种方法，重要的是我们对基金和基准的两次return contribution分析所采用的精确度要相同。因此，在Campisi模型研究和搭建的起步阶段，我们先只在整体层次进行计算，后续如果有更高精细度的需求，可以用模块化的形式继续改进。

（2）对基准进行return contribution分析

为了避免2中所提到的一系列问题，需要调整收入效应和利差效应的计算方式，具体流程如下：

**收入效应=持有期票息/买入价格**，类似股票的dividend yield。买入价格即期初指数点位。持有期票息=面值\*持有期票息率，而持有期票息率=年票息率\*持有期时长，比如基金期初年票息率为5%，那么年报和一季报之间对应研究期的持有期票息率就是1.25%。注意票息、面值、价格要对应，尤其是在基金和基准整体层面，相对单只债券更容易对应错误。这种计算方式实际上并不依赖于持有期真实收到的票息，而是把整体票息率在时间跨度上平均摊开。相比基于真实票息偿付来计算，这种方式实际上更合理，一是因为基金和基准组合中含有大量债券，付息时间可能非常分散，逐一追踪所需消耗的工作量巨大，而且基金和基准组合应该是集中几次分红，而不是每当组合中某债券偿付票息就分红一次；二是因为即便是对单只债券而言，在两次付息期间全价也会由于应计利息而增长，若用真实付息计算，此期间就没有收入效应，这显然是不合理的。

**国债效应=-期初久期\*国债收益率变化**。根据期初久期，在期初和期末的国债收益率曲线上找到对应的点，作差即可得到对应久期的国债收益率变化。有两点细节：一是Campisi并没有考虑随着时间流逝，在持仓不变的假设下久期会不断变短，从而相应调整期末所使用的国债收益率期限，而是期初期末的期限都与期初久期对应即可。二是国债收益率曲线本质上是离散、分段线性的，可能需要通过线性插值来找到和期初久期对应的收益率。

**总收益率用指数点位计算即可。**

**利差效应=总收益率-收入效应-国债效应**，即利差效应是倒算得出的，以避免对各行业、公司利差变化的繁琐计算；或者说，利差效应这里是以余项形式存在，为了保证归因模型各部分之和等于总收益，有点类似Brinson模型中的交互收益。这是Campisi模型的巧妙之处。在不涉及基准只对基金本身做return contribution，或者是return attribution分析中第一步对基准做return contribution时，没有择券效应，利差效应是余项；只有return attribution分析中第二步对基金做return contribution，以及后续要解释alpha的时候，才会出现择券效应，这时候择券效应才是配平的余项，下一部分中会详细说明。

**利差变化=-利差效应/期初久期**，这里我们进一步倒算出基准所对应的利差变化。

（3）对基金进行return contribution分析

基金的收入效应和国债效应与基准完全相同，不再赘述，只需相应使用基金参数，如复权单位净值等即可。

择券效应和利差效应此时应当作为一个整体来理解：如果基金完全被动跟踪基准，它们就应当具有相同的利差、利差变化和利差效应。**如果基金与基准的利差效应不同，就说明基金经理偏离了基准，进行了择券的决策，从而提供了利差效应的差异，这种差异即为择券效应**。基于这个逻辑，基金的利差效应和择券效应计算方法如下：

**利差效应=-期初久期\*基准利差变化**，此处利差效应不再是余项，因而不再是倒算得出，但注意使用的利差变化应当是基准return contribution倒算所得，而不是基金的利差变化。

**总收益率用复权单位净值计算即可。**

**择券效应=总收益率-收入效应-国债效应-利差效应**。总收益率-收入效应-国债效应也可视为只对基金做return contribution时所得基金的实际利差效应，那么择券效应也可以理解为基金和基准实际利差效应的差值。

（4）Alpha分析

两次单独的return contribution做完之后，我们就可以进行return attribution分析，将基金的超额收益分解到四个效应所提供的alpha上，非常直观：

**收入效应alpha=基金收入效应-基准收入效应，国债和利差计算方法完全相同。**

**择券效应alpha=基金择券效应**，因为基准无择券效应。

（5）算例

这里提供一个简单、虚构的算例，以具体展示计算流程，假设持有期为半年，可参见Excel表格。

|  |  |  |
| --- | --- | --- |
| 输入参数 | 基金 | 基准 |
| 期初久期 | 4 | 6 |
| 票息率 | 4% | 5% |
| 期初全价 | 103 | 105 |
| 总收益率 | 1.2% | 1.5% |

基金收入效应=4%\*100\*0.5/103=1.94%，同理基准收入效应为2.38%，对减得到收入效应alpha为-0.44%。

与期初久期对应，假设持有期4年期国债收益率下降了20bps，6年期下降了30bps，因此基金国债效应=-0.2%\*(-4)=0.8%，同理基准国债效应为1.8%，对减得到国债效应alpha为-1%。

接着，基准利差效应=基准总收益率-基准收入效应-基准国债效应=1.5%-2.38%-1.8%=-2.68%，基准利差变化=-2.68%/(-6)=0.45%，即基准的利差上升了45bps。

那么基金利差效应=-4\*0.45%=-1.79%，基金择券效应=1.2%-1.94%-0.8%+1.79%=0.25%，利差效应alpha由两个利差效应对减得到0.89%，择券效应alpha就是0.25%。

在模块化的设计思路中，以上过程就是Campisi模型的核心单元，已经可以直接具体地使用，后续工作需要的就是输入参数的可获得性，如果无法直接获得，需要考虑手动计算，比如国债收益率插值模块、组合久期和票息率计算等。就国债收益率而言，Wind上能获取到的期限已经比较完整，有0年（隔夜），1、2、3、6、9月，1-10年，15、20、30、40、50年，数据来源于中债估值中心。中债估值中心在构建到期收益率曲线时使用的是Hermite多项式法，我们在Campisi模型中插值时使用分段线性即可，计算复杂度不高，又由于中债估值中心提供的权限较完整而能得到较高的精确度。

以上分析将债券基金和其基准视为两个long-only且fully-invested的组合，Campisi模型和现有券商报告中也暗含如此的假定，但这种简化可能更适用于基准：债券基金一般会使用杠杆来增强收益，杠杆成本和运作对模型理论上应当具有一定影响。但实际中杠杆不会非常高，2014年证监会第104号令《公开募集证券投资基金运作管理办法》要求债券基金80%以上的基金资产需要投资于债券，同时总资产不能超过净资产的140%。兴证金工计算得出的2020Q1纯债型基金仓位中位数为116.8%，杠杆水平中位数为122.15%。暂时忽略这个问题有一定合理性，但在有杠杆情况下的理论模型修正不失为一个后续研究方向。

1. 期初久期和票息率输入

基准的期初久期和票息率部分情况下Wind可直接获取得到，基金则任何时刻都无法直接获取。因此我们需要手动估计基准在无法直接获取时和基金的期初久期和票息率，常用方法有基于持仓数据（holdings-based）和基于净值数据（returns-based）两大类。

（1）持仓法

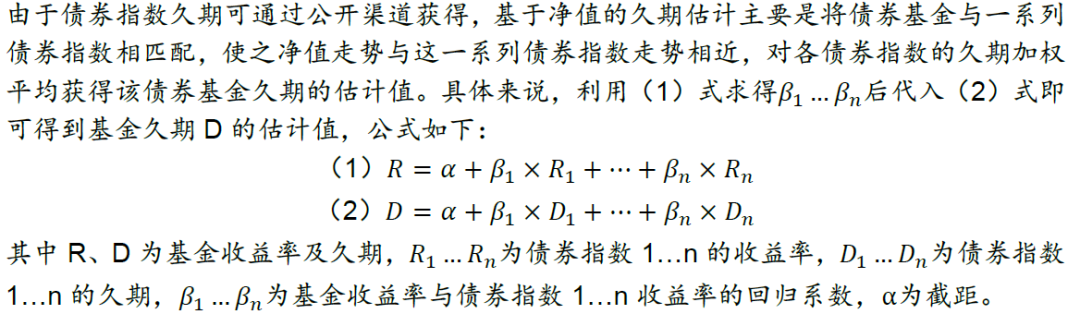
理论上，只要有完整的持仓数据，就可以通过加权平均法计算某个债券组合的期初久期和票息率，组合久期应当用市值加权法，票息率应当用面值加权法。

实践中证监会只要求债券基金在报告中披露前五大重仓债券，我们绝大多数时候都无法获取完整信息，只能将可见重仓券权重标准化后作为一个组合计算其久期，来近似基金久期。这是行业中从持仓出发测算的一般方法，在披露方式不改变的情况下，可改进空间不大。

虽然ABS和处于转股期的可转债的持仓信息是可全部获取的，但是这类资产持仓比例一般来说是非常小的，将其考虑进来充分计算的边际收益并不高。重仓券标准化近似的准确性明显取决于集中度：对持仓较集中的债券基金来说，这种方法的近似较好。基准在这个问题上相对更易处理，如果基准是某债券指数，有可能可以直接从Wind上获取到久期和票息率，若无法直接获取，也能得到完整持仓信息。

（2）净值法

持仓法的三大难点在于无法获取完整持仓信息、时效性不足以及无法得知期间调仓情况。净值法的一种思路是将研究期基金净值收益率回归到一系列债券指数上，再用回归系数去加权平均债券指数的久期，如下图所示：



由于其本质为回归，净值法自然就要求问题的结构在研究期内不发生结构性变化，具体来说就是基金久期在研究期内稳定；也有可能碰到多重共线性等问题，但是净值法只能用来估计久期，而无法应用在票息率上。操作上，同样可以用回归系数来加权平均指数的票息率，但没有理由预期这样得到的结果和真实情况以及持仓法估计值有任何接近之处，四只基金的2020Q2截面测试中也体现出了这样的特点。

海通金工研报中提供了另外一种净值法思路，但是是基于其先前构建的债基七因子体系的，无法单独使用。核心思想是其七因子模型中的Level因子表示了系统性风险，可近似理解为久期因子。因此，在每个研究期内，将一系列披露久期的债券指数先进行七因子分析，得到每个指数的Level因子暴露；然后将每个指数的久期回归到其Level因子暴露上，得到当期Level因子暴露和久期之前的关系，表现为一个一元回归方程久期=a+b\*Level因子暴露。然后，将所要分析的债券基金再做七因子分析，得到基金Level因子暴露，用前面估计的回归方程来预测其久期。

这种思路可能相对上述的简单回归更有理论基础，但有赖于七因子体系，因而当前只是简要提及一个可能的改进思路，后面可回头再考虑。

（3）基于财务数据

兴证金工研报中提到用基金季报数据来估计久期和收入效应，其中久期估计基于基金报告中提供的利率敏感性分析数据，收入效应则使用报告期内债券利息收入除以期初市值。但由于Wind接口无法提供细致到这种层面的财务数据，这种方法大规模操作起来可能不太方便，因此目前也是先简要提及，最终完善、改进模型时如果有需要可再考虑。

（4）测试主要结论

使用了广发政策性金融债（006869.OF）、华安年年红A（000227.OF）、鹏华丰融（000345.OF）、南方双元A（000997.OF）四只基金测算了2020Q2横截面的Campisi模型，主要分为核心业绩归因、票息率与久期测算两部分，其中南方双元A由于其业绩基准中证全债指数Wind上未提供期初久期、票息率等数据，核心业绩归因部分暂时无法进行。

票息率与久期测算同时使用了持仓法和净值法两种方法，且净值法考虑了中债全价、净价、财富三种指数，以横向比较各种方法和specification的表现，主要结论如下：

首先，票息率的测算确实只能用重仓券来近似计算，净值回归法如所预期的一样不可靠。全价、净价、财富三种指数测算所得的票息率之间差异极大，而且还出现了负数，这是明显不合理的。后续Campisi模型实现中，票息率测算部分我们均采用重仓券标准化近似。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 票息率 | 持仓法 | 净值法全价 | 净值法净价 | 净值法财富 |
| 华安年年红A | 4.732347 | -0.36931 | 0.420244 | 12.43731 |
| 鹏华丰荣 | 6.357107 | 2.557331 | 2.971477 | -4.99471 |
| 广发政金债 | 3.827407 | 1.855467 | 2.199874 | 2.174777 |
| 南方双元A | 3.43449 | 1.88284 | 2.194395 | 2.21263 |

其次，在重仓券权重和越大时，持仓法测算久期显然误差会越小。广发政金债五只重仓券占基金资产净值比例达到了103.09%，接近long-only以及fully-invested，持仓法和净值法估计的结果较接近，且此时我们认为持仓法的估计应当更接近真实情况。三种指数选择相比，全价指数与持仓法的估计最接近。鹏华丰融虽然五只重仓券占净值比只有33.79%，测算结果之间差异也不太大，全价指数在三种指数中与持仓法最接近。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 久期 | 持仓法 | 净值法全价 | 净值法净价 | 净值法财富 |
| 华安年年红A | 0.218357 | 1.327451 | 0.998213 | 1.719458 |
| 鹏华丰融 | 1.161266 | 1.362183 | 2.082092 | 1.657466 |
| 广发政金债 | 2.100063 | 2.028131 | 2.622705 | 2.648234 |
| 南方双元A | 2.927202 | 6.265097 | 6.600879 | 6.662891 |

华安年年红A与南方双元A的持仓法结果则与净值法结果有较大差异。注意到华安年年红A的五只重仓券占净值比为33.96%，但五只重仓券期初久期均低于0.4，该基金剩下的持仓似乎不太可能久期也低于0.4。换句话说，其五只重仓券刚好是即将到期的债券。顾名思义其基金经理有可能是通过重仓超短债、同时分散投资中长债来在实现长期收益的同时又实现每年都“红”，从而导致了持仓法的估计误差。但是，这两只基金净值法三种指数的估计结果之间是差不多的，一定程度上能够增强净值法此时的可信度。

一个粗略的选取规则可考虑在前五大重仓券占比高于50%时使用持仓法，否则使用净值法。如果基金明确了是国债、政策性银行债等利率债基金，指数使用中债-国债及政策性银行债指数系列，否则使用中债-信用债总指数系列；且基于广发政金债这一五只重仓券占比为103%的特例的观察，倾向于使用全价指数。

Python操作文档

1. 类和方法

（1）YieldCurveInterpolator

根据线性插值法计算某期限（以年为单位）的国债收益率变化。

\_\_init\_\_方法：传入参数startdate，enddate，“YYYY-MM-DD”格式，向Wind拉取对应日期21个期限的国债收益率并计算每个期限对应的差。

interpolate方法：传入参数tenor，return一个数值插值结果，单位%。

（2）ReturnContributor

在基金或基准层面做return contribution分析，给出收入、国债、利差、择券（基准择券效应为0）效应。

\_\_init\_\_方法：传入参数fund，yieldchange，spreadchange=None，其中fund必须是一个fund类的对象，yieldchange是国债收益率变化；如果传入的fund是基金，就必须额外提供基准利差变化spreadchange。

\_\_calincome方法：计算收入效应，单位%。

\_\_caltreasury方法：计算国债效应，单位%。

\_\_calspread方法：计算利差效应，单位%。

\_\_calselection方法：计算择券效应，单位%。

四个效应计算方法不return任何值，计算结果直接作为对象属性，且在\_\_init\_\_中被调用；生成完ReturnContribution对象后只需直接访问属性即可获得各个效应的值。

（3）ReturnAttributor

基金和基准已经做完了return contribution分析之后，进一步给出阿尔法分析。

\_\_init\_\_方法：传入参数fundc，bmkc，其中fundc必须是一个基金的ReturnContribution对象，bmkc必须是一个基准的ReturnContribution对象。

getfundc方法：return一个长度为5的list，前四个元素是基金的return contribution分析结果，依次为收入、国债、利差、择券，第五个元素是总收益率，单位均为%。

getbmkc方法：return一个长度为5的list，是基准的return contribution分析结果，依次为收入、国债、利差、择券（在基准的情形下就是0），第五个元素是总收益率，单位均为%。

getalphas方法：return一个长度为5的list，是最终的阿尔法分析结果，依次为收入、国债、利差、择券和总alpha，单位均为%。

（4）DebtFund

用于描述基金或基准组合的一个类，包含期初久期、票息率、面值、总收益率等组合参数作为对象的属性。

\_\_init\_\_方法：传入参数fcode，startdate，enddate，indicator，per；fcode为基金或基准的Wind代码，日期格式为YYYY-MM-DD，indicator取benchmark或fund，per为持有或研究期长度

\_\_caltotalret方法：从Wind拉取持有期收益率，基金用现金分红净值增长率，单位%。

\_\_calstartdirtyprice方法：从Wind拉取期初全价，基金用单位净值，基准指数用全价指数的期初点位。

\_\_getdata方法：对于基金，从Wind拉取前五大重仓券的名称、代码、数量、期初全价、久期、票息率等信息，并且计算市值权重、面值权重，用字典打包存储在对象属性中。

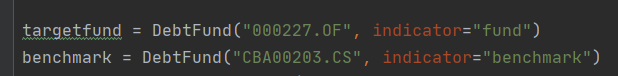
\_\_calmod方法：对于基金用市值加权法计算组合久期，对于基准指数通过Wind拉取或手动输入。

\_\_calcrate方法：对于基金用面值加权法计算票息率，对于基准指数通过Wind拉取或手动输入。

\_\_calpval方法：对于基金用持仓数据计算期初全价所对应的组合面值，对于基准指数用债券定价公式、票息率、剩余期限、期初全价、到期收益率倒推。

1. 使用方法

（1）生成两个DebtFund对象，一个基金，一个基准。



（2）给定起止日，生成国债收益率插值器对象。



（3）先对基准做return contribution，再对基金做，因为基金有赖于基准生成的利差变化作为参数传入，顺序不能调换。



（4）最后用两个ReturnContributor对象做阿尔法分析，得出结果

