**淘宝医药/娱乐生活指数构建方案——程序流程说明**

By 卡尔(2015-03-07)

**（一）回溯需要用到的数据表（共同）**

1- 基础数据表（读取到本地）

(1) 交易日期列表

(2) 行情表: 只取A股数据

(3) 自由流通市值表

注: 其中(2)(3) 用于计算流通市值

2- 回溯时间段设定:

(1) 回溯日期

(2) 调仓日期（每个月月末）

注:

1. 医药指数调仓日期: "31dec2008"d to "31dec2014"d;
2. 娱乐指数调仓日期: "31jan2011"d to "30sep2014"d;

**（二）医药指数**

* 指数构建基本想法:

(1) 根据富国因子(个股层面）和淘宝因子（需从行业层面->个股层面）计算出个股的综合因子。

(2) 选取综合因子得分>0的前N只股票构成指数成份股，并采取得分加权。

**一、参数表**

(1) taobao\_pct: 淘宝因子权重（目前设定为0.2）

(2) threshold: 成份股中最多N只股票（N=50)

**二、指数构建流程**

**1. Step1: 淘宝/富国因子准备工作**

**(1) 输入:**

1. taobao\_mapping

🡪 行业和个股映射表，两个指数共用。需定期人工维护。

* end\_date:
* stock\_code: 字符型(32)
* stock\_name：字符型(32)
* indus\_code：数字型(8)
* indus\_name: 字符型(32)
* indus\_wt: 数字型(8)，当一个股票属于多个行业时，用于设定不同行业的相对权重。 缺省值为1。表示各个行业之间等权。
* fund\_name: 字符型(32)，取值"TAOBAO\_MEDICINE"
* into\_time: 入库时间

注: 医药指数中，个股和行业之间存在一对多的关系。即同一天内，一只股票可能存在多个记录。

1. alipay\_out

🡪 淘宝输出表，两个指数共用（数据表由淘宝提供）

* day\_id
* company: fullgoal
* fg\_level2: 连接条件等于taobao\_indus\_code.indus\_code
* idx\_key: "fg\_factor" (只取该值下的条目)
* idx\_value：淘宝行业因子
* gmt\_create

1. taobao\_indus\_code

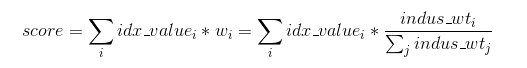
* 行业代码和行业简称的映射关系，两个指数共用。基本不变
* indus\_code: 数字型(8)
* indus\_name：字符型(32)
* into\_time： datetime20.
* fund\_name: 文本(32)，取值"TAOBAO\_MEDICINE"

提示: (a)(b)中只选取indus\_code或fg\_level2属于医药行业的记录（即taobao\_indus\_code中fund\_name = “TAOBAO\_MEDICINE”的行业代码）

**(2) 计算过程**

(2-1) 基于taobao\_mapping，从淘宝行业因子计算个股淘宝因子

1. 计算个股原始得分



其中idx\_value为股票所属行业的淘宝得分。

(2-2) 从富国因子得分表提取中证医药行业成份股及其得分。

目前该步骤的结果存在tinysoft.fg\_csi800\_factor中（已标准化后），需要问方旻如何得到该表

(2-3) 得分标准化

1. **以富国因子得分表中的成份股作为备选池**。注意，taobao\_mapping中可能出现不在该表中的股票。这些股票不进入备选股票池。（注：这里也可以进行提前确认，保证taobao\_mapping中的恰好就是所有的备选股票池）
2. 在备选池内对淘宝和富国因子的原始得分进行标准化

🡪 利用文汉之前提供给我的存储过程

注: 当前tinysoft.fg\_csi800\_factor中需要剔除"600851", "002614", "600490", "300061" 这四只股票。另外，600849在2010/2/28日还有因子值，但这一天这只股票已经退市(流通市值缺失)，需予以剔除。

* 判断是否退市/或暂停上市的方法，在回溯中只依靠close/factor/freeshare，是否有其他方式?

**(3) 输出:**

(a) taobao\_score\_s: （仅包括备选股票池）

个股因子得分，两个指数共用。

* end\_date
* stock\_code
* stock\_name
* taobao\_score：淘宝因子
* fg\_score：富国因子
* tot\_score：综合得分 🡪 待填入内容
* taobao\_pct: 淘宝得分权重（参数值）🡪 待填入值
* fund\_name: "TAOBAO\_MEDICINE"
* into\_time: 入库时间

1. taobao\_score\_i: （仅包括备选行业）

行业因子，两个指数共用。

* end\_date
* indus\_code: 富国行业代码
* indus\_name: 富国行业简称
* taobao\_score: 淘宝因子
* fg\_score: 富国因子（医药指数该项缺失）
* tot\_score：综合得分（医药指数该项缺失）
* taobao\_pct: 淘宝得分权重（参数值）（医药指数该项缺失）
* fund\_name: "TAOBAO\_MEDICINE"
* into\_time: 入库时间

**2. Step2: 确定综合因子得分**

**(1) 输入:**

(a) taobao\_score\_s: 个股因子得分

**(2) 计算过程:**

(a) 综合得分tot\_score（只针对个股因子得分表）



注意：计算时，淘宝/富国得分缺失都设定为0。

**(3) 输出:**

（a） taobao\_score\_s: 填写以下字段内容

* tot\_score：综合得分
* taobao\_pct: 淘宝得分权重（参数值）

**3. Step3: 确定成份股及其权重**

**(1) 输入:**

(a) taobao\_score\_s: 个股因子得分

**(2) 计算过程:**

(2-1) 剔除所有综合因子得分(tot\_score)<=的股票

(2-2) 在所有综合因子得分>0的股票中，按从高到低排序，选取得分前N只股票（至多）。注意: 若综合因子得分相同，则优先考虑富国因子。

(2-3) N只股票按照综合因子得分加权，确定权重。

**(3) 输出:**

（a) taobao\_stock\_pool

🡪 调仓后的股票池，两个指数共用。

* end\_date：调整日
* effective\_date：生效日
* stock\_code
* stock\_name
* weight
* fund\_name
* into\_time

**（三）娱乐生活指数**

* 指数构建基本想法:

(1) 根据富国因子（需从个股层面->行业层面）和淘宝因子（行业层面）计算出行业的综合因子。

(2) 指数由两部分组成: 50%基准部分 + 50%增强部分。（50%称为增强比例）

其中:

(2a) 基准部分: 所有成份股按照流通市值加权构建。

(2b) 增强部分: 只筛选出综合因子得分前50%行业(即行业得分大于所有行业的中位数)中的股票。这些行业之间按照综合因子得分加权配置，行业内部个股按照流通市值加权配置。

最终，个股权重由两部分的权重按照增强比例加权。且控制最大权重为10%。

**一、参数表**

(1) taobao\_pct: 淘宝因子权重（目前设定为0.5）

(2) improve\_pct: 增强比例（目前设定为0.5）

(3) stock\_upper: 个股最大权重（目前设定为0.1）

**二、指数构建流程**

**1. Step1: 淘宝/富国因子准备工作**

**(1) 输入:**与医药指数相同。

(a) taobao\_mapping

(b) alipay\_out

(c) taobao\_indus\_code

提示: (a)(b)中只选取indus\_code或fg\_level2属于医药行业的记录（即taobao\_indus\_code中fund\_name = “TAOBAO\_ENTERTAIN”的行业代码）

**(2) 计算过程**

(2a)

1. 从alipay\_out中提取淘宝行业因子得分，得到

**(2) 计算过程**

(2-1) 基于taobao\_mapping得到成份股。

(2-2) **以taobao\_mapping中的成份股作为备选池**。根据富国个股因子得分表，计算行业的富国得分。

目前该步骤的结果方旻是计算后直接输出成Excel给我的，需要问方旻如何得到该表的结果。

（2-3）根据行业淘宝/富国因子得分，得到个股淘宝/富国因子得分。（即同个行业中的股票，得分相同。单个股票只归属于一个行业。这个与医药指数的成份股特征是不同的）

**(3) 输出:** 与医药指数相同。

(a) taobao\_score\_s: （仅包括备选股票池）

1. taobao\_score\_i: （仅包括备选行业）

其中fund\_name = “TAOBAO\_ENTERTAIN”。

**2. Step2: 确定综合因子得分**

**(1) 输入:**

(a) taobao\_score\_s: 个股因子得分

(b) taobao\_score\_i: 行业因子得分

**(2) 计算过程:**

(a) 综合得分tot\_score（同时针对两个表）



注意：计算时，淘宝/富国得分缺失都设定为0。

**(3) 输出:**

（a） taobao\_score\_s: 填写字段内容

* tot\_score：综合得分
* taobao\_pct: 淘宝得分权重（参数值）

（b） taobao\_score\_i: 填写字段内容

* tot\_score：综合得分
* taobao\_pct: 淘宝得分权重（参数值）

注意：这个步骤为了与医药指数相似，在个股层面上也进行操作。但在同个行业中，个股的得分是一样的。另外，娱乐生活指数中，行业因子得分表中也许填写对应字段内容。（医药指数不需要改步骤）

**3. Step3: 确定成份股及其权重**

**(1) 输入:**

(a) taobao\_score\_s; 个股因子得分

(b) taobao\_score\_i: 行业因子得分

(c) taobao\_mapping: 需要用到个股到行业的映射

**(2) 计算过程:**

（2-1) 计算基准部分的个股权重：所有成份股按照流通市值加权配置。确定基准部分个股权重w\_b

(2-2) 计算增强部分的个股权重:

1. 选出行业得分前50%行业（大于所有行业的分的中位数即可），且要求行业得分>0。
2. 行业之间按照行业因子得分加权配置。行业内个股按照流通市值加权配置。确定增强部分个股权重w\_i。（未出现在增强行业中的成份股，该权重为0）

(2-3) 计算指数中个股权重

 (2-4) 调整个股权重，使得最大权重为&stock\_upper. (超过部分，在其他成分股中按各自的原始权重配置）

**(3) 输出:**

（a) taobao\_stock\_pool: 与医药指数相同。