

2018.10

宏观因子聚合择时模型

摘要：

- 任何市场在同一时刻都受到经济、社会等多种不同因素的影响，它们共同作用影响了市场的走向。因此结合多类因子的数据建立模型能够预测更准确预测市场走向。
- 以沪深 300 为代表的股指受宏观经济和周期性影响较大，宏观因子聚合择时策略通过分析宏观经济因子的表现对沪深 300 指数为标的资产的股指期货进行择时择量交易。我们通过宏观经济形势因子构造了 13 个宏观经济指标。
- 基于宏观因子聚合择时模型，自 2016 年 10 月 24 日到 2018 年 6 月 29 日实现了累计收益率 40.49%，夏普 0.19，最大回撤 4.90%。

团队名 Golden Five

曹宏键

邮箱：chi5chi5@163.com

李唯一

邮箱：504596213@qq.com

王奕能

邮箱：wyn9816@situ.edu.cn

辛干

邮箱：charlesxin@situ.edu.cn

郑奇波

邮箱：zhengqibo96@gmail.com

宏观因子聚合择时模型

目录

1. 问题概览.....	2
1.1. 背景介绍.....	2
1.2. 问题分析.....	2
2. 模型介绍.....	2
2.1. 指标构建.....	2
2.1.1. 因子的选取	2
2.1.2. 因子权重的确定	4
2.1.3. 因子的聚合	5
2.2. 市场信号的生成	5
2.2.1. 收盘价变动信号	5
2.2.2. 交易量信号	6
2.3. 市场信号的预测	6
3. 数据获取及处理	7
3.1. 数据来源.....	7
3.2. 数据处理.....	8
3.2.1. z-score 标准化:	8
3.2.2. 缺失值填充:	9
3.2.3. 工作日匹配:	9
3.2.4. 去除异常值:	9
3.2.5. 指数加权平均:	9
4. 策略构建及回测表现	9
4.1. 仅根据收益率指标建仓	9
4.1.1. 策略描述:	9
4.1.2. 策略表现:	10
4.2. 综合收益率指标和成交量指标.....	10
4.2.1. 策略描述:	10
4.2.2. 成交量指标:	11
4.2.3. 实际仓位	11
4.2.4. 策略表现:	12
4.3. 策略小结.....	12
5. 总结与展望	13

1. 问题概览

1.1. 背景介绍

中国证监会有关部门负责人 2010 年 2 月 20 日宣布,证监会已正式批复中国金融期货交易所沪深 300 股指期货合约和业务规则,至此股指期货市场的主要制度已全部发布。2010 年 2 月 22 日 9 时起,正式接受投资者开户申请。公布沪深 300 股指期货合约自 2010 年 4 月 16 日起正式上市交易。迄今我国股指期货市场已经有了相当程度的发展,作为中国资本市场改革前沿的新兴产品,它有着巨大的发展潜力。

股指期货市场对于投资者乃至整个中国资本市场都是至关重要的。投资者可以通过在股票市场和股指市场反向操作达到规避风险的目的。同时,它也具有发现价格的功能,通过在公开、高效的期货市场中众多投资者的竞价,有利于形成更能反映股票真实价值的股票价格;并且股指期货具有交易成本低、杠杆倍数高、指令执行速度快等优点,对信息的反应更快。另外股指期货的双向交易机制改变了传统工具只能通过价格上涨获取收益的单向盈利模式,加之采用保证金交易制度使交易成本大幅降低,因此也被机构投资者广泛用作资产配置的手段。

因此针对股指期货市场的整体走势构建有效的指标是十分必要的。

1.2. 问题分析

影响期货市场的因素有许多,要想预测市场走势就需要充分考虑到这些因素。由于这些不同种类的因素对于最终市场趋势的影响都有着不同的机理,而这些机理也是不明确的,所以我们选择采用数据驱动模型,对于找到的 100 个因子的数据进行分析,并进一步构建出 13 个大类指标用于最终的市场走势预测。

最后我们需要根据这 13 个宏观经济指标的情况来制定具体的交易策略进行回测,我们把这个看作是一个分类问题,利用决策树模型,我们从每日的 13 个指标的数据中得到最终的交易信号。

2. 模型介绍

2.1. 指标构建

2.1.1. 因子的选取

本题要求我们设计经济形势指示器来对宏观经济形势的变化做出合理的预测。众所周知,中国经济近十年来实现了突飞猛进的发展,最近几年的发展趋势稍微放缓。2015 年的股灾以及 2018 年以来,由于贸易战等因素的持续影响,国内股票市场今年以来一直呈现持续疲软的态势。今年以来,国家逐渐放开主要金融机构外资占比限制,加强金融监管,并进行贸易战,对国内的股票市场也产生了长远的影响。因此报告中我们将会采用多个宏观经济形势的指标来进行沪深 300 期货的择时择量交易。

报告中我们设计了如下表所示的 13 个经济指标指示器,由一百个宏观经济因子构成,这些指标中包括先行指标、传统指标、商品价格指标、汇率指标、债市指标、货币资金指标、创新指标、政策指标、贸易指标、股票市场指标、自然社会指标、宏观经济指数

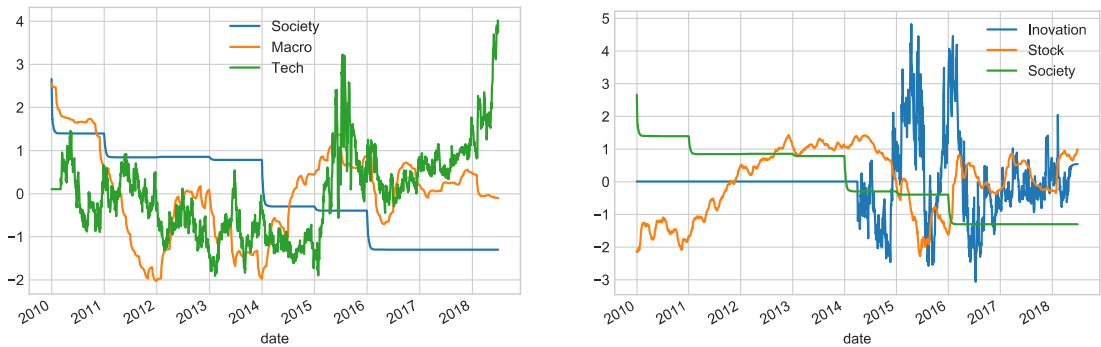
以及 HS300 技术指标等来对交易策略进行合理的调控。通过对这 13 个指标作为自变量的数据再次进行线性回归得到的权重系数，以及最终交易策略判断决策树的学习结果进行进一步分析，我们能够得出这 13 个指标相对的重要性大小。

根据再次拟合的权重大小以及决策树判断的先后顺序及分裂后的 GINI 值大小，我们可以看出这 13 个指标中自然社会指标、传统指标、汇率指标和商品价格指数对宏观经济形势预测的贡献率更大。

指标名称	指标含义	构成指标因子个数	指标线性拟合权重	T检验Pr(> t)	备注
先行指标	先行指数的成分指标	11	-0.80887	9.74E-06	先行指数成分
传统指标	反映传统经济性形势的指标	18	18.0457	<2E-16	
商品价格指数	反映主要商品价格变化比率的指标	5	16.86843	<2E-16	包括期货价格
汇率指标	世界主要币种指数和人民币汇率的加权	4	14.57366	<2E-16	人民币间接标价
债市指标	反映债券市场风险和收益的指标	5	3.85924	<2E-16	
货币资金指标	货币资金用途指标	2	2.38472	2.42E-06	
创新指标	余额宝情绪指数	1	0.19919	0.0103	14年以后数据
政策指标	税收等政策指标	4	0.04449	0.8579	
贸易指标	进出口数量和价格指数	5	2.04698	6.30E-07	
股票市场指标	股票市场总体趋势指标	6	6.08902	<2E-16	
自然社会指标	自然灾害、人口数量、农作产品等指标	5	19.04398	<2E-16	年度数据
宏观经济指数	宏观经济及其预期指数	10	3.4301	<2E-16	
HS300技术指标	沪深300技术指标	24	3.03226	<2E-16	来源于TA-Lib库

表 1：宏观经济指标概览

同时我们对这十三类指标在过去八年时间内的变化趋势做了简单的可视化分析：



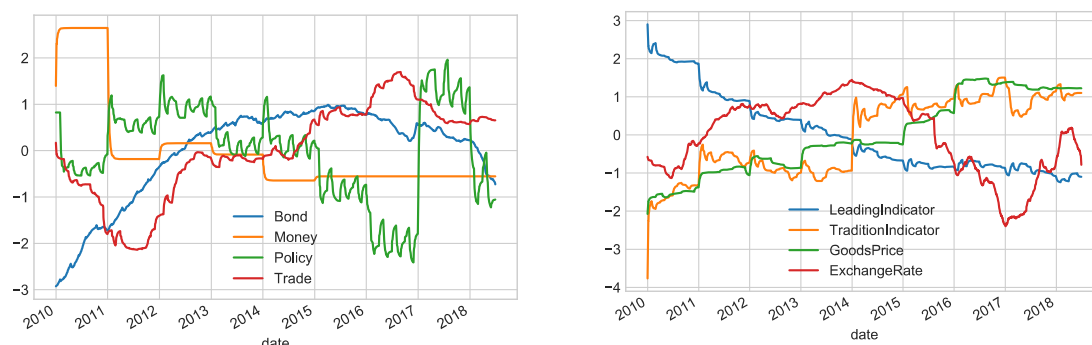


图 1：宏观经济指标走势

每个经济指标显示器分别由数量不等的因子按照一定的权重构成，下文中我们将会详细解释每个指标的构造方式和权重设置方法。多个指标依照决策树组合成交易策略-进行择时择量的买多和卖空。我们的策略结果显示：

	累计收益率	最大回撤	最大回撤发生时间段		收益回撤比	Sharpe Ratio
总计	40.49%	-4.90%	2017-04-06	2017-05-10	8.27	0.19
2016	4.36%	-1.27%	2016-12-19	2016-12-21	3.42	0.31
2017	27.04%	-4.90%	2017-04-06	2017-05-10	5.52	0.25
2018	9.09%	-2.65%	2018-02-08	2018-03-12	3.43	0.14

累计夏普：	0.19
累计收益率：	40.49%
累计最大回撤率：	-4.90%

表 2：交易策略结果简述

2.1.2. 因子权重的确定

首先我们需要确定 105 个不同因子 $\{x_i\}_k (k = 1, 2, \dots, 100)$ 对于收益率 $\{y_i\}$ 的贡献大小，为此我们将收益率作为因变量，将 105 个不同的因子作为自变量，首先进行多元线性拟合，并根据拟合结果中对应的系数大小来确定每个因子的权重。

令对应的系数为 $\{\omega_k\}$ ，则预测的收益率可以表示为：

$$\hat{y}(\omega, x) = \omega_0 + \sum_k \omega_k x_k$$

对于这个线性回归模型，求解系数 ω 我们选择采用普通最小二乘法。

目标是最小化预测值各点与实际值之间距离的平方和，即残差平方和，定义为 RSS：

$$RSS = \sum_i \left(\hat{y}_i - y_i \right)^2 = \sum_i (\omega_0 + \omega x_i - y_i)^2$$

求解这个问题可以直接寻找残差平方和函数的极值点，分别对系数求偏导并求出令导数等于 0 的点即为目标问题的解：

$$\frac{\partial RSS}{\partial \omega} = 2 \sum_i \omega (\omega_0 + \omega x_i - y_i)$$

$$\frac{\partial RSS}{\partial \omega_0} = 2 \sum_i (\omega_0 + \omega x_i - y_i)$$

直接利用矩阵运算可以得到因子收益矩阵的简洁的表达式：

$$W = (X'X)^{-1}X'Y$$

$$\text{where: } Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{pmatrix}$$

$$X = \begin{pmatrix} x_1 & x_1 & \cdots & x_{1,105} & 1 \\ & \ddots & & & 1 \\ & & \ddots & & 1 \\ x_i & x_{i,100} & \cdots & x_{i,105} & 1 \end{pmatrix}$$

$$W = (\omega_1 \ \omega_2 \ \cdots \ \omega_{105} \ \omega_0)$$

2.1.3. 因子的聚合

根据上一步中105个因子对收益率的线性回归，我们根据贡献性舍弃其中5个因子，最终可以获得其中100只因子的因子收益率 W ：

$$W = (\omega_1 \ \omega_2 \ \cdots \ \omega_{100} \ \omega_0)$$

对于这100个因子，我们事先根据具体指标的意义和性质对其进行了分类，总共分类到之前所述的13个指标。而对于每个分类指标下的因子，按照因子贡献率的大小确定其在分类指标下的权重大小，可以依据 W 进行加权和计算构造不同分类指标的数值 $F_t(t = 1, 2, \dots, 13)$ ：

$$F_t = \sum_{x_j \text{ considered in } F_t} \omega_j x_j$$

通过这样的方式，我们构造出13个指标的数值，由于每个指标中含有的小因子数以及权重并不相同，所以通过以下方式进行数据的标准化处理：

$$\tilde{F}_t = \frac{F_t - \mu}{\sigma}$$

$$\text{where } \mu = \frac{\sum_i F_{t,i}}{n}, (n \text{ is the number of data})$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_i (F_{t,i} - \mu)^2}$$

通过这样的处理，我们就获得了13个指标下的数值结构，之后我们又利用这些指标进行线性回归，得到它们对应的权重。

2.2. 市场信号的生成

2.2.1. 收盘价变动信号

考虑到我们模型样本点数目有限（2064 个），准确预测收盘价或是收盘价变化率难度较大（在实践中也验证了这个观点），在经过多次实验之后，我们最终决定预测收盘价的变动方向 D ，进而将预测问题转化成为分类问题。

考虑到我们的操作频率是周以及手续费等因素,我们根据定义收盘价变化率以及变动方向为:

$$R_i = \frac{(p_{i+5} - p_i)}{p_i} \times 100\%$$

- 当 $-1\% < R_i < 1\%$, 我们定义 $D_i = 0$;
- 当 $1\% < R_i$, 我们定义 $D_i = 1$;
- 当 $R_i < -1\%$, 我们定义 $D_i = -1$ 。

2.2.2. 交易量信号

除了收盘价变动方向的这一因素外,我们还考虑到如果交易量上升,则表示市场表现活跃,热度上升,但同时潜在的下跌风险也增大,反之亦然。因此我们利用前面使用过的线性拟合来对交易量 $Volume$ 进行预测。

$$\ln(Volume(\omega, x)) = \omega_0 + \sum_k \omega_k x_k$$

2.3. 市场信号的预测

为了进行回测,我们使用决策树对整理出的 13 个指标进行学习,得到最终的三个交易信号,分别是以-1,0,1 来表示。对于交易量指标,我们利用日交易量对数值对 13 个指标进行多元线性拟合预测未来交易量走势。由于前文已经介绍过多元线性模型,在这里我们具体介绍一下决策树模型

决策树是属于机器学习监督学习分类算法中比较简单的一种,决策树是一个预测模型;他代表的是对象属性与对象值之间的一种映射关系。树中每个节点表示某个对象,而每个分叉路径则代表的某个可能的属性值,而每个叶结点则对应从根节点到该叶节点所经历的路径所表示的对象的值。

决策树的训练我们选择CART算法,最终生成的是一个典型的二叉决策树。CART算法中为了表示每个节点对属性进行分裂后子样本的纯净度大小,定义了GINI值,对于样本集S:

$$GINI(S) = 1 - \sum P_k^2$$

where P_k is the frequency of the k^{th} attribute in S

对于含有N个样本的样本集S,根据属性A的第i个属性值,将样本划分为两部分,划分完成后可以计算新的GINI值Gain_GINI:

$$Gain_GINI_{A,i}(S) = \frac{n_1}{N} GINI(S_1) + \frac{n_2}{N} GINI(S_2)$$

where n_i is the number of samples in S_i

对于属性A,分别计算任意属性值将数据集划分成两部分之后的Gain_GINI,选取其中的最小值,作为属性A得到的最优二分方案:

$$\min_{i \in A} (Gain_GINI_{A,i}(S))$$

对于样本集S,计算所有属性的最优二分方案,选取其中的最小值,作为样本集S的最优二分方案:

$$\min_{A \in \text{Attributes}} \left(\min_{i \in A} (\text{Gain_GINI}_{A,i}(S)) \right)$$

所得到的属性A及其i个属性值，即为样本集S的最优分裂属性以及最优分裂属性值。

最终我们得到如下的决策树结构，图中可以看到每个叶结点的GINI值，判断标准以及样本在该节点对应于三类别的数量。

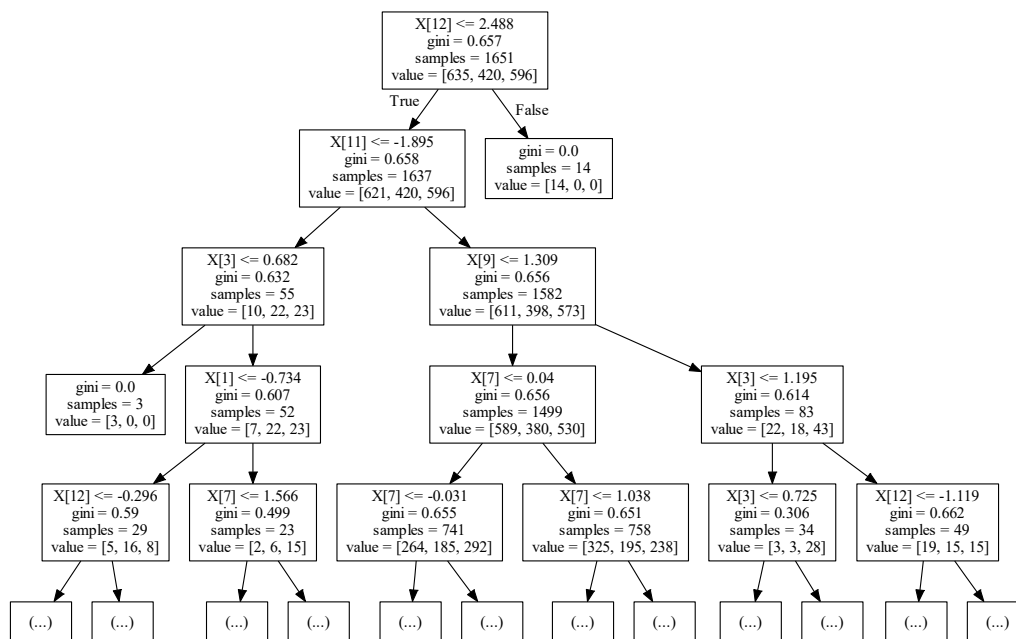


图 2: 决策树前五层结构

对决策树对每个指标的判断先后顺序以及对应的每一类的样本数进行分析,也能够得到每个指标对于最终结果的影响程度大小。

3. 数据获取及处理

3.1. 数据来源

本题要求预测整个经济市场的状况。我们选择以中国宏观经济市场作为研究对象。从而,我们对因子检验的样本空间选取为全国 2009-2018 年各类指标。

针对整体宏观经济，我们首先选择了可以预测短期经济总体情况的先行指标，包括了货币供应量、财政预算收入、能源消费量等。其次是直接反应宏观经济总体情况的传统指标以及重要的宏观经济指数，包含了 GDP、GDP 增长率、失业率、居民消费指数等。由于金融业在现代经济占据核心地位，是调控宏观经济的重要杠杆，因此我们重点从金融市场选择指标，分析测试其影响。我们针对股票市场、外汇市场、债券市场等部分，各自选取了对应的因子，反映其收益与波动情况。金融市场之外，考虑到几类重要商品在国民经

济中占据重要地位，我们也选取了这些商品的价格等数据作为因子，其中包括原油、商品房、煤炭和农产品等。世界各经济体互相沟通影响，我们也把国际贸易与外商投资作为指标之一。宏观经济除了受经济自身因素影响外，还受自然环境与社会多重因素的影响。因此，我们加入了自然社会、政策等作为又一大类指标。除了这些较为传统的指标，我们还选取了余额宝情绪指数作为创新指标，以此反应散户情绪与股市间的交集关系。最终我们选取了 100 个宏观经济因子，并且基于此构建了 13 个经济指标。分类因子构建情况如下图所示，100 个因子的详细构造方式、数据来源以及因子权重见附录。

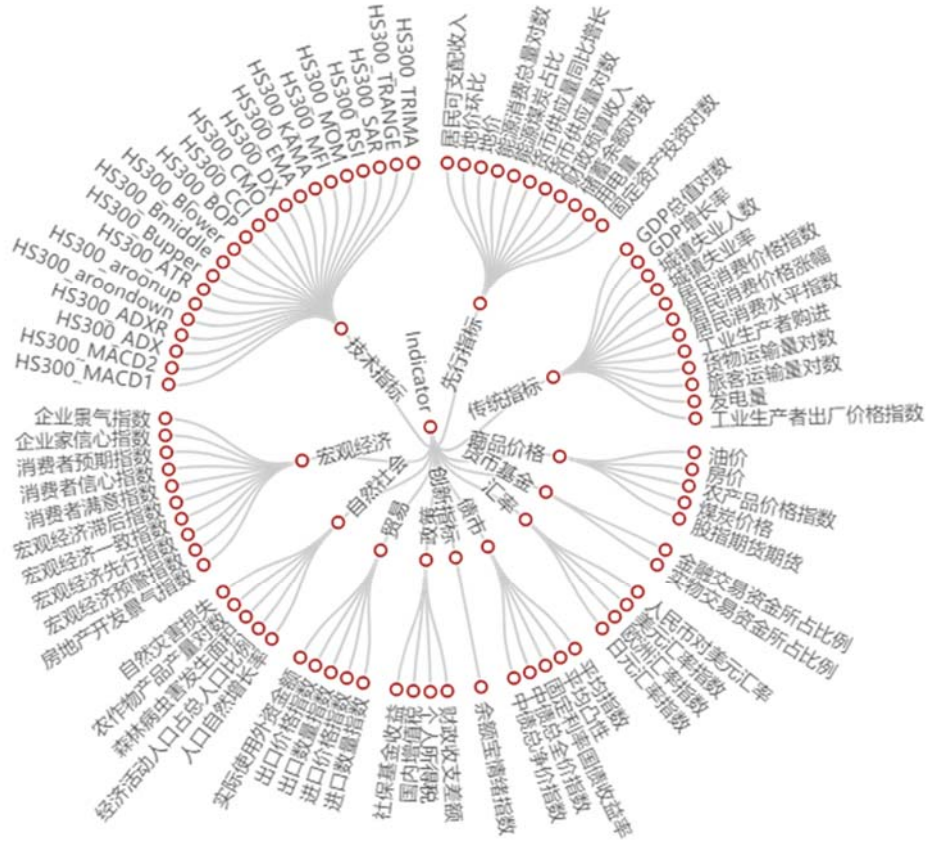


图 3：指标组成结构

3.2. 数据处理

考虑到数据本身存在数量级不同、数据缺失、异常值等问题，我们需要对数据进行初步处理，处理环境为 python 3.6.4。

3.2.1. z-score 标准化：

在我们的评价体系中，由于各评价因子的性质不同，量纲和数量级不同，如果直接永远指标值进行分析，就会影响我们对数值较高的指标在综合分析中权重的判断，相对削弱了其他因子的作用，因此，为了保证结果的可靠性，需要对原始指标数据进行标准化处理。这里我们选择是比较常用 z-score 标准化，经过处理后的数据因子均值为 0，标准差为 1：

$$x' = \frac{x_i - \bar{x}}{s}, \text{ 这里 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

3.2.2. 缺失值填充:

在我们的评价体系中,各评价因子由于公布频率不同(日频,月评,季频,年频)以及数据来源有限等原因,数据存在缺失的问题,由于大多数缺失是由于公布频率较低导致的,所以在这里我们采用的是前向填充(将缺失值替换为前一个有效值)的方法来替换缺失值。

3.2.3. 工作日匹配:

由于我们的目标指数是沪深 300,所以我们根据沪深 300 的日期索引提取出评价因子中工作日的日期即可。

3.2.4. 去除异常值:

考虑数据在传播过程中失真以及统计误差等原因,评价因子数据中存在个别异常值,在本次报告中,我们采用的是拉依达准则法(3σ)来识别异常值并将其剔除:我们设μ与σ分别表示单个指标的数学期望和标准差,我们认为在实际情况中出现μ+3σ以及μ-3σ的概率是很小的,因此对于大于μ+3σ的数据,我们视为μ+3σ处理;同理,对于小于μ-3σ,视作μ-3σ处理。

3.2.5. 指数加权平均:

考虑到因子的动量效应,在使用因子的过程中,我们经常需要用到因子加权平均来综合考虑近期因子与远期因子的影响,在这次报告中,我们采用了指数加权移动平均(Exponential Weighted Moving Average)法:以当前日期以及前四天作为一个时间窗口,计算出日期的权重向量,进而用于平滑数据,对于边界上的点我们也相应使用了四天期、三天期权重向量进行平滑。

$$\frac{[e^0, e^{0.25}, e^{0.5}, e^{0.75}, e^1]}{\sum_{i=0}^4 e^{i/4}} = [0.11, 0.15, 0.19, 0.24, 0.31]$$

4. 策略构建及回测表现

由于机器学习对大量训练集的需要,我们在有限的日期时间段上划分前 80%作为训练集,后 20%作为预测集。最终回测时间段为 2016 年 10 月至 2018 年 7 月。所有交易频率均为周。

4.1. 仅根据收益率指标建仓

4.1.1. 策略描述:

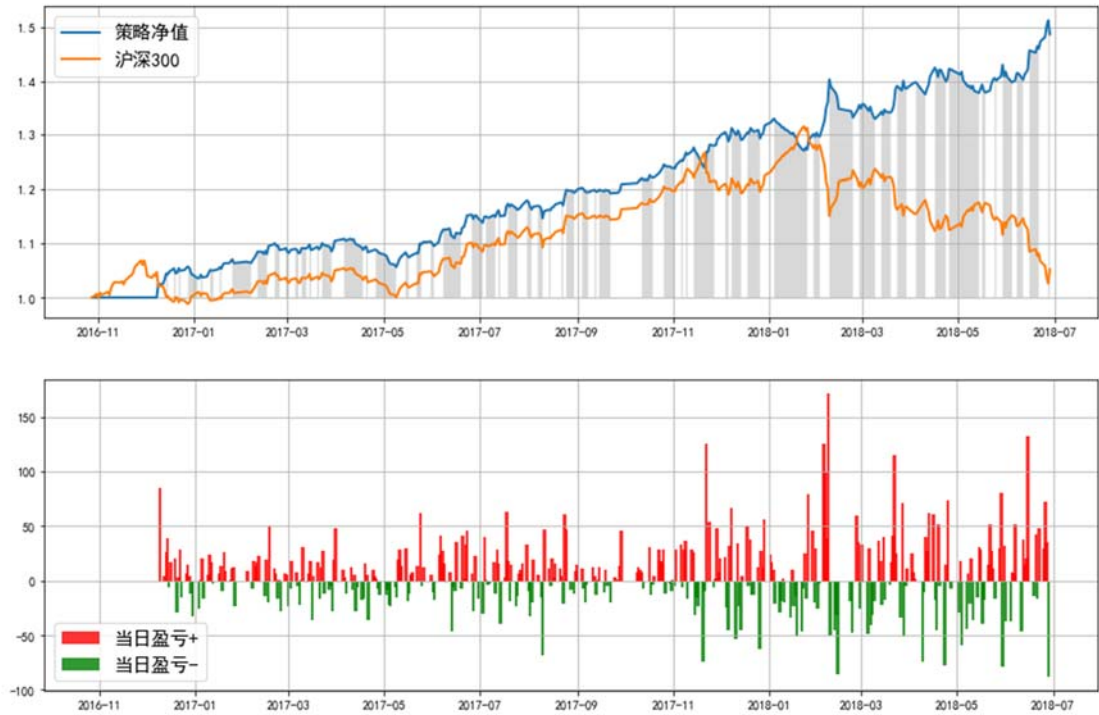
以沪深 300 为标的,通过开多和开空 IF 期货合约,达到相应的做多做空的目的。

当收益率指标对未来行情做出正面预测时(预测值为 1),开多一张 IF 期货合约;反之,

开空一张 IF 期货合约。当合约到期时，自动平仓并根据当前时间的指标信号重新开仓。通过上述操作，使仓位完全符合收益率指标的指示。

回测中选取股指期货交易手续费万分之七，回测框架基于 Python。

4.1.2. 策略表现：



注：上图纵轴为策略净值（初始值为 1.00），下图纵轴为当日盈亏（无乘数）。计算方式均为单利。

图 4：策略一净值与当日盈亏变化图

	累计收益率	最大回撤	最大回撤发生时间段		收益回撤比	Sharpe Ratio
总计	48.67%	-5.20%	2018-02-08	2018-03-12	9.36	0.18
2016	4.37%	-1.27%	2016-12-19	2016-12-21	3.44	0.31
2017	27.39%	-4.71%	2017-04-06	2017-05-10	5.82	0.25
2018	16.90%	-5.20%	2018-02-08	2018-03-12	3.25	0.18

表 3：策略一的收益指标

4.2. 综合收益率指标和成交量指标

不难发现，收益率指标对市场行情大方向的判断是非常准确的。基于上一部分的策略，下一步我们综合考虑成交量指标，来进一步增强策略表现。

4.2.1. 策略描述：

首先，通过收益率指标，判断未来走势。

其次，通过成交量指标，决定相应的多空手数。当成交量指标处于高点，标的指数也处于高点时，考虑相应的下跌风险；当成交量指标处于低点，标的指数下行时，考虑长期熊市的可能。

最后，构造相对成交量

$$rvol = \frac{vol - \min(vol)}{\max(vol) - \min(vol)}, \quad rvol_2 = \frac{\max(vol) - vol}{\max(vol) - \min(vol)}$$

最终的仓位取 $pos = indicator \cdot rvol_2 \cdot (1 + 2 \cdot rvol)$ (绝对值最大不超过 1)

当仓位不足 100%时，对剩余现金做年化收益率 3%的现金管理。

4.2.2. 成交量指标:

下图展示了成交量指标在回测区间内的预测值和真实值对比。可以看出，在 2016 年和 2018 年，预测的成交量指标非常准确；而在 2017 年，预测成交量显著小于真实值。推测原因在于，2017 年的牛市吸引了太多跟风炒作的资金，从而拉高成交量偏离其应有的值。



图 5：策略二的成交量指标

4.2.3. 实际仓位

通过上述的仓位构造方法，我们得到在回测区间上仓位的图像，整体换手率很低，仓位长期比较稳定。

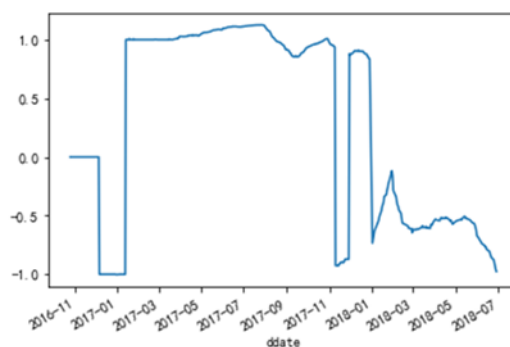
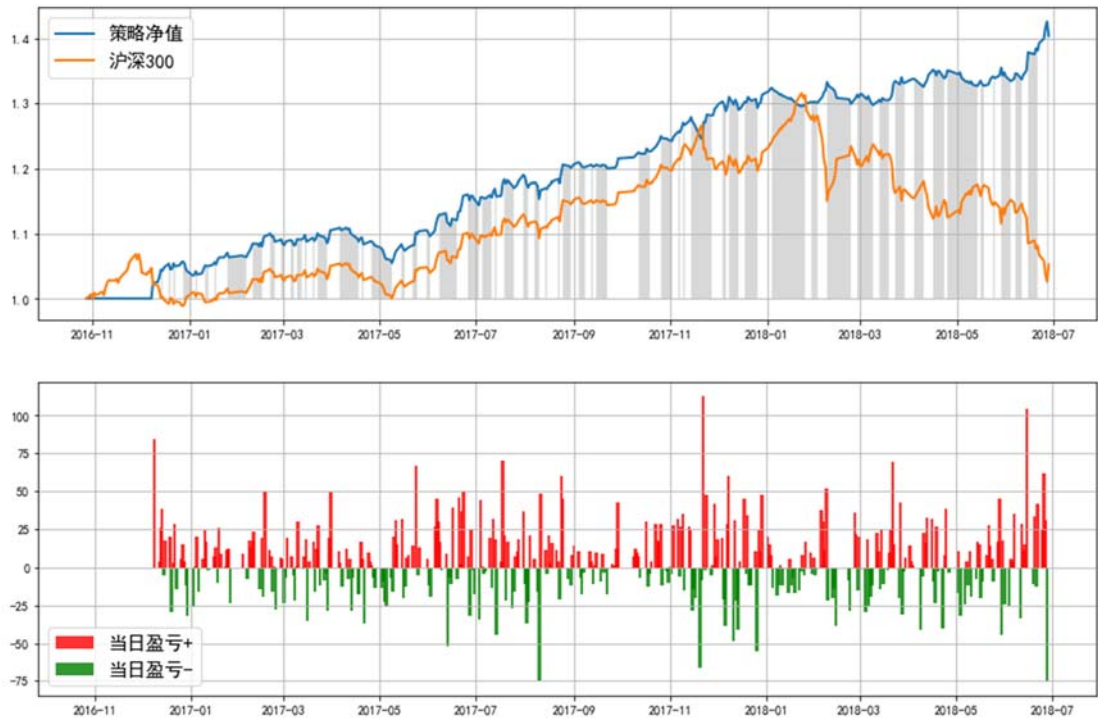


图 6：策略二的实际仓位变化

4.2.4. 策略表现：



注：上图纵轴为策略净值（初始值为 1.00），下图纵轴为当日盈亏（无乘数）。计算方式均为单利。

图 7：策略二净值与当日盈亏变化图

	累计收益率	最大回撤	最大回撤发生时间段		收益回撤比	Sharpe Ratio
总计	40.49%	-4.90%	2017-04-06	2017-05-10	8.27	0.19
2016	4.36%	-1.27%	2016-12-19	2016-12-21	3.42	0.31
2017	27.04%	-4.90%	2017-04-06	2017-05-10	5.52	0.25
2018	9.09%	-2.65%	2018-02-08	2018-03-12	3.43	0.14

表 4：策略二的收益指标

策略二引入了成交量指标，在对市场大趋势有了总体把握的前提下，进行了一定的仓位管理。相较于策略一，策略二的收益率略有下降，但最大回撤显著减小，夏普比率也有相应提高，收益率曲线明显更加平稳。

4.3. 策略小结

本策略本质上属于宏观趋势策略，仅根据模型指标的指示按周调仓，因而策略净值会一定程度上跟随市场波动。但是在 16-18 年的回测上可以看出，我们构建的收益率指标对市场趋势的判读具有很好的准确性和预见性。例如在 18 年初，收益率指标已经综合考虑 17 年的各类宏微观指标。预见到未来可能的熊市，提前开始开空。

5. 总结与展望

我们从宏观经济因子出发，筛选了 100 个反映宏观经济形势的各类因子（具体见附件），并且在 100 个因子的基础上构建了 13 个宏观经济分类指标。这些经济分类指标反映长期或短期以来市场经济形势的总体发展态势，涵盖债券市场、货币市场、汇率市场和衍生品市场等国内主要或者重要金融市场，也包含工业、农业、采矿业等实体经济市场，还收纳了包括人口、自然灾害在内的一些社会科学指标。13 个宏观经济分类指标系统，有重点地描述了国内经济发展整体态势，有助于模型全面准确地学习和描述金融市场未来的趋势。

在现有因子的基础上，可改进之处包括：数据的质量，包括如何提高一些获取的宏观数据的频率等等；数据的内容上，亦可以考虑增加全球宏观经济数据，以反映诸如中美“贸易战”之类会对市场产生重大利空或利好的事件。

当然，使用宏观数据必须考虑的缺陷在于，现实中宏观经济数据的公布和获取不可避免的存在滞后现象。这种延迟通常在 1 个月左右，虽然可能因此错过及时调整仓位的机会，但我们认为这种程度的滞后不足以导致模型预测出现很大的偏差。

对于多维数据信息的学习，我们目前采用的方法主要步骤是先进行多元线性拟合，再根据每一个维度数据的具体含义，手动将它们分类为 13 个指标（也可看作数据降维），再按照每一个因子对于最终市场收益率的权重将其加权求和得到最终用于判断的市场走势和交易策略的指标的值。然后再利用构造的指标数据通过决策树模型来预测最终的交易信号。

根据决策树模型，我们选取前 80% 时间段数据集作为训练集，后 20% 作为回测集，进行策略构建与回测。我们首先仅以收益率指标建仓构建了策略一，然后综合考虑成交量指标构建了策略二。结果显示，相比策略一，策略二收益率略有下降，但最大回撤比显著下降，夏普比率也有相应提高。根据策略表现，我们构建的收益率指标可以较好得预测市场变动情况，为交易提供信号。

附录一：

表 5：因子分类构建

经济指数	序号	经济因子	公布频率	数据来源	实际数据	处理方式	备注
先行指标	27	居民可支配收入	季度	统计局			
	26	地价环比增长率	季度	国泰安			
	25	地价	季度	国泰安			
	29	能源消费总量	年	国泰安		取对数	
	30	能源煤炭占比	年	国泰安			
	32	货币供应量同比增长率	月	国泰安			
	33	货币供应量	月	国泰安		取对数	
	31	财政预算收入	月	国泰安			
	23	储蓄余额	月	国泰安		取对数	
	28	用电量月度数据	月	国泰安			
	24	固定资产投资	月	国泰安		取对数	
传统指标	2	GDP	季度	统计局		取对数	
	1	GDP 增长率	季度	统计局			同比
	4	城镇失业人口	年	国泰安			
	5	城镇失业率	年	国泰安			
	6	居民消费价格指数	年	国泰安			
	7	居民消费价格涨幅	月	国泰安			环比
	8	居民消费水平指数	年	国泰安			
	10	工业生产者购进价格指数	月	统计局			
	18	货物运输量	月	统计局		取对数	
	11	旅客运输量	月	统计局		取对数	
	3	发电量月度数据	月	统计局			
	9	工业生产者出厂价格指数	月	统计局			
	12	第一产业贡献率	季度	统计局			
	13	第一产业贡献率-累计值					
	16	第二产业贡献率	季度	统计局			
	17	第二产业贡献率-累计值					
	14	第三产业贡献率	季度	统计局			
	15	第三产业贡献率-累计值					
商品价格 指标	36	油价	日	英为财经	WTI 原油现货 价格（美		美元计价

					元)		
	35	房价	年	国泰安	商品房售价		
	34	农产品价格指数	季	国泰安	农产品总价格指数		
	37	煤炭价格	月	国泰安	煤炭开采洗选业出厂价格指数		
	22	股指期货基差	日	国泰安			
汇率指标	69	人民币对美元汇率	日	国泰安			间接标价
	72	美元汇率指数	月	国泰安			
	71	欧元汇率指数	月	国泰安			
	70	日元汇率指数	月	国泰安			
债市指标	19	平均久期	日	国泰安			移动平均
	20	平均凸性	日	国泰安			
	21	固定利益国债收益率	日	国泰安			
	61	全价指数	日	Wind	中债总全价指数		
	62	净价指数	日	Wind	中债总净价指数		
货币资金指标	103	金融交易资金占总资金比例	年	统计局			总资金采用实物和金融加和
	102	实物交易资金占总资金比例	年	统计局			
创新指标	58	余额宝情绪指数	日	Wind	余额宝情绪指数（2014年3月31日）		只有14年以后数据
政策指标	63	财政收支差额	月	Wind	财政收支差额当月值	归一化	
	64	个人所得税	月	Wind	个人所得税税收收入当月值	归一化	
	65	国内增值税	月	Wind	国内增值税税收收入当月值	归一化	
	49	社保（社保基金收益）	年	Wind	全国社保基	归一化	

					金基金收益		
贸易指标	51	贸易余额（进出口）	月	Wind	进口数量指数(HS2):总指数		
	52		月	Wind	进口价格指数(HS2):总指数		
	53		月	Wind	出口数量指数(HS2):总指数		
	54		月	Wind	出口价格指数(HS2):总指数		
	50	外商直接投资			实际使用外资金额		
股票市场指标	57	超额收益率		Wind	Wind 全 A 超额收益率		
	59	股东回购比率	月	Wind	上证所股票回购金额累计值		初始值较小不能归一化
	56	市净率	日	Wind	市净率(Wind 全 A)		
	55	滚动市盈率	日	Wind	滚动市盈率(Wind 全 A)		
	60	IPO（中国公司在国内外）	月	Wind	境内外当月筹资金额	归一化	
	48	期权期货成交量	月	Wind	全国当月期货成交量	归一化	
自然社会指标	97	人口自然增长率	年	统计局			
	100	经济活动人口占总人口比例	年	统计局			
	99	森林病虫害灾害	年	统计局	森林病虫害灾害发生面积		
	98	农作物产品产量	年	统计局		取对数	
	101	自然灾害损失	年	统计局			
宏观经济	44	房地产开发景气指数	月	国泰安			数据库下载

指数	43	宏观经济预警指数	月	国泰安			数据库下载
	41	宏观经济先行指数	月	国泰安			数据库下载
	40	宏观经济一致指数	月	国泰安			数据库下载
	42	宏观经济滞后指数	月	国泰安			数据库下载
	46	消费者满意指数	月	国泰安			数据库下载
	45	消费者信心指数	月	国泰安			数据库下载
	47	消费者预期指数	月	国泰安			数据库下载
	38	企业家信心指数	季度	国泰安			数据库下载
	39	企业景气指数	季度	国泰安			数据库下载
HS300 技术指标	73	HS300_vol	日				成交量
	74	HS300_MA_10	日				10 日均线
	75	HS300_MACD1	日				MACD(fast)
	76	HS300_MACD2	日				MACD(slow)
	77	HS300_ADX	日				Average Directional Movement Index
	78	HS300_ADXR	日				Average Directional Movement Index
	79	HS300_aroondown	日				Aroon (down)
	80	HS300_aronup	日				Aroon (up)
	81	HS300_ATR	日				Average True Range
	82	HS300_Bupper	日				Bollinger Bands (upper)
	83	HS300_Bmiddle	日				Bollinger Bands (middle)
	84	HS300_Blower	日				Bollinger Bands (lower)
	85	HS300_BOP	日				Balance Of Power
	86	HS300_CCI	日				Commodity Channel Index
	87	HS300_CMO	日				Chande Momentum Oscillator
	88	HS300_DX	日				Directional Movement Index
	89	HS300_EMA	日				Exponential Moving Average
	90	HS300_KAMA	日				Kaufman Adaptive Moving Average
	91	HS300_MFI	日				Money Flow Index
	92	HS300_MOM	日				Momentum
	93	HS300_RSI	日				Relative Strength Index
	94	HS300_SAR	日				Parabolic SAR
	95	HS300_TRANGE	日				True Range
	96	HS300_TRIMA	日				Triangular Moving Average

附录二：

表 6：因子有效性检验

序号	因子权重	标准差	t	P> t	序号	因子权重	标准差	t	P> t
1	-1.08	1.36	-0.79	0.42	51	0.07	0.34	0.21	0.84
2	-15.37	3.26	-4.72	0.00	52	1.26	1.03	1.22	0.22
3	2.74	1.20	2.28	0.02	53	0.95	0.39	2.46	0.01
4	43.32	7.66	5.66	0.00	54	2.55	0.84	3.05	0.00
5	-19.90	4.66	-4.27	0.00	55	-8.17	2.08	-3.92	0.00
6	1.78	2.95	0.60	0.55	56	2.28	2.92	0.78	0.44
7	2.12	1.11	1.92	0.06	57	-9.60	2.39	-4.03	0.00
8	18.42	5.72	3.22	0.00	58	1.10	0.39	2.81	0.01
9	-1.53	6.03	-0.25	0.80	59	-7.65	1.05	-7.31	0.00
10	-6.61	7.63	-0.87	0.39	60	-0.72	0.61	-1.17	0.24
11	-7.03	2.91	-2.41	0.02	61	4.80	10.77	0.45	0.66
12	55.29	23.25	2.38	0.02	62	-15.44	9.65	-1.60	0.11
13	-35.97	9.90	-3.64	0.00	63	2.33	0.57	4.09	0.00
14	264.00	87.00	3.03	0.00	64	-2.87	1.25	-2.29	0.02
15	-369.64	68.75	-5.38	0.00	65	-3.60	1.29	-2.79	0.01
16	221.13	81.63	2.71	0.01	69	-12.55	1.39	-9.02	0.00
17	-321.70	64.85	-4.96	0.00	70	2.33	1.82	1.28	0.20
18	4.63	0.97	4.78	0.00	71	3.25	0.97	3.37	0.00
19	-14.19	2.91	-4.87	0.00	72	11.08	3.03	3.66	0.00
20	9.46	5.16	1.83	0.07	73	-4.83	10.10	-0.48	0.63
21	-9.30	1.84	-5.06	0.00	74	55.34	8.40	6.59	0.00
22	-2.85	0.95	-3.00	0.00	75	14.47	1.97	7.36	0.00
23	1.84	10.99	0.17	0.87	76	-17.73	1.91	-9.29	0.00
24	-2.49	1.34	-1.86	0.06	77	-0.99	0.49	-2.00	0.05
25	-47.98	10.15	-4.73	0.00	78	1.25	0.45	2.78	0.01
26	0.42	0.65	0.65	0.52	79	-0.10	0.28	-0.37	0.71
27	-11.73	1.49	-7.85	0.00	80	-0.35	0.29	-1.20	0.23
28	-3.32	0.93	-3.58	0.00	81	-2.32	1.01	-2.30	0.02
29	-40.02	7.87	-5.08	0.00	82	-11.18	12.43	-0.90	0.37
30	12.26	13.60	0.90	0.37	83	23.03	23.19	0.99	0.32
31	-1.77	1.01	-1.75	0.08	84	-18.95	11.48	-1.65	0.10
32	0.21	0.57	0.37	0.72	85	-0.75	0.26	-2.87	0.00
33	36.35	6.22	5.84	0.00	86	-0.50	0.42	-1.18	0.24
34	6.87	1.03	6.70	0.00	87	-0.92	0.31	-3.01	0.00
35	40.25	7.35	5.47	0.00	88	0.88	0.27	3.27	0.00
36	-1.32	0.91	-1.46	0.14	89	-41.74	10.23	-4.08	0.00
37	5.32	2.84	1.88	0.06	90	-15.59	3.92	-3.98	0.00
38	2.87	1.36	2.11	0.04	91	-0.64	0.30	-2.17	0.03

39	4.96	1.42	3.49	0.00	92	-1.64	0.61	-2.66	0.01
40	22.43	4.57	4.91	0.00	93	-0.92	0.31	-3.01	0.00
41	13.37	1.93	6.94	0.00	94	-2.85	1.52	-1.88	0.06
42	-12.65	3.74	-3.38	0.00	95	1.71	0.48	3.59	0.00
43	-6.28	2.24	-2.80	0.01	96	24.00	4.73	5.07	0.00
44	-11.71	1.39	-8.42	0.00	97	-12.34	5.34	-2.31	0.02
45	110.67	48.05	2.30	0.02	98	0.92	8.41	0.11	0.91
46	-42.57	21.10	-2.02	0.04	99	-3.50	7.46	-0.47	0.64
47	-67.85	29.62	-2.29	0.02	100	-12.26	3.10	-3.96	0.00
48	-1.18	0.51	-2.32	0.02	101	7.46	5.99	1.25	0.21
49	-13.94	4.57	-3.05	0.00	102	-4.88	3.92	-1.25	0.21
50	-4.08	1.58	-2.58	0.01	103	4.88	3.92	1.25	0.21