

量化专题报告

宏观逻辑的量化验证：动态因子模型

本报告尝试解决宏观逻辑的结合与配权问题。宏观因素影响资产价格的逻辑与数据链条有很多，各链条影响程度不同。传统的配权方法如：1) 简单结合法；2) 时间加权最小均方误差法；3) 收缩估计法；4) 时变参数权重法，无法解决链条相关性导致的共线性问题，且从逻辑角度也较难解释和理解。

动态因子模型为宏观逻辑配权提供了解决方案。动态因子模型假设宏观变量是一个由“公共因子”和“特质扰动项”组成的随机变量，通过对“公共因子”的估计能将大量宏观变量进行降维，从而解决宏观数据大 N 小 t 导致的参数无法估计的问题，以及宏观变量共线性的问题。其通常用于领先指标研究、经济周期研究、经济指标预测等方面。

本报告详细阐述了动态因子模型的构建。动态因子模型从模型的形式角度分为动态与静态两类，从模型的假设角度分为严格与近似两类。从实际的应用角度，近似假设下的静态形式比较适用于大 N 小 t 问题的建模，在此形式下“公共因子”的估计可以通过主成分分析进行。

通过动态因子模型的降维，大量宏观变量可用于预测资产收益率。我们构建了从基础数据处理到最终收益率预测的完整框架。以 400 个宏观变量构建了包含 787 个宏观因子的数据库，并展示了包括大类资产、风格因子、行业因子等部分预测模型的样本内外预测效果。

关于噪音与信息损失的思考。为了更好的理解“公共因子”的逻辑，我们采用 Occam 剃刀的原则精简“公共因子”中的宏观因子，能够排除一些噪音干扰。动态因子模型以部分信息的扭曲和损失为代价换来了模型精炼和逻辑的强化，因此我们建议增加主成分遍历和特质波动检验环节弥补信息损失。

风险提示：模型建立基于历史数据，在政策、经济结构等因素变化的情况下模型可能失效。模型建立从数据出发，数量关系未必对应因果关系。

作者

分析师 叶尔乐

执业证书编号：S0680518100003

邮箱：yeerle@gszq.com

分析师 刘富兵

执业证书编号：S0680518030007

邮箱：liufubing@gszq.com

相关研究

- 1、《量化周报：市场的中期上涨趋势更加稳健》
2019-04-07
- 2、《量化分析报告：量化方法测算 A 股壳价值》
2019-04-02
- 3、《量化周报：50、300 确认周线级别上涨》2019-03-31
- 4、《量化专题报告：多因子系列之四-对价值因子的思考和改进》2019-03-25
- 5、《量化周报：市场本周处于牛熊分界线》2019-03-24

内容目录

1. 宏观经济的动态因子模型.....	4
1.1. 宏观逻辑的结合与配权问题.....	4
1.2. 动态因子模型简介.....	4
1.3. 动态因子模型构建.....	6
1.3.1 模型的形式：动态与静态.....	6
1.3.2 模型的估计：严格与近似.....	6
2. 基于动态因子模型的宏观-资产关系建模.....	7
2.1. 资产收益率预测模型结构.....	7
2.2. 模型的定阶.....	7
2.3. 数据的选择.....	8
2.4. 数据的处理.....	8
2.5. 模型构建步骤.....	9
3. 部分显著模型分析.....	10
3.1. 大类资产：沪深 300.....	10
3.2. 大类资产：黄金.....	12
3.3. 大类资产：工业品.....	13
3.4. 风格因子：波动率.....	15
3.5. 风格因子：杠杆率.....	16
3.6. 行业因子：建筑.....	18
3.7. 行业因子：家电.....	20
3.8. 行业因子：食品饮料.....	21
4. 关于噪音与信息损失的思考.....	23
4.1. 宏观因子的 Occam 剃刀.....	23
4.2. 主成分选取的顺序.....	23
4.3. 降维造成的信息损失.....	24
风险提示.....	24

图表目录

图表 1: 不同宏观预测模型对比.....	5
图表 2: 宏观经济指标库.....	8
图表 3: 宏观指标处理模块.....	9
图表 4: 动态因子预测模型框架.....	10
图表 5: DFM 预测模型信号.....	10
图表 6: DFM 预测模型样本内择时表现.....	11
图表 7: DFM 预测模型样本外择时表现.....	11
图表 8: 宏观变量最终权重.....	11
图表 9: DFM 预测模型信号.....	12
图表 10: DFM 预测模型样本内择时表现.....	12
图表 11: DFM 预测模型样本外择时表现.....	12
图表 12: 宏观变量最终权重.....	13
图表 13: DFM 预测模型信号.....	14
图表 14: DFM 预测模型样本内择时表现.....	14

图表 15: DFM 预测模型样本外择时表现.....	14
图表 16: “公共因子”组成.....	15
图表 17: DFM 预测模型信号	15
图表 18: DFM 预测模型样本内择时表现.....	16
图表 19: DFM 预测模型样本外择时表现.....	16
图表 20: “公共因子”组成.....	16
图表 21: DFM 预测模型信号	17
图表 22: DFM 预测模型样本内择时表现.....	17
图表 23: DFM 预测模型样本外择时表现.....	17
图表 24: 宏观变量最终权重.....	18
图表 25: 能源与黑色、有色金属工业增加值.....	18
图表 26: DFM 预测模型信号	19
图表 27: DFM 预测模型样本内择时表现.....	19
图表 28: DFM 预测模型样本外择时表现.....	19
图表 29: “公共因子”组成.....	20
图表 30: DFM 预测模型信号	20
图表 31: DFM 预测模型样本内择时表现.....	21
图表 32: DFM 预测模型样本外择时表现.....	21
图表 33: “公共因子”组成.....	21
图表 34: DFM 预测模型信号	22
图表 35: DFM 预测模型样本内择时表现.....	22
图表 36: DFM 预测模型样本外择时表现.....	22
图表 37: “公共因子”组成.....	23
图表 38: 动态因子模型建模过程中的噪音、信息损失、信息扭曲问题对策	24

1. 宏观经济的动态因子模型

1.1. 宏观逻辑的结合与配权问题

在上一篇报告中，我们利用状态匹配法初探了宏观经济指标和资产价格变化的映射关系。这类方法都存在一个问题：单一宏观经济指标的状态划分信息量有限，使得资产价格判断的波动性较大，而多个宏观经济指标的状态划分叠加虽然信息量增加了但是划分完后的样本数据过少使得规律的显著性大打折扣。因此这类模型亟需解决一个重要的问题：**在数据量较小的情况下，多个单一模型/指标预测结果的结合问题。**

多个模型预测结果的结合（Forecasting Combination）方式有很多种，在连续的预测模型中常用的方法有如下几种：

1) 简单结合法

简单结合法即取多个预测结果的均值、截尾均值或者中位数作为最终预测值；

$$w_{it} = 1/n$$

2) 时间加权最小均方误差法

此方法在计算最终预测值的时候使用每个预测模型历史时间加权的均方误差（Means squared forecast error） m_{it} 的倒数，其中 ρ 为折现因子。

$$w_{it} = \frac{m_{it}^{-1}}{\sum_{j=1}^n m_{jt}^{-1}}, m_{it} = \sum_{s=T_0}^{t-h} \rho^{t-h-s} (Y_{s+h}^h - \hat{Y}_{is+h|s}^h)^2$$

3) 收缩估计法

此方法通常将权重往某个先验的权重进行收缩估计，一般这个先验权重取等权重。

$$w_{it} = \lambda \hat{w}_{it} + (1 - \lambda)(1/n)$$

4) 时变参数权重法

此方法允许权重作为一个随机过程，比如允许其遵循随机游走的变化方式，时变参数权重法的估计可以通过 Kalman 滤波进行。

$$w_{it} = w_{it-1} + \eta_{it}$$

从各种研究结果来看，预测合成的结果往往比单个预测的结果要好，同时简单合成方法一般和复杂合成效果差不多甚至更好。但这样的合成方式的逻辑解释都是基于某种数据的处理方式，很难找到实际的经济含义，同时对于各个预测可能的共线性问题（逻辑集中），数据量不足等问题也没有很好的解决办法，相对来说较为粗糙。本报告将探讨一种由 Geweke（1977）提出，由 Stock 与 Watson 等人发展起来的多变量宏观预测方法，这种方法适用于利用大量经济数据（往往数据个数 N 相对数据量 t 较大）进行预测。

1.2. 动态因子模型简介

相比于过去，当前的信息处理与储存能力大幅提升，使得宏观数据能够以一个合理的成本大批量的产生，从而诞生了一个全新的宏观研究领域，即利用大量的时间序列来预测某几个关键经济变量。传统的模型只能处理少量的自变量，比如向量自回归（VAR）模型最多也仅能处理 10 个以下的自变量。这就涉及到变量选择的问题，而变量的选择势必会损失一部分数据的信息。

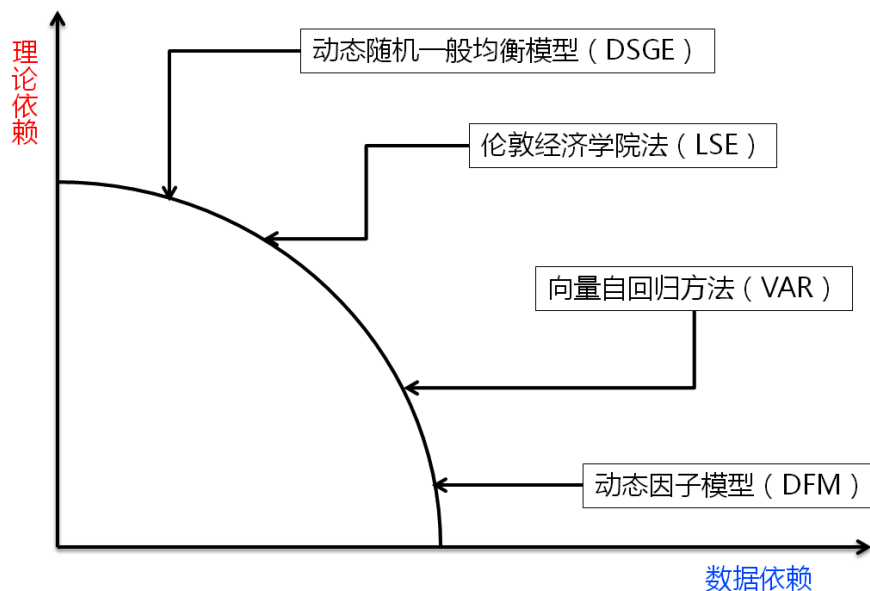
另一种思路则将所有有用信息都导入模型，通过组合去噪得到数据集的主要变化因素再对因变量进行相应的预测。宏观经济学中的现代动态一般均衡模型假设存在少量不可观测变量能够解释宏观变量的大部分波动，而且这些不可观测变量可看作宏观变量的一种“公共因子”。

这种思想的代表就是**动态因子模型（DFM, Dynamic Factor Model）**，在利用 DFM 进行预测时分为两步，首先用大量宏观变量估计出“公共因子”，然后用这些“公共因子”去预测。这样的“公共因子”存在一定解释上的难题，然而仅聚焦于其预测能力避免了这个问题。对这种“公共因子”的一种解释是美国国家经济研究局（NBER）在做商业周期分析时提出的，称为**扩散指数（Diffusion Indexes）**。

利用 DFM 进行预测的好处在于，在数据个数 N 相对数据量 t 较大的情况下，回归 OLS 可能无解，或者误差非常大（估计误差与 N/t 成正比），同时在不进行共线性处理的情况下回归得到的系数误差较大。而 DFM 通过压缩自变量，使得大量的自变量转化为互相正交的少数几个“公共因子”，既解决了数据量的问题，同时也解决了共线性的问题。参考学术研究的成果，DFM 相比于单变量自回归模型、较少变量向量自回归模型、领先指标模型等都表现更好。

利用 DFM 进行预测的不足在于，解释度最高的几个“公共因子”尚且有一定显著的经济逻辑，而其余因子未必能找到，且模型的形式决定了“公共因子”不适宜解释为某种宏观周期的波动。相比较而言动态随机一般均衡模型（DSGE）更具理论依赖性，其高度依赖经济学理论，但是模型的构建过程较为繁琐并且维护成本高，二来其假设过强使得对先验逻辑的要求极高，相比之下我们认为数据依赖的 DFM 更适合用来搭建宏观量化系统。

图表 1：不同宏观预测模型对比



资料来源：国盛证券研究所

1.3. 动态因子模型构建

1.3.1 模型的形式：动态与静态

动态因子模型的经典形式如下：

$$\begin{aligned} X_{it} &= \lambda_i(L)f_t + e_{it}, i = 1, \dots, n \\ \Gamma(L)f_t &= \eta_t \end{aligned}$$

X_{it} 代表宏观变量， f_t 为 $q \times 1$ 的向量，代表 q 个公共动态因子， $e_t = (e_{1t}, \dots, e_{nt})'$ 代表特质扰动项， η_t 为噪声项， $\lambda_i(L)$ 、 $\Gamma(L)$ 为有限阶数滞后多项式。模型假设宏观变量 X_{it} 由不可观测的“公共因子” f_t 以及“公共因子”所不能解释的特质扰动项 e_{it} 构成，并且两者在所有滞后情况下都不相关： $E(f_t e_{is}) = 0, \forall i, s$ 。 $\lambda_i(L)$ 可以称为“动态因子载荷”。

模型的动态性是指“公共因子”本身以动态过程的结构进行建模。如果我们假设 f_t 不具有自相关的结构： $E(f_t f_s) = 0, \forall t \neq s$ ，那么动态因子模型即变为**静态因子模型 (SFM, Static Factor Model)**。

1.3.2 模型的估计：严格与近似

模型经典的假设较为严格，需要保证不同宏观变量的特质扰动项 e_{it} 之间互不相关。

严格 DFM 假设：

$$E(e_{it}e_{jt}) = 0 \text{ for } i \neq j$$

在此基础之上我们可以使用带参的估计方法，例如极大似然估计 (MLE)，其中的似然函数可以使用 Kalman Filter 计算得到，而动态因子也可以由 Kalman Smoother 估计得到。由于随着 n 的增长待估参数和计算量都显著增长，极大似然估计一定程度上受到数据量的限制。同时，我们无法保证提取的“公共因子”组能够解释宏观变量所有的公共驱动因素(把公共部分剔干净)，从而也就无法保证不同宏观变量的特质扰动项的不相关假设。因此在实际应用中我们偏向于使用近似的模型假设与其他估计方式。

近似 DFM 假设：

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sum_{j=1}^n |E(e_{it}e_{jt})| < \infty$$

近似的模型假设允许特质扰动项之间存在一定的相关性，只需这个相关性随着 n 的增大对 X_t 的总体方差贡献有限就行。并且近似的 DFM 在 n 较大的情形下可以使用主成分分析 (PCA) 的方法进行估计，在 n 趋于无穷大的时候，PCA 被证明是和传统因子分析和 MLE 等价的 (Chamberlain and Rothschild, 1983)。

具体的来说我们可以将 DFM 改写为 SFM 的形式：

$$X_t = \Lambda F_t + E_t$$

其中 X_t 为 X_{it} 的向量形式，而 $F_t = [f_t' f_{t-1}' \dots f_{t-p+1}']$ ，其维数为 $1 \times r$ ($r \leq pq$)，本质是将 $\lambda_i(L)f_t$ 项进行展开， Λ 为相应系数矩阵， E_t 为特质扰动项的向量形式。由于 F_t 与 E_t 在任意滞后阶数上都互不相关，因此 X_t 的协方差矩阵可以写为：

$$\Sigma_{XX} = \Lambda \Sigma_{FF} \Lambda' + \Sigma_{EE}$$

当 n 趋于无穷大的时候, 假设 Σ_{EE} 的特征值是 $O(1)$ 的, $\Lambda\Lambda'$ 的特征值是 $O(n)$ 的, 那么 Σ_{XX} 的前 r 个特征值是 $O(n)$ 的, 而其余的特征值是 $O(1)$ 的, 这意味着 X_t 的前 r 个主成分可以作 Λ 的估计。具体证明在本报告中不再展开。

从计算的角度看, PCA 方法的一大好处就在于数据量不足的情况下, 其对于很大的 N 也能轻松的计算, 而且其可以较为便利的处理缺失或者不同频的数据集。

2. 基于动态因子模型的宏观-资产关系建模

2.1. 资产收益率预测模型结构

资产收益率不仅受到外部宏观因素的影响, 部分资产也有较强的自相关性, 因此我们在资产预测模型构建中除了动态因子, 也将资产收益率的滞后项也加入进来:

$$\begin{aligned} y_{t+1} &= \alpha + \beta(L)f_t + \gamma(L)y_t + \epsilon_{t+1} \\ X_{it} &= \lambda_i(L)f_t + e_{it} \end{aligned}$$

其中 y_{t+1} 代表我们需要预测的资产收益率, X_{it} 代表宏观变量, f_t 代表公共动态因子, $\lambda_i(L)$ 、 $\beta(L)$ 和 $\gamma(L)$ 为有限阶数滞后多项式, α 为截距项。我们采用近似 DFM 的假设构建模型, 从而模型的整体估计可以分为两步:

- 一、 利用 PCA 将 X_t 降维到少量“公共因子” f_t ;
- 二、 将 y_{t+1} 相对 f_t 、 y_t 及其滞后项以及一个截距项 α 进行回归, 最终得到模型的参数估计: $\hat{\alpha}$ 、 $\hat{\beta}(L)$ 、 $\hat{\gamma}(L)$ 。

2.2. 模型的定阶

模型的参数集合一共由三个参数组成: 公共因子个数 k , 公共因子滞后阶数 m , 因变量滞后阶数 p 。模型由复杂到简单有三种形式:

- 1、 **DI-AR, Lag 模型**
 $1 \leq k \leq 4, 1 \leq m \leq 3, 0 \leq p \leq 3$, 具体参数由 BIC 准则来定;
- 2、 **DI-AR 模型**
 $1 \leq k \leq 4, m = 1, 0 \leq p \leq 3$, k 、 p 的选择由 BIC 准则来定;
- 3、 **DI 模型**
 $1 \leq k \leq 4, m = 1, p = 0$, k 的选择由 BIC 准则来定。

传统的 BIC 准则目的是达到似然函数最大化和模型参数个数最小化的平衡, 使得模型有较好的拟合效果, 同时减少过拟合概率。但是在**资产收益率方向预测**中, 模型的优化目标(最大化似然函数)和我们的预测目标(最大化预测方向的正确率)是不完全相同的, 因此会出现 BIC 最小而预测方向的正确率不是最高的情况。因此我们的模型定阶不完全使用 BIC, 同时考虑样本内预测方向的正确率。

2.3. 数据的选择

由于 DFM 比较擅长处理大 N 相对小 t 的问题，有用的信息量越多越好，本报告整理了 15 大类，共 400 个宏观经济数据，据此构建我们的宏观因子库。

图表 2: 宏观经济指标库

宏观指标类别	宏观指标个数	宏观指标举例
经济增长	26	GDP:不变价:当季同比、GDP:不变价:第一产业:当季同比、GDP:不变价:第二产业:当季同比
工业增加值	52	工业增加值:当月同比、工业增加值:采矿业:当月同比、工业增加值:制造业:当月同比、
工业产品产量	57	产量:原煤:当月同比、产量:焦炭:当月同比、产量:发电量:当月同比
景气调查	68	PMI、PMI:生产、PMI:新订单
价格	50	CPI:当月同比、CPI:食品:当月同比、CPI:非食品:当月同比、
货币	18	M0:同比、M1:同比、M2:同比
固定资产投资	11	固定资产投资完成额:累计同比、房地产开发投资完成额:累计同比、新增固定资产投资完成额:累计同比、
进出口贸易	16	出口金额:当月同比、进出口金额:当月同比、银行结售汇差额:当月值
国内消费	3	社会消费品零售总额:当月同比、社会消费品零售总额:商品零售:当月同比、社会消费品零售总额:餐饮收入:当月同比
融资	20	债券发行量:合计:当月值、债券发行量:政府债券:当月值、债券发行量:地方政府债:当月值
利率	26	活期存款利率、定期存款利率:3 个月、定期存款利率:6 个月
就业	5	城镇登记失业率、城镇领取失业保险金人数、城镇新增就业人数:累计同比
国际收支	9	经常账户:差额:当季值、经常账户:货物和服务:差额:当季值、经常账户:初次收入:差额:当季值、
工业企业经济效益	23	工业企业:主营业务收入:当季同比、工业企业:利润总额:当季同比、工业企业:国有及国有控股企业:利润总额:当季同比
房地产	16	100 大中城市:供应土地数量:当月值、本年购置土地面积:累计同比、土地购置费:累计同比

资料来源: 国盛证券研究所

待预测的资产方面，我们选择了包括大类资产、风格因子组合、行业因子组合在内多种资产进行回测。

2.4. 数据的处理

由于宏观数据公布的频率和方式各不相同，在输入 DFM 之前需要进行一定的调整，调整的原则在于需要将非平稳的序列转化为平稳的序列，并且序列的值有显著的意义，要么是值的绝对值高低有显著意义，比如利率的高低对应融资环境的宽紧，要么是值代表了某种变化趋势，比如用序列差分代表宏观指标变化的方向。针对不同类型和发布频率的数据我们采用了以下 11 种处理模块：

图表 3: 宏观指标处理模块

处理方法	方法简写	适合数据类型
原序列	ori	平稳序列，绝对值高低有明显区分意义的序列
差分序列	diff	差分后平稳序列，变化有明显区分意义的序列
月平均序列	ave	高频且代表某种水平的序列
月求和序列	sum	高频且具有累计影响的序列
4 期移动平均差分序列	ma4+diff	季度数据需要剔除季节效应
HP 滤波差分序列	hp1+diff	趋势相对扰动更重要的序列
3 期移动平均+滤波差分序列	ma3hp1+diff	需要剔除月度效应且趋势相对扰动更重要的序列
4 期移动平均+滤波差分序列	ma4hp1+diff	需要剔除月度效应且趋势相对扰动更重要的序列
6 期移动平均+滤波差分序列	ma6hp1+diff	需要剔除月度效应且趋势相对扰动更重要的序列
11 期移动平均差分序列	ma11+diff	每年只公布 11 个数据的序列，需要剔除月度效应
12 期移动平均差分序列	ma12+diff	需要剔除月度效应的序列

资料来源：国盛证券研究所

同时由于模型的输入需要面板数据，对于低频的数据我们采用**向后填值**的方式将数据填充完整。处理完成后，我们共得到 787 个宏观经济指标时间序列，每个数据点仅由其之前可得到的信息计算得到。

2.5. 模型构建步骤

第一步：宏观因子选择

由于因子的降维只与因子本身有关，因此把与资产相关的和不相关的因子一起提取公共因子后，其在对资产有预测作用的方向的暴露不一定高，这会给模型带来很多噪音，利用动态因子模型进行预测，一定要输入“有用”的宏观信息。因此我们首先需要对于每一类我们需要预测的资产，分别进行“有用”宏观因子的筛选。

我们通过上篇报告《宏观逻辑的量化验证：映射关系混沌初开》提出的 **ANOVA 方差检验法与择时策略检验法**对每类资产各选择出满足显著性条件的关系，然后将选取的宏观序列进行去重等预处理作为模型的输入。

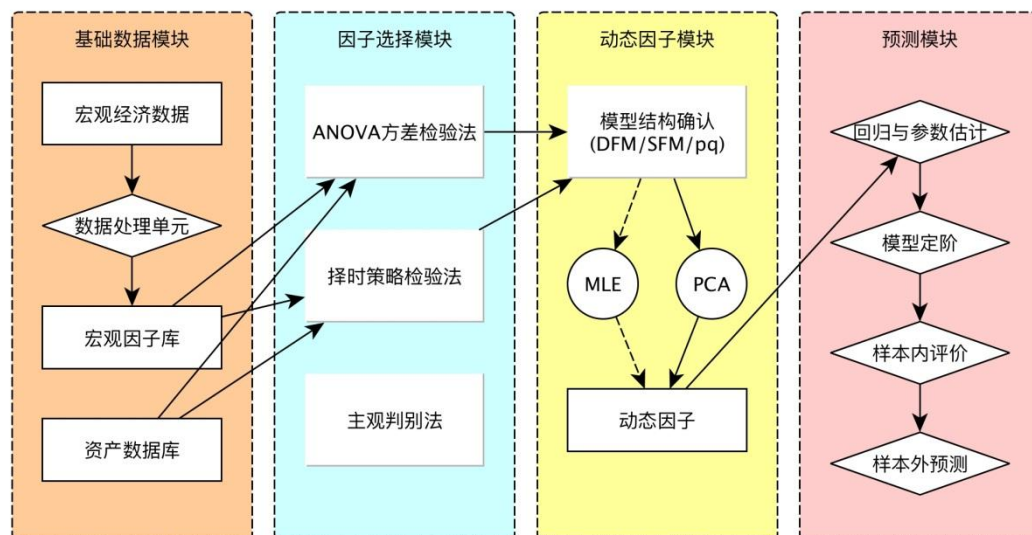
第二步：公共因子的提取

将输入的宏观序列进行降维，利用 PCA 提取对应解释度最大的部分主成分。由于对于量纲不同的指标，直接计算 PCA 得到的权重受到量纲影响，逻辑意义会受到扭曲，因此为了便于理解主成分的逻辑意义，PCA 之前需要对变量进行标准化，使得权重对比的意义更加的纯粹清晰。

第三步：预测模型的构建

我们对公共因子与待预测资产建立带有外生变量的自回归模型。在模型的估计中，我们在一定参数范围内遍历参数组 (k, m, p) 进行模型训练，寻找满足定阶条件的模型。利用所选择的模型对全样本进行样本内预测，获得的预测值转化为对方向的判断（正值对应看多，负值对应看空），利用预测值进行模拟择时，通过择时得到的结果评价模型的优劣。

图表 4: 动态因子预测模型框架



资料来源: 国盛证券研究所

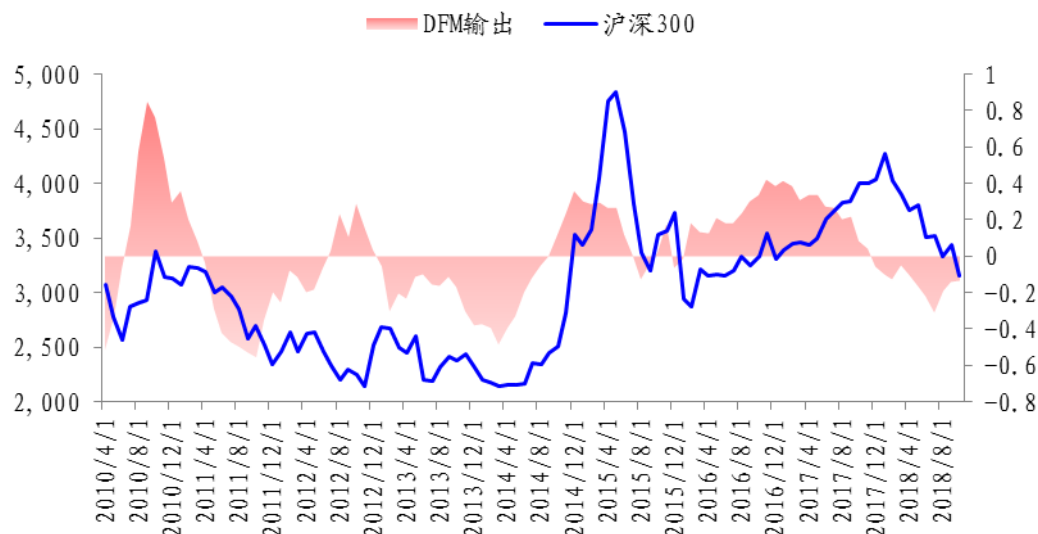
3. 部分显著模型分析

3.1. 大类资产：沪深 300

在权益资产中，我们以沪深 300 作为中国权益市场的代理资产。沪深 300 的投资者成分复杂，收益率决定因素较多，自相关性弱，因而模型定阶的结果 $k=3$, $m=1$, $p=0$ ，也就是说沪深 300 的收益率需要 3 个“公共因子”才能相对较好的进行解释。

$$y_{t+1}^{\text{沪深 300}} = 0.54 - 0.04f_{1,t} + 0.05f_{2,t} - 0.16f_{3,t}$$

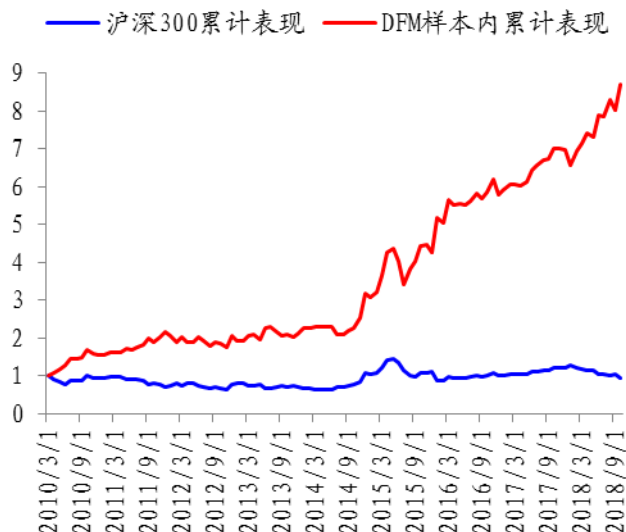
图表 5: DFM 预测模型信号



资料来源: Wind, 国盛证券研究所

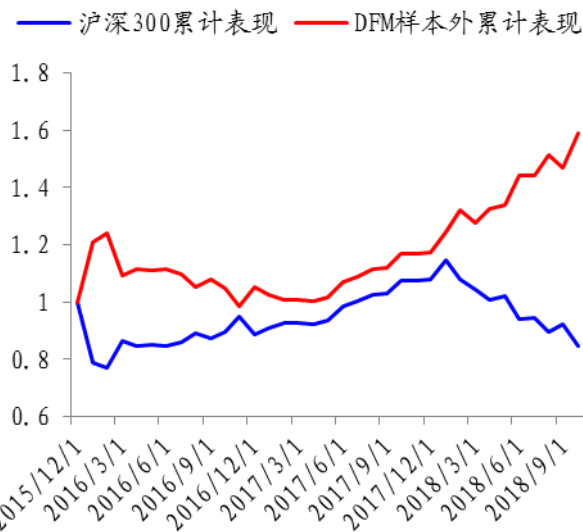
在样本内 DFM 模型胜率 63.11%，同时我们也做了样本外测试：将样本分为 2016 年之前和 2016 年之后两段，通过每月滚动重计算动态因子以及预测模型进行预测，样本外胜率可达 61.76%。

图表 6: DFM 预测模型样本内择时表现



资料来源: Wind, 国盛证券研究所

图表 7: DFM 预测模型样本外择时表现



资料来源: Wind, 国盛证券研究所

将回归系数代入代表“公共因子”的主成分，最终的指标配权如下。可以看到在预测沪深 300 收益率时，**进出口以及外资**有相当正向的影响，**利率**以及上一篇报告中提到的**彩电产量**具有负向影响。除此之外，模型主要涉及**工业企业经济效益**、**价格指数**等宏观变量。

图表 8: 宏观变量最终权重



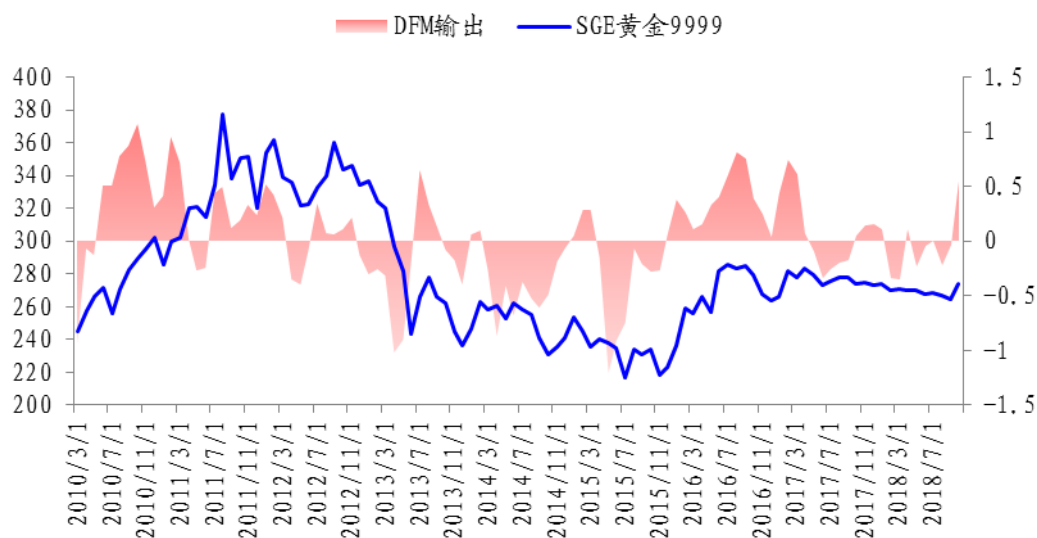
资料来源: Wind, 国盛证券研究所

3.2. 大类资产：黄金

我们以上海黄金交易所的黄金 9999 合约作为黄金的代理资产。国内的黄金合约同样受到诸多复杂因素的影响，从宏观的角度来看黄金主要受到美元、利率、通胀、风险事件等因素的影响，作为国内的黄金合约又受到汇率的影响。模型定阶的结果 $k=5$, $m=1$, $p=1$, 也就是说黄金 9999 合约的收益率需要 5 个“公共因子”以及自身的 1 阶滞后项才能相对较好的进行解释。

$$y_{t+1}^{\text{SGE 黄金 9999}} = 1.35 - 0.02f_{1,t} + 0.01f_{2,t} + 0.003f_{3,t} + 0.21f_{4,t} - 0.38f_{5,t} - 0.21y_t$$

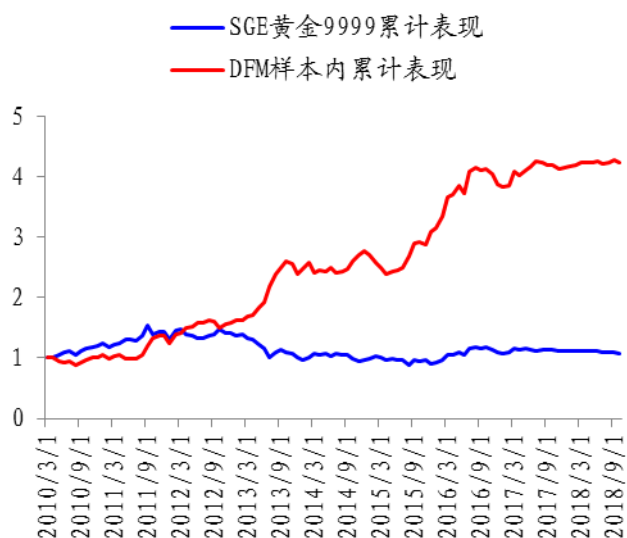
图表 9: DFM 预测模型信号



资料来源: Wind, 国盛证券研究所

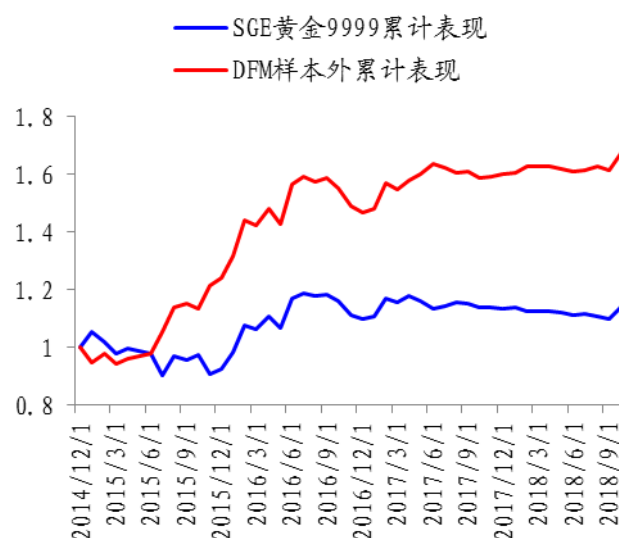
在样本内 DFM 模型胜率 68.93%，同时我们也做了样本外测试：将样本分为 2015 年之前和 2015 年之后两段，通过每月滚动重计算动态因子以及预测模型进行预测，样本外胜率可达 60.87%。

图表 10: DFM 预测模型样本内择时表现



资料来源: Wind, 国盛证券研究所

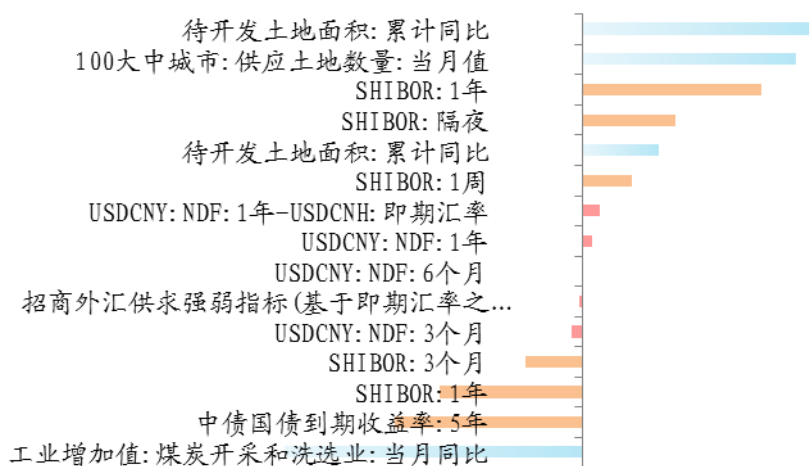
图表 11: DFM 预测模型样本外择时表现



资料来源: Wind, 国盛证券研究所

将回归系数代入代表“公共因子”的主成分，最终的指标配权如下。可以看到黄金价格受到土地供应、短期资金成本、长端利率等的影响比较明显，背后可能蕴含的逻辑为：土地供应增加预示着一定的经济偏弱或是财政紧张问题，黄金避险作用提升；短期资金成本较高，市场流动性风险高，黄金避险作用提升；长端利率下降预示经济预期悲观，黄金避险作用提升。我们的数据目前仅包含国内宏观因素，黄金的海外宏观因素我们将在未来纳入到模型中。

图表 12: 宏观变量最终权重



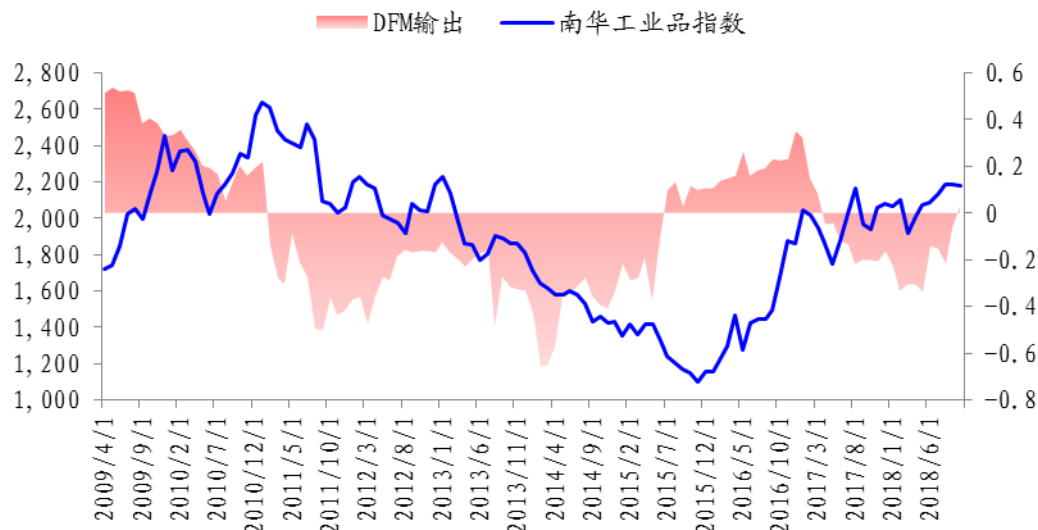
资料来源: Wind, 国盛证券研究所

3.3. 大类资产：工业品

我们以南华工业品指数作为工业品的代理资产。模型定阶的结果 $k=1, m=2, p=0$ ，也就是说对于工业品来说只需要一个因子即可解释大部分波动，只不过因子的影响滞后性较强。南华工业品指数与“公共因子”的 1 阶滞后略微正相关，但是与 2 阶滞后有较强的负相关关系。

$$y_{t+1}^{\text{南华工业品}} = 1.16 + 0.02f_{1,t} - 0.15f_{1,t-1}$$

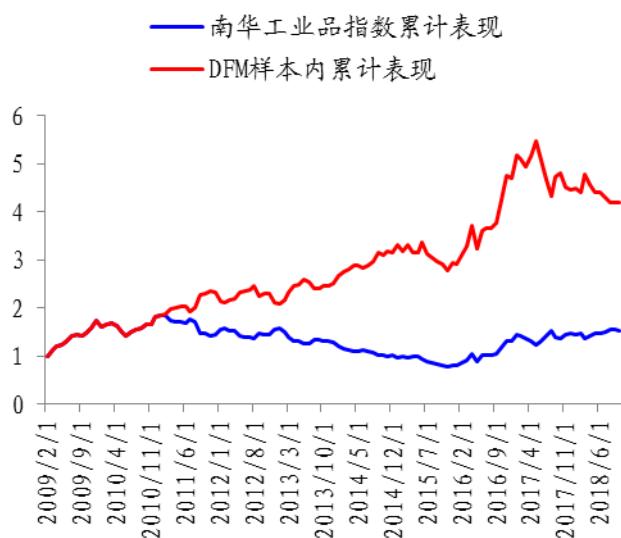
图表 13: DFM 预测模型信号



资料来源: Wind, 国盛证券研究所

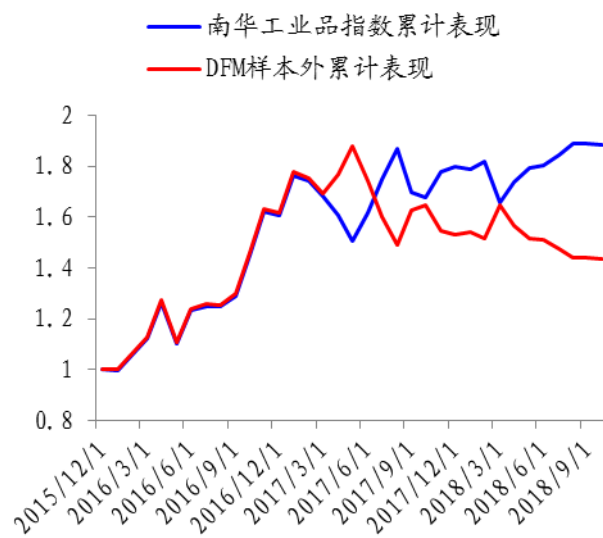
在样本内 DFM 模型胜率 60.87%，同时我们也做了样本外测试：将样本分为 2016 年之前和 2016 年之后两段，通过每月滚动重计算动态因子以及预测模型进行预测，样本外胜率为 47.06%。

图表 14: DFM 预测模型样本内择时表现



资料来源: Wind, 国盛证券研究所

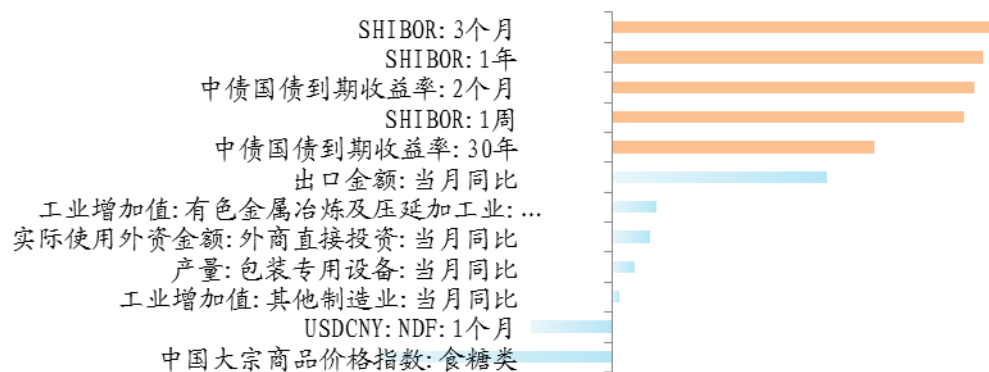
图表 15: DFM 预测模型样本外择时表现



资料来源: Wind, 国盛证券研究所

“公共因子”当中利率占据绝对的权重，利率较低阶段工业品上涨概率更高，利率较高的时候工业品下跌的概率更高。但不论是样本内的 2017 年、2018 年，还是分样本内外测试的样本外，模型的表现都较为一般。我们认为这主要受到供给侧改革和环保限产能的影响，但此类政策性的影响对工业品定价的影响期限有限，我们认为长期来看模型仍具有一定的适用性。

图表 16: “公共因子”组成



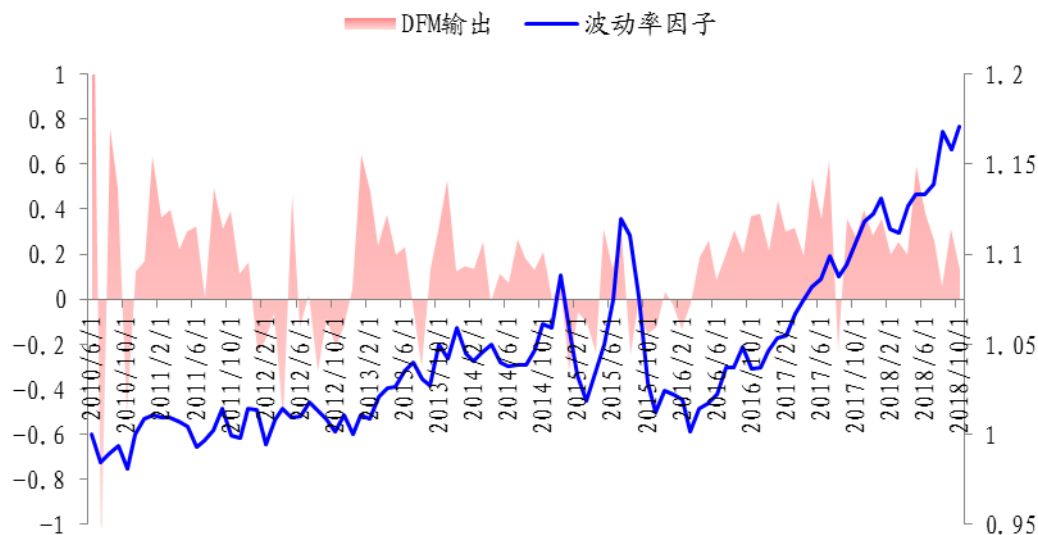
资料来源: Wind, 国盛证券研究所

3.4. 风格因子: 波动率

我们以 Barra 波动率因子收益率作为波动率因子的代理资产。模型定阶的结果 $k=1, m=3, p=0$, 其 1 阶系数绝对值与 2 阶系数绝对值相当, 因此受到“公共因子”变化程度的影响较大。

$$y_{t+1}^{\text{波动率因子}} = 0.10 + 0.41f_{1,t} - 0.32f_{1,t-1} + 0.05f_{1,t-2}$$

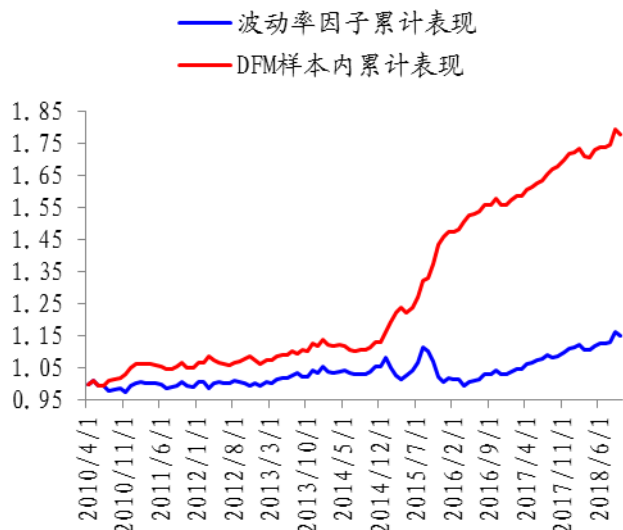
图表 17: DFM 预测模型信号



资料来源: Wind, 国盛证券研究所

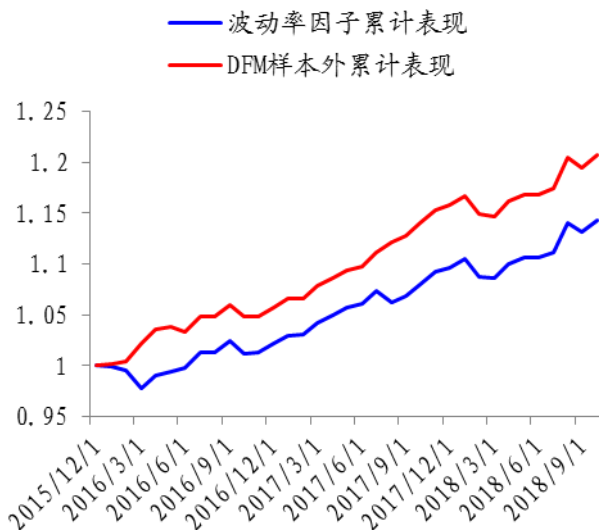
在样本内 DFM 模型胜率 71.29%, 同时我们也做了样本外测试: 将样本分为 2016 年之前和 2016 年之后两段, 通过每月滚动重计算动态因子以及预测模型进行预测, 样本外胜率为 79.41%。但由于样本外波动率因子本身波动不大, 还需继续观察验证。

图表 18: DFM 预测模型样本内择时表现



资料来源: Wind, 国盛证券研究所

图表 19: DFM 预测模型样本外择时表现



资料来源: Wind, 国盛证券研究所

“公共因子”的组成成分较为纯粹,基本为工业企业经济效益指标与企业景气指数。也就是说在经济基本面“加速变好”(1 阶滞后系数为正,2 阶滞后系数为负),经济基本面的变化好于线性外推预期的情况下,低波动的股票具有更高的超额收益。

图表 20: “公共因子”组成



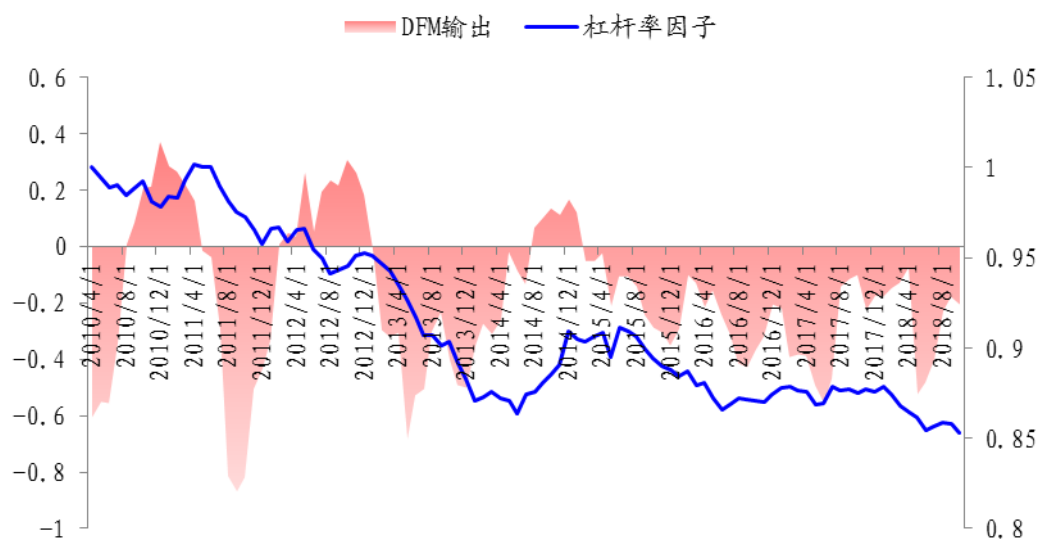
资料来源: Wind, 国盛证券研究所

3.5. 风格因子: 杠杆率

我们以 Barra 杠杆率因子收益率作为杠杆率因子的代理资产。模型定阶的结果 $k=2, m=1, p=0$, 因此其由两个“公共因子”决定。

$$y_{t+1}^{\text{杠杆率因子}} = -0.09 + 0.05f_{1,t} + 0.12f_{2,t}$$

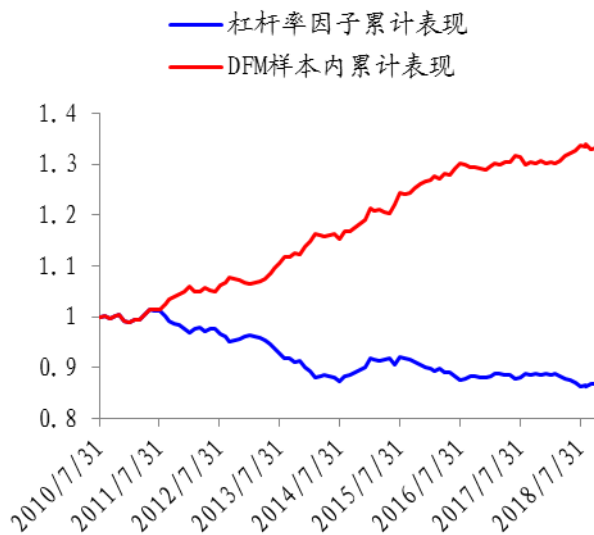
图表 21: DFM 预测模型信号



资料来源: Wind, 国盛证券研究所

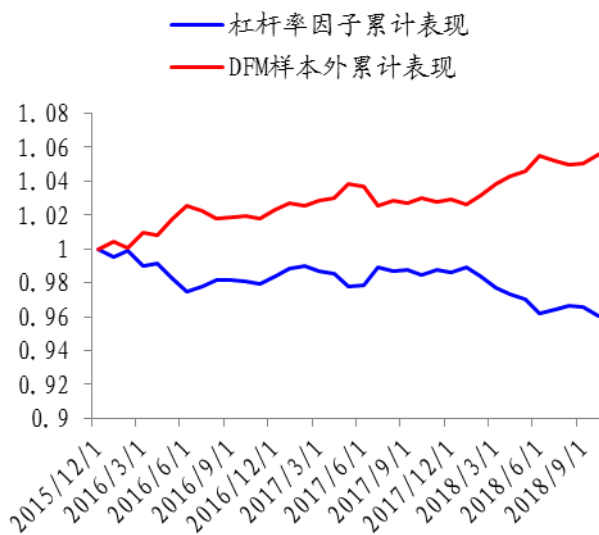
在样本内 DFM 模型胜率 65.00%，同时我们也做了样本外测试：将样本分为 2016 年之前和 2016 年之后两段，通过每月滚动重计算动态因子以及预测模型进行预测，样本外胜率为 61.76%。

图表 22: DFM 预测模型样本内择时表现



资料来源: Wind, 国盛证券研究所

图表 23: DFM 预测模型样本外择时表现



资料来源: Wind, 国盛证券研究所

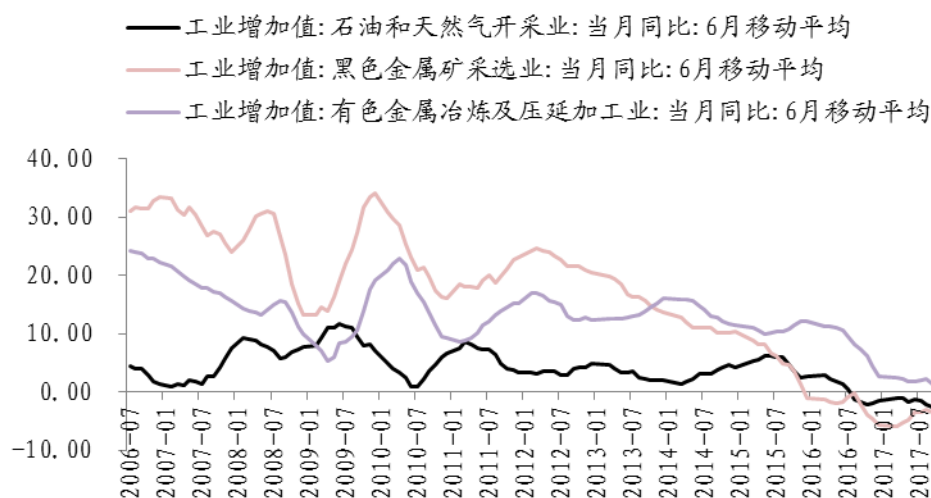
将回归系数代入代表“公共因子”的主成分，最终的指标配权如下。可以看到杠杆率因子的影响因素中，主要有：**工业企业经济效益指标，进出口指标，行业效益差（能源行业-黑色、有色金属行业）**等。当整体工业与进出口表现良好的情况下，高杠杆率的企业能够获取更高的收入，而当工业与进出口趋弱的情况下，低杠杆率的企业其应对风险的能力更强。能源行业与黑色、有色金属行业的工业增加值长期负相关，其差值可能与某种经济周期有关。

图表 24: 宏观变量最终权重



资料来源: Wind, 国盛证券研究所

图表 25: 能源与黑色、有色金属工业增加值



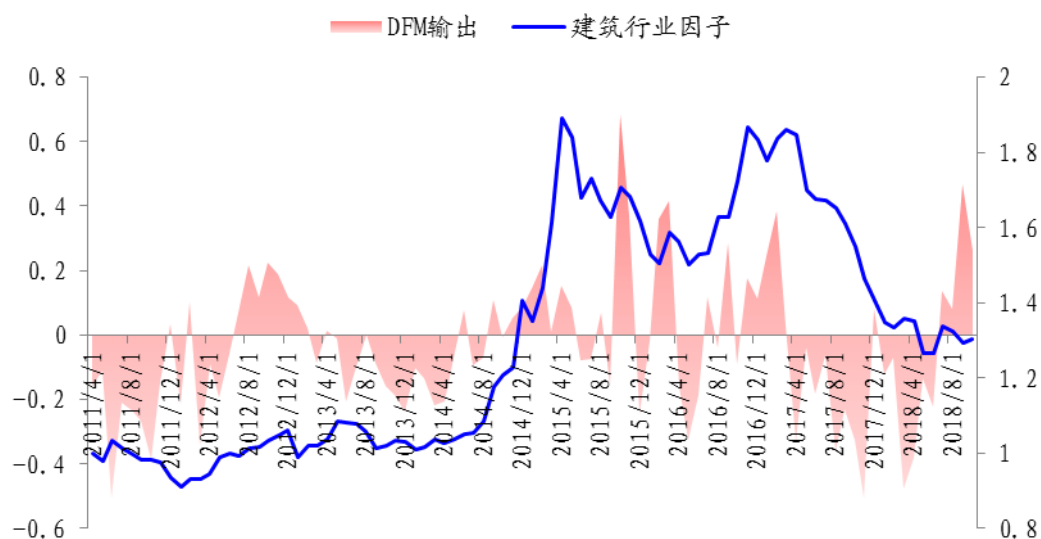
资料来源: Wind, 国盛证券研究所

3.6. 行业因子: 建筑

我们以建筑（中信）减去沪深 300 的收益率作为建筑行业因子的代理资产。模型定阶的结果 $k=1$, $m=2$, $p=1$, 因此其由一个“公共因子”决定, 并受到自身滞后项的影响, 具有一定的动量特征。

$$y_{t+1}^{\text{建筑行业因子}} = 0.01 - 0.03f_{1,t} + 0.13f_{1,t-1} + 0.16y_t$$

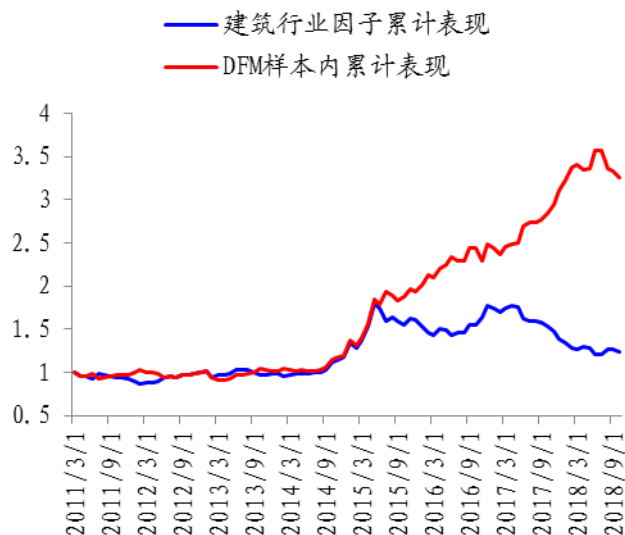
图表 26: DFM 预测模型信号



资料来源: Wind, 国盛证券研究所

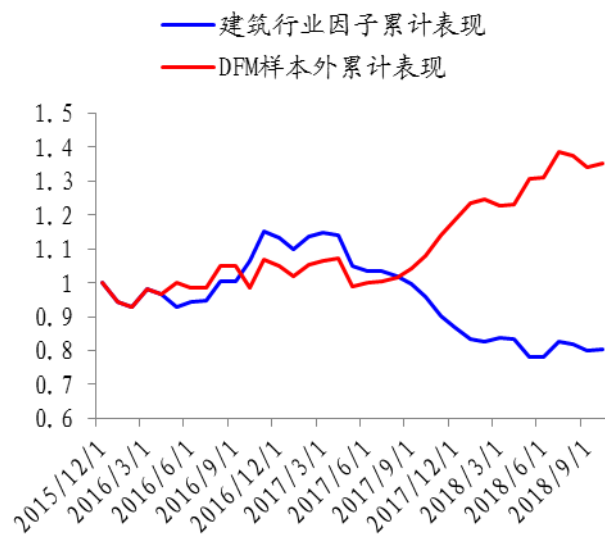
在样本内 DFM 模型胜率 68.13%，同时我们也做了样本外测试：将样本分为 2016 年之前和 2016 年之后两段，通过每月滚动重计算动态因子以及预测模型进行预测，样本外胜率为 64.71%。

图表 27: DFM 预测模型样本内择时表现



资料来源: Wind, 国盛证券研究所

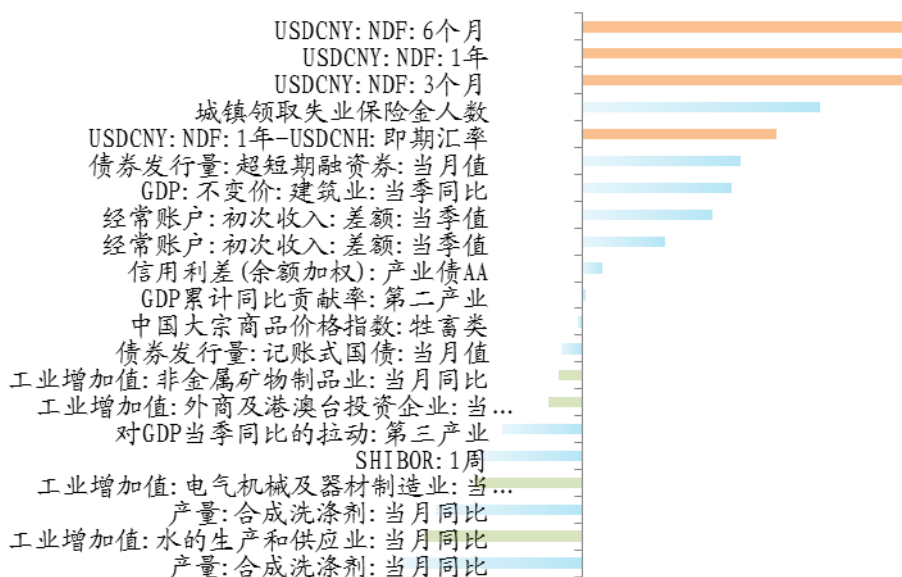
图表 28: DFM 预测模型样本外择时表现



资料来源: Wind, 国盛证券研究所

“公共因子”中正向贡献的主要为汇率和出口，负向贡献的主要为各类工业增加值和合成洗涤剂的产量。建筑企业签订海外合同时往往约定使用国际通行货币或者当地货币进行结算，实际情况中以美元结算居多。多数建筑公司存在较大汇率风险敞口，因此在人民币贬值的时候，建筑行业的将获得更多的汇兑收益。建筑行业的另一条主线与地产相同，即在工业不景气的时期历史一般会刺激地产和基建，因此利好建筑行业。

图表 29: “公共因子”组成



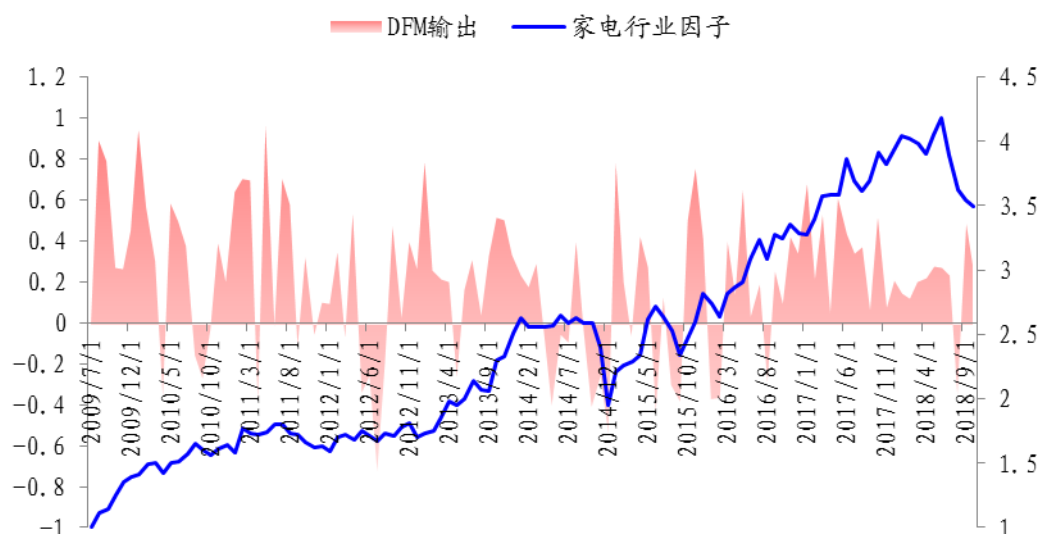
资料来源: 国盛证券研究所

3.7. 行业因子: 家电

我们以家电（中信）减去沪深 300 的收益率作为家电行业因子的代理资产。模型定阶的结果 $k=1$, $m=2$, $p=1$, 因此其由一个“公共因子”决定, 并受到自身滞后项的影响, 具有一定的反转特征。

$$y_{t+1}^{\text{家电行业因子}} = 0.15 + 0.34f_{1,t} - 0.26f_{1,t-1} - 0.01y_t$$

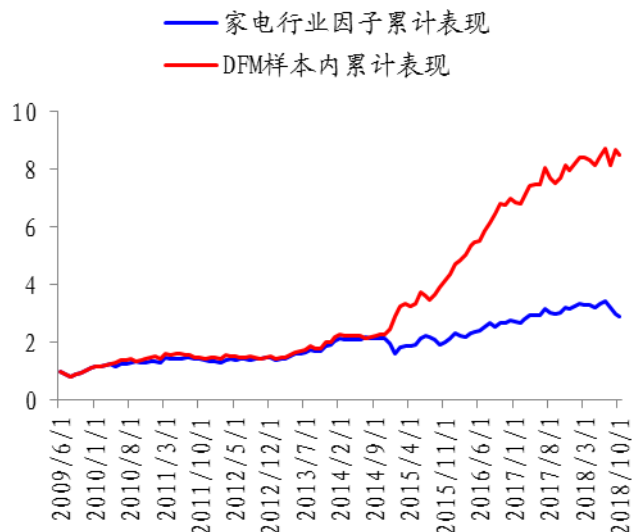
图表 30: DFM 预测模型信号



资料来源: Wind, 国盛证券研究所

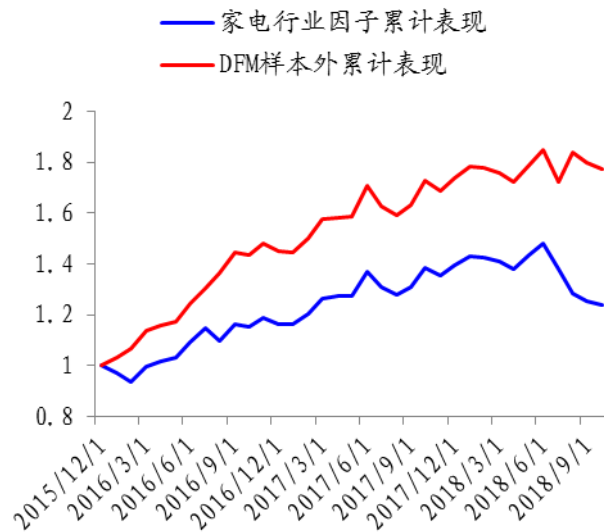
在样本内 DFM 模型胜率 68.75%，同时我们也做了样本外测试：将样本分为 2016 年之前和 2016 年之后两段，通过每月滚动重计算动态因子以及预测模型进行预测，样本外胜率为 64.71%。家电行业的超额收益本身就较高，动态因子预测模型通过避免了几次家电超额收益的大幅回撤使得收益更加显著。

图表 31: DFM 预测模型样本内择时表现



资料来源: Wind, 国盛证券研究所

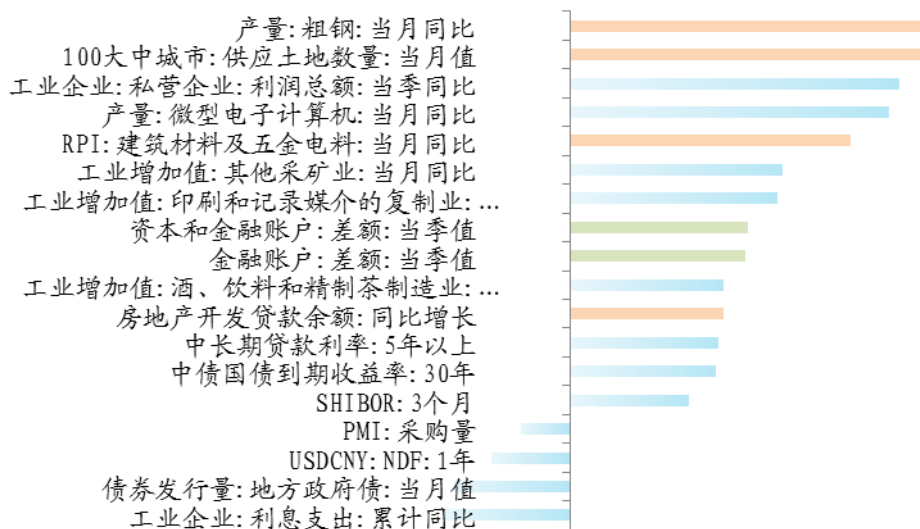
图表 32: DFM 预测模型样本外择时表现



资料来源: Wind, 国盛证券研究所

“公共因子”中权重较高的主要有：地产周期指标、外资净流入、工业企业经济效益等。

图表 33: “公共因子”组成



资料来源: Wind, 国盛证券研究所

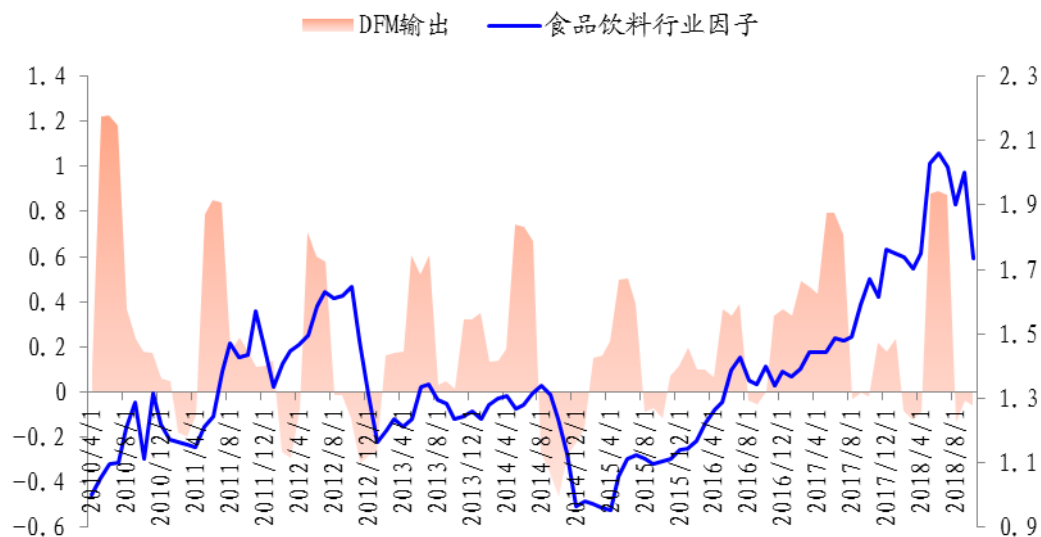
3.8. 行业因子: 食品饮料

我们以食品饮料（中信）减去沪深 300 的收益率作为食品饮料行业因子的代理资产。模

型定阶的结果 $k=1, m=1, p=1$ ，因此其由一个“公共因子”决定，并受到自身滞后项的影响，具有一定的动量特征。

$$y_{t+1}^{\text{食品饮料行业因子}} = 0.20 - 0.21f_{1,t} + 0.01y_t$$

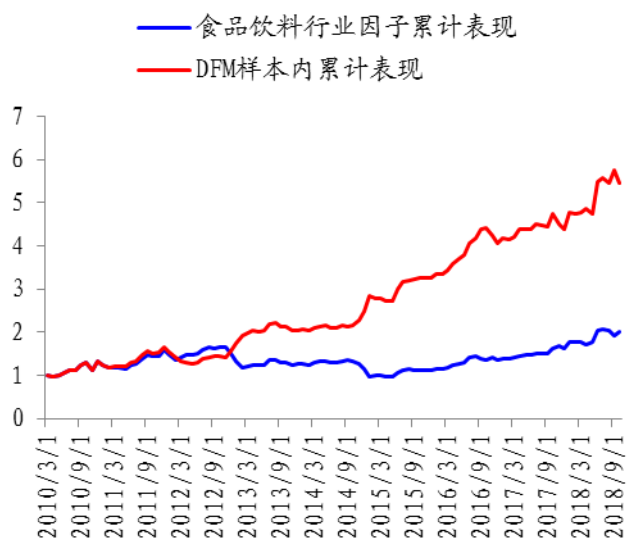
图表 34: DFM 预测模型信号



资料来源: Wind, 国盛证券研究所

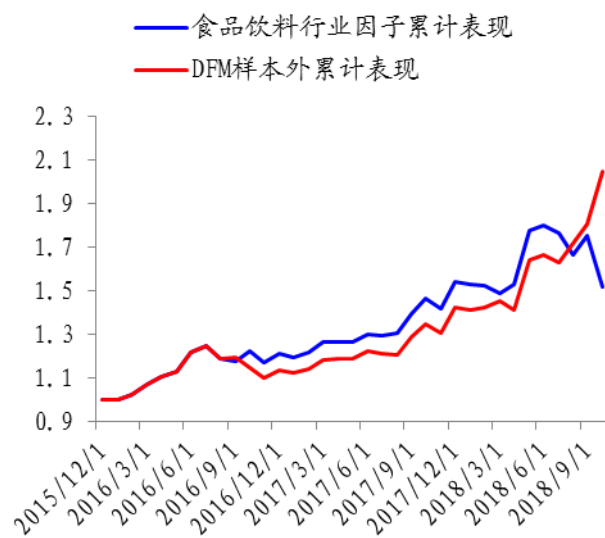
在样本内 DFM 模型胜率 66.99%，同时我们也做了样本外测试：将样本分为 2016 年之前和 2016 年之后两段，通过每月滚动重计算动态因子以及预测模型进行预测，样本外胜率为 70.59%。

图表 35: DFM 预测模型样本内择时表现



资料来源: Wind, 国盛证券研究所

图表 36: DFM 预测模型样本外择时表现

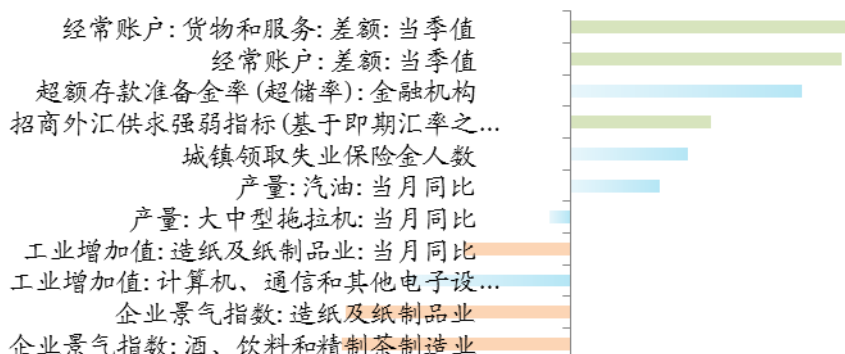


资料来源: Wind, 国盛证券研究所

“公共因子”中权重较大的主要有：食品饮料相关行业景气度、出口相关指标。食品饮料主要与酒、饮料和精制茶制造业、造纸及纸制品业有直接的联系，而与出口成负向关

系，出口变差经济有下行压力的时候，食品饮料作为防御性行业受益。

图表 37: “公共因子”组成



资料来源: Wind, 国盛证券研究所

4. 关于噪音与信息损失的思考

4.1. 宏观因子的 Occam 剃刀

在宏观因子选择的时候，由于阈值的设定或者想要结合尽量多的信息，我们往往会纳入较多的宏观因子进行建模，然而引入过多未在逻辑上充分考虑过的宏观因子容易带来较多的噪音，同时使得模型的结果更不易理解。因此我们需要对我们输入的宏观因子组进行精简。

Occam 剃刀的原则是：“如无必要，勿增实体”。因此我们采取算法为分别检验“丢掉”某一宏观因子后的模型拟合效果，如果某一宏观因子的缺失对模型拟合效果有较大的影响，则保留宏观因子；如果某一宏观因子的缺失对模型拟合效果没有任何影响，或者影响微乎其微，我们倾向于删除此因子。这样的处理在我们第三节的部分模型展示中，能够将宏观因子组删减 20%-50%，具有较明显的精简作用，这也代表在前几大主成分中，部分因子的贡献较小。当然我们并不能保证删减后的模型是最好的模型，只是在相同模型参数下，模型的可读性将会变得更高。

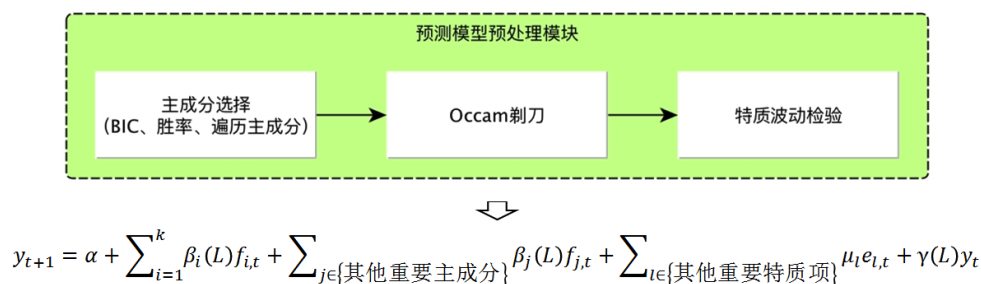
4.2. 主成分选取的顺序

由于我们在宏观因子选择的时候，通过几种检验的方式，已经保证了入选的宏观因子对待预测资产有一定的预测作用，而最大几个特征值对应的主成分确实也能解释这组宏观因子的大部分波动，所以大部分情况从前往后挑选主成分没有太大的问题。然而如果入选的宏观因子代表的驱动因素相关性并不高，部分含有重要信息的宏观因子无法在前几大主成分体现，而被归入后面的主成分，则从前往后挑选主成分的时候可能由于自由度的增加过多而被放弃。因此我们建议在挑选主成分的时候，不妨检验一下全部主成分的预测能力，以免遗漏重要预测信息。

4.3. 降维造成的信息损失

DFM有其假设的合理性，但是从数据角度来看，如果入选的宏观因子其具有预测作用的信息并非与其他因子的公共部分 f_t ，而是宏观因子的特质部分 e_{it} ，那么通过主成分降维的“平滑”反倒会使得有用的信息被抹平，就会造成预测信息丢失的问题。因此我们建议，在挑选完主成分之后，对于每一个原始宏观因子，都对已选主成分进行回归，将残差部分放入模型检验其预测能力，避免因降维带来过多的预测信息损失。

图表 38: 动态因子模型建模过程中的噪音、信息损失、信息扭曲问题对策



资料来源：国盛证券研究所

风险提示

模型建立基于历史数据，在政策、经济结构等因素变化的情况下模型可能失效。模型建立从数据出发，数量关系未必对应因果关系。

免责声明

国盛证券有限责任公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告的信息均来源于本公司认为可信的公开资料，但本公司及其研究人员对该等信息的准确性及完整性不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，可能会随时调整。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司力求报告内容客观、公正，但本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。

本报告版权归“国盛证券有限责任公司”所有。未经事先本公司书面授权，任何机构或个人不得对本报告进行任何形式的发布、复制。任何机构或个人如引用、刊发本报告，需注明出处为“国盛证券研究所”，且不得对本报告进行有悖原意的删节或修改。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的任何观点均精准地反映了我们对标的证券和发行人的个人看法，结论不受任何第三方的授意或影响。我们所得报酬的任何部分无论是在过去、现在及将来均不会与本报告中的具体投资建议或观点有直接或间接联系。

投资评级说明

投资建议的评级标准		评级	说明
评级标准为报告发布日后的 6 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准。	股票评级	买入	相对同期基准指数涨幅在 15%以上
		增持	相对同期基准指数涨幅在 5%~15%之间
		持有	相对同期基准指数涨幅在 -5%~+5%之间
		减持	相对同期基准指数跌幅在 5%以上
	行业评级	增持	相对同期基准指数涨幅在 10%以上
		中性	相对同期基准指数涨幅在 -10%~+10%之间
		减持	相对同期基准指数跌幅在 10%以上

国盛证券研究所

北京

地址：北京市西城区锦什坊街 35 号南楼

邮编：100033

传真：010-57671718

邮箱：gsresearch@gszq.com

南昌

地址：南昌市红谷滩新区凤凰中大道 1115 号北京银行大厦

邮编：330038

传真：0791-86281485

邮箱：gsresearch@gszq.com

上海

地址：上海市浦明路 868 号保利 One56 10 层

邮编：200120

电话：021-38934111

邮箱：gsresearch@gszq.com

深圳

地址：深圳市福田区益田路 5033 号平安金融中心 101 层

邮编：518033

邮箱：gsresearch@gszq.com