

金工研究/深度研究

2018年05月29日

林晓明 执业证书编号：S0570516010001
研究员 0755-82080134
linxiaoming@htsc.com

黄晓彬 执业证书编号：S0570516070001
研究员 0755-23950493
huangxiaobin@htsc.com

相关研究

- 1 《金工：华泰单因子测试之财务质量因子》
2018.05
- 2 《金工：华泰单因子测试之资金流向因子》
2018.05
- 3 《金工：Smartbeta 在资产配置中的优势》
2018.05

周期理论与机器学习资产收益预测

华泰金工市场周期与资产配置研究

基于资产周期状态与市场表现的关系规律，采用机器学习挖掘预测逻辑

本文首先介绍美林时钟等宏观择时模型，分析其根据经济周期状态的划分进行资产配置的原理，同时指出其应用于中国等新兴市场投资时遇到的挑战。其次，简要回顾华泰金工周期系列研究基于市场统一周期规律提出的周期三因子资产定价模型，以及机器学习挖掘市场规律的原理。最后，采用机器学习的方法发现资产所处的周期状态与其未来市场表现的内在逻辑，实现对资产收益排序的概率预测，并通过对全球市场和中國市场的实证研究，证明该方法对指导资产配置的有效性。

市场周期运动的主导特征是基于资产周期状态预测其未来表现的基础

美林时钟模型为宏观择时到资产配置的投资实践提供了较为清晰的思路和扎实的理论依据，具有启发意义。但美林投资时钟模型并不是一个实时、定量的交易策略，其有效与否的关键在于对经济周期状态的判断是否正确。华泰金工周期系列研究通过傅里叶变换、联合谱估计等信号处理方法，发现并证实了市场中广泛存在 42 个月、100 个月和 200 个月左右的共同周期。以此为基础，提出了华泰量化投资时钟“周期三因子定价与资产配置模型”。通过计算金融资产同比序列与其周期三因子的回归拟合值、拟合值的增加值等，实现对资产周期状态比较精确且全面的测度。

机器学习能有效挖掘资产周期状态与未来表现间的内在逻辑

资产收益率分布的不对称性或导致对资产短期收益表现的确定性预测结果出现与实际的偏差，概率结果比确定性结果常常更为有效。同比序列受短期噪声影响小，适合作为周期状态“特征变量”的计算基础。机器学习模型能以概率方法建立起资产同比周期状态与未来表现间的非线性联系，并给出收益排序的概率预测；仿真测试证明本文机器学习模型对挖掘上述联系的有效性。此外，为避免模型对参数选择的敏感性，本文引入了集成学习法，计算多种可行参数组合的预测结果，采用“少数服从多数”的原则，平均后确定最终结果，降低模型对参数依赖，更全面有效的利用历史规律。

实证结果证明周期理论与机器学习的研究方法具有不同市场的普适性

基于周期理论和机器学习方法预测结果的资产配置实证显示，策略应用于全球和中国市场均有较好表现。与基准的等权配置模型相比，基于周期理论和机器学习方法的策略在年化收益、最大回撤等风险收益指标均有明显提升，充分证明了机器学习能够挖掘市场周期规律并实现更加有效的预测。由于市场周期规律的普遍存在，机器学习挖掘投资规律的方法也具有普适性。本文方法应用于全球股债资产配置，在 2004 年 1 月至 2018 年 3 月的样本区间集成学习综合各种参数集的预测结果，取得年化 20.5%，最大回撤 14.1%，夏普比 1.36 的表现；应用于中国大类资产配置的表现也明显好于基准。

风险提示：本文基于华泰金工周期系列研究对全球各类经济金融指标长达近百年样本的实证检验结果，确定周期长度。然而市场存在短期波动与政策冲击，就每轮周期而言，暂无法判断具体长度。周期长度只是估计值，可能存在偏差；历史规律存在失效风险。

正文目录

本文导读	3
市场周期运动的主导特征是基于资产周期状态预测其未来表现的基础	5
“美林时钟”关于经济周期状态与资产配置的分析逻辑	5
华泰金工周期系列研究实现对资产周期状态更加精确全面的测度	6
机器学习挖掘资产周期状态与其未来市场表现的内在逻辑，实现对收益排序的概率预测	8
机器学习的基本原理：以二元分类的逻辑回归为例	8
周期状态“特征变量”的选择：周期三因子同比序列拟合值与拟合值的增加值	10
短期收益率预测：概率结果比确定性结果更有效	12
机器学习：通过挖掘资产周期规律与其市场表现的内在逻辑预测其收益排序	13
集成学习：避免对参数选择的敏感性，更全面有效的利用历史规律	14
实证分析：周期理论与机器学习的研究方法具有对不同市场的普适性	16
预测先行指标的选择：“特征变量”组合的方式与原因分析	16
仿真测试：机器学习在挖掘资产周期状态与未来表现逻辑关系具有有效性	18
策略回测：基于机器学习预测结果的资产配置策略业绩优势明显	20
全球主要股债细分资产指数实证结果	21
全球大类资产等权指数实证结果	24
中国主要大类资产指数实证结果	27
2018 下半年的资产表现排序的预测结论	29
风险提示	32
附录一：采用同比序列作为资产收益预测的数据基础的原因	33
附录二：资产收益率分布的不对称性致概率预测好于确定性预测	37
附录三：机器学习之 Softmax 模型原理	38

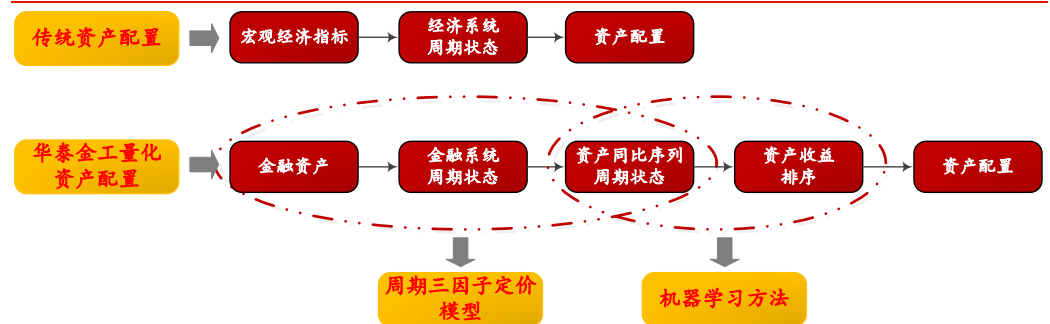
本文导读

Brinson 等人（1986）的研究指出，有效的大类资产配置是投资成功的关键。2004 年，美林时钟的提出，为从宏观择时到资产配置的研究提供了理论依据。然而，美林时钟由于对周期状态的判断、以及对周期状态与资产价格之间映射关系的构建缺乏定量标准，在应用于美国以外的市场如中国等新兴市场时有效性受到挑战。华泰金工提出周期三因子资产定价模型，实现对资产周期状态的定量测度。在此基础上，本文融合机器学习方法，提出一种根据资产所处的周期状态预测其收益表现的定量方法。

本文将在第一部分介绍以美林时钟为代表的宏观择时模型并回顾华泰周期三因子定价模型。美林时钟由美国投行美林证券于 2004 年提出，其核心思想——经济具有周期性，在不同经济周期下配置合适的资产以获取投资回报，很好地补充了其他大类资产配置模型对宏观经济研究的不足，将经济周期与资产配置策略动态地联系起来，为投资者提供了从经济周期分析到资产配置的研究框架。在对 1973 年至 2004 年美国市场数据的实证检验中，该模型很好地诠释了历史上美国大类资产轮动的规律，但其应用于现阶段的中国市场投资时似乎并不十分有效。究其原因，可能有两点：一是美林时钟仅通过经济增长和通货膨胀来识别经济周期的市场分析逻辑在中国等新兴经济体并不完全适用；二是其对宏观经济状态和金融资产价格之间映射关系的构建缺乏定量的标准。

华泰金工在美林时钟所提出的研究框架下，引入信号处理手段，对金融经济系统周期状态和金融资产价格进行量化测度。本系列研究首先通过信号处理、时间序列分析、统计分析等手段，发现并证明了金融经济系统普遍存在三大共同周期，即 42 个月左右的基钦周期，100 个月左右的朱格拉周期和 200 个月左右的库兹涅茨周期，这三大周期对各项金融经济指标普遍具有较强的解释力度，使得基于周期的量化资产配置成为可能。进一步，提出周期三因子资产定价模型：通过对多项金融资产和宏观经济指标的周期相位分析，推断系统所处的周期状态，进而预测系统中各指标和资产价格未来的涨跌变化。相比于传统宏观择时模型的定性分析，该模型为周期状态的识别和资产价值的判断提供了定量的测度方法，且对全球各项金融资产具有普遍的适用性。

图表1： 量化资产配置模型与传统模型思路对比



资料来源：华泰证券研究所

然而，周期三因子模型也有其局限性。一是该模型仅能得出对资产同比的预测，而不能得出对下期收益的直接观点；二是该模型直接应用于资产配置时，对指标的选择较为敏感。针对这两个问题，本文采用机器学习方法，模仿经济学家基于经济周期状态预测资产投资机会的决策过程，并充分发挥机器学习挖掘历史规律的优势，实现对下一期资产收益排序的预测。在前期研究中我们发现，资产同比序列所处的周期状态与资产下期收益间存在密切的联系，但该联系并非简单的一一映射。本文以资产所处周期状态作为确定下期收益的先行指标，进而利用机器学习中多元分类器（Softmax 模型）建立起先行指标与资产收益表现的相关关系，从而得出关于资产下期收益的概率性预测。这种概率性预测不直接提供关于资产收益的精确数值，但是可以较为准确地给出资产周期状态和收益排序之间的关系。为了降低模型对指标选取的依赖性，本文选取 Bagging 集成学习方式，综合使用多种先行指标集得出预测结论，以“少数服从多数”的预测原则，提高了模型的稳定性。至此，本文完成了对周期三因子模型的改进，提出一种基于机器学习方法的资产收益表现的量化预

测方法。该方法的原理和应用步骤将在本文第二部分具体说明。

本文第三部分对上述量化预测方法的有效性进行实证检验，并应用该方法对未来的资产价格变化方向进行预测。实证检验发现，该策略对全球大类资产配置和中国的大类资产配置均有较好的表现，较线性预测方法有明显提升。本文方法应用于全球股债资产配置的实证，在 2004 年 1 月至 2018 年 3 月样本区间，集成学习综合各种参数集的预测结果，取得年化 20.5%，最大回撤 14.1%，夏普比 1.36 的表现；而直接应用线性预测逻辑的策略结果仅能达到年化收益率 7.1%，最大回撤 22.5%。这充分证明了机器学习方法能够深入挖掘市场周期规律并实现有效的资产收益预测。最后，应用以上研究，本文给出 2018 年下半年的资产配置建议。

市场周期运动的主导特征是基于资产周期状态预测其未来表现的基础

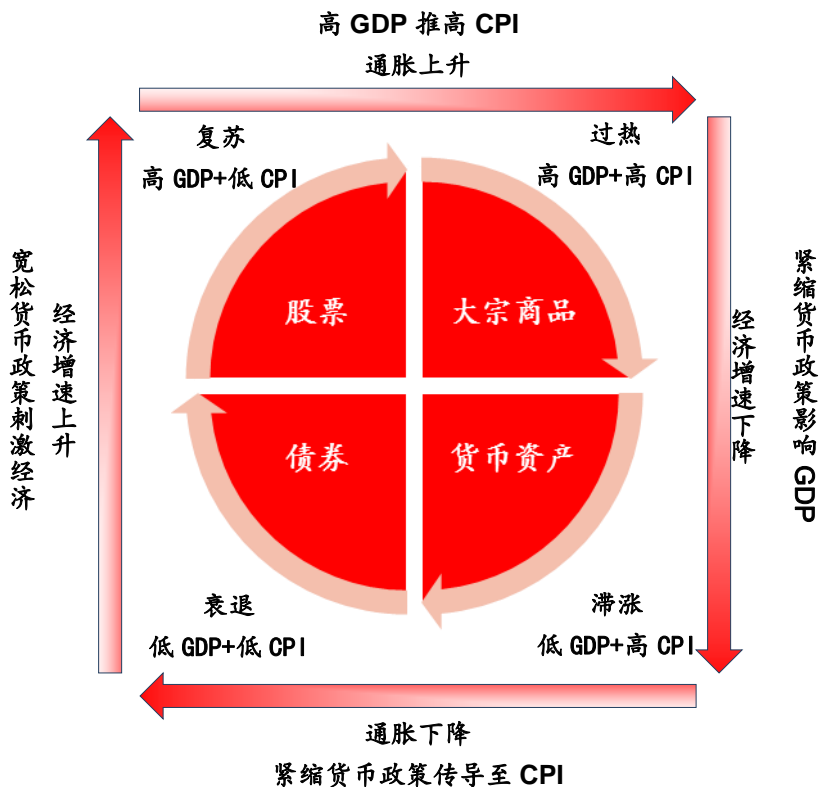
2004 年，美林证券提出宏观择时研究的基石——美林时钟模型，为宏观择时到资产配置的研究提供了较为清晰的思路和扎实的理论依据，对大类资产宏观择时的研究具有启发意义。但美林投资时钟模型并不是一个实时、定量的交易策略，其有效与否的关键在于对经济周期状态的判断是否正确。因此，美林时钟作为大类资产宏观择时研究的基石，虽然提供了逻辑合理的框架，但其框架中最为核心的经济周期的判断以及宏观经济与金融资产价值间映射关系缺乏定量判断标准，这就导致了美林时钟模型在指导中国等新兴市场的投资时面临挑战。

受到美林时钟模型的启发，以及考虑其存在的完善空间，华泰金工周期系列研究通过傅里叶变换、联合谱估计等信号处理方法，证明了全球主要国家市场变量，包括股票市场指数、债券指数、大宗商品指数等资产价格数据，以及 CPI、PPI 等宏观经济变量普遍存在 42 个月，100 个月和 200 个月左右的三个共同周期。并且，进一步的实证研究发现，42 个月，100 个月和 200 个月左右的周期信号能够比较准确的判断市场运动方向，并对资产的未来表现有较好的预测作用。以此为基础，进一步提出了华泰量化投资时钟“周期三因子定价与资产配置模型”。通过从多项金融资产和宏观经济指标中提取三个周期信号（或称“周期三因子”信息），推断系统所处的周期状态，从而预测系统中各指标和资产价格未来的涨跌变化。相比于传统宏观择时模型的定性分析，该模型对周期的识别和对资产价值的判断更具有实践价值，对全球各项金融资产具有普遍的适用性。

“美林时钟”关于经济周期状态与资产配置的分析逻辑

2004 年，美国投行美林证券（Merrill Lynch）发表了《The Investment Clock》一文，文中通过对美国超过 30 年的数据统计分析，提出了将资产轮动及行业策略与经济周期联系起来的资产配置模型，美林投资时钟由此正式走入大家的视野。

图表2：“美林时钟”轮动机制



资料来源：美林证券、华泰证券研究所

美林投资时钟的主要原理是根据经济增长趋势和通货膨胀趋势，将经济周期划分为复苏、过热、滞胀、衰退四个阶段，将资产类别划分为债券、股票、大宗商品和现金四大类，并分别给出了每个周期阶段适合的投资标的：

在衰退阶段（低 GDP+低 CPI），经济增长率低于潜在增长率，市场需求疲软，产出缺口持续扩大，超额的生产能力和下跌的大宗商品价格使得通胀率更低。政府不得不采取宽松的货币政策（减息）及积极的财政政策（减税）以刺激经济增长。此时，收益率曲线急剧下行，债券是最佳选择。

在复苏阶段（高 GDP+低 CPI），刺激政策开始起作用，GDP 增长率加速，产出缺口缩小，因为空置的生产能力还未耗尽，通胀率还在继续下行。而周期性的生产能力扩充强劲，需求端回暖加上生产成本下行，企业盈利能力大幅上升，这个阶段投资股票是最佳选择。

在过热阶段（高 GDP+高 CPI），经济增长率超过潜在经济增长率，市场需求旺盛，企业开始面临产能约束，企业生产能力增长减慢，导致通胀抬头。中央银行加息以求将经济拉回到可持续的增长路径上来，使得收益率曲线上行，债券表现往往较差。经济活动的过热以及较高的通货膨胀使大宗商品成为收益最高的资产。

在滞胀阶段（低 GDP+高 CPI），经济增长率依然超过潜在增长率，但呈减速趋势，产出正缺口逐渐减小，而通胀仍然处于上行阶段。企业成本日益上升，为了保持盈利不得不提高产品价格，导致价格上涨并进一步推升通胀，企业业绩开始出现停滞甚至下滑，股价上涨乏力。而通胀位于高位导致利率也维持高位，这就限制了债券市场的回暖步伐。此时持有现金才是最佳选择。

美林投资时钟的核心贡献在于指出了经济波动具有周期性，而且通过数据实证表明，如果能在不同经济周期运行状态下配置合适的资产/行业，可以获得显著的超额回报。但美林投资时钟模型并不是一个实时、定量的交易策略，其有效与否的关键在于对经济周期的判断是否正确。因此，美林时钟作为大类资产宏观择时研究的基石，虽然提供了逻辑合理的框架，但其框架中最为核心的经济周期的判断以及宏观经济与金融资产价值间映射关系缺乏定量判断标准，这就导致了美林时钟模型在指导投资时存在挑战。比如它只针对经济周期和货币周期做出投资某种大类资产的判断，没有考虑当期资产本身的价格水平是否合适，容易出现高回撤风险；再比如在后危机时代，各个国家央行频频调整货币政策，试图改变经济下行的趋势，美林时钟模型往往不能及时反映货币政策变动可能引发的市场方向短期变化，导致其有效性降低；既往研究中关于经济周期的划分以及大类资产收益率的测算口径均没有统一的标准，因此针对资产轮动现象的解释具有一定主观性，等等。根据已有的研究和检验结果来看，美林时钟模型在美国市场的表现要好于在中国等其他市场的表现。

美林时钟对大类资产宏观择时的研究具有启发意义，但宏观择时研究仍面临许多挑战，包括经济周期的确定，经济周期与资产表现的关系规律等。自美林时钟提出后，学者们从未停止对经济周期识别、宏观经济与资产价格映射关系的探索，然而传统的研究由于缺乏有效的数据挖掘分析工具，往往很难取得大的突破。

华泰金工周期系列研究实现对资产周期状态更加精确全面的测度

通过对美林时钟的分析，我们可以发现：传统的宏观择时方法是建立在经验性地对经济周期数据与资产收益率映射关系进行研究的基础之上的。因此，传统的宏观择时方法存在两方面不足：一是模型预测结果过于依赖经验性的经济周期划分方式，二是缺乏定量的针对宏观经济状态与金融资产价值映射关系的研究。不同的人依据不同标准，很有可能会划分出不同的周期状态；即便对于周期有大致统一的认识，关于周期状态与具体资产收益表现的对应关系也难以定量刻画。这种基于经济周期状态判断的资产轮动规律并不是十分稳定可靠。正因如此，我们才需要正视金融经济系统的复杂性，引入信号处理、机器学习、统计等数据挖掘工具，实现对资产周期状态更加精确且全面的测度。

华泰金工“周期系列研究”通过傅里叶变换、联合谱估计等信号处理方法，证明了全球主要国家市场变量，包括股票市场指数、债券指数、大宗商品指数等资产价格数据，以及 CPI、PPI 等宏观经济变量普遍存在 42 个月、100 个月和 200 个月左右的三个共同周期。这三个周期的长度和古典经济学研究中的基钦周期、朱格拉周期、库兹涅茨周期基本一致，且通过大量市场数据均得以验证。根据华泰周期系列深度报告《金融经济系统周期的确定》的研究结果，对于大多数经济金融指标，周期三因子能够解释同比序列 70% 以上的波动。下表展示了 17 个全球主要市场指数同比序列的周期三因子线性回归模型的拟合结果，可以看到，对于绝大多数指数，其周期三因子回归模型的可决系数 R^2 均在 70% 以上，说明模型能够准确的把握资产同比序列的历史走势，且能对其未来的走势起到一定的预测作用。

图表3： 全球主要市场指数同比序列周期三因子回归结果

	短周期回归系数 β_1	中周期回归系数 β_2	长周期回归系数 β_3	可决系数 R^2
上证综指	0.574	0.240	0.030	0.624
标普 500	0.187	0.133	0.023	0.711
富时 100	0.159	0.114	0.011	0.670
法国 CAC40	0.221	0.162	0.024	0.714
德国 DAX	0.231	0.147	0.015	0.662
日经 225	0.218	0.200	0.036	0.678
恒生指数	0.245	0.110	0.018	0.496
美国总回报指数	0.023	0.018	0.008	0.525
泛欧总回报指数	0.098	0.121	0.017	0.605
中国总回报指数	0.047	0.034	0.008	0.760
CRB 食品现货	0.196	0.142	-0.049	0.690
CRB 食用油现货	0.310	0.182	-0.038	0.688
CRB 家禽现货	0.220	0.130	0.023	0.686
CRB 金属现货	0.352	0.227	0.036	0.724
CRB 工业现货	0.210	0.108	0.015	0.721
CRB 综合现货	0.194	0.089	0.011	0.722
CRB 纺织现货	0.122	0.070	0.009	0.660

资料来源：华泰证券研究所

因此，我们提出了“全球金融经济体由统一的三周期驱动”的核心假设，并开启了经济周期与资产配置研究的全新视角，即华泰量化投资时钟“周期三因子模型”——将 42 个月、100 个月、200 个月的短周期、中周期、长周期抽象为三因子，用以解释资产价格的运动变化，并预测未来走势。在华泰周期系列深度报告《周期三因子定价与资产配置模型》中，我们通过量化周期三因子定价模型形成关于资产收益率的观点，不同于 CAPM 等模型直接对资产的价格进行预测，量化周期三因子模型着眼于资产价格同比序列中蕴含的周期信号，通过发掘周期信号与资产收益的关系规律，形成对资产下期走势的预测观点。

然而，周期三因子模型也有其局限性。一是该模型仅能得出对资产同比的预测，而不能得出对下期收益的直接观点；二是应用该模型指导资产配置时，配置结果对于模型中的参数变化较为敏感。针对这两个问题，下文将采用机器学习方法模仿经济学家基于经济周期状态预测资产投资机会的决策过程，并充分发挥机器学习挖掘历史规律的优势，实现对下一期资产收益排序的预测。

机器学习挖掘资产周期状态与其未来市场表现的内在逻辑，实现对收益排序的概率预测

本节将对文章所提出的资产收益预测的方法进行详细的阐述。首先，我们以二分类的逻辑回归为例介绍机器学习的基本原理：机器学习模型实际上是通过一种非线性的方式，建立起特征变量 X 与因变量 Y 所属类别之间的联系（在资产收益的预测上，我们将某种资产收益排序是否第一定义为一种类别，比如有两种资产：股票、债券，当股票涨幅排序第一时，定义为市场状态属于类别 1；当债券涨幅排序第一时，定义为市场状态属于类别 2，以此类推。如果预测的资产类别有 N 种，那么就是多元分类的问题，相应的市场状态就有 N 类）。接着，分析和验证了资产同比序列的周期状态能够作为衡量资产未来收益的特征变量；与此同时，由于资产短期收益率分布的有偏性，我们认为基于线性回归模型的资产收益确定性预期结果不能有效预测资产未来表现，概率预期结果相较于确定性结果更具有实际应用价值。

本研究考虑根据现有信息，对于资产未来表现的概率预期结果进行估计，而以逻辑回归为代表的机器学习分类模型能够在给定特征变量 X 的前提下，变量 Y 属于类别的概率估计。然而，逻辑回归仅能处理二元分类问题，为了实现同时配置多类资产的目标我们需要采用多元分类方法。

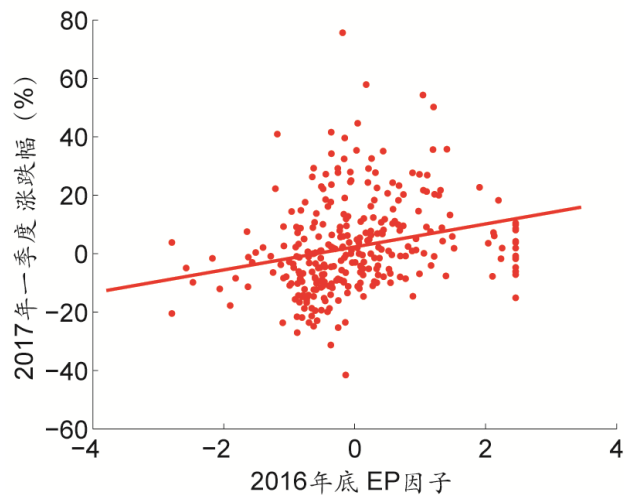
Softmax 模型（Soft+Max，读者可以从其命名方式大致猜到这是一个概率模型）是逻辑回归的多元扩展，该模型能够有效的处理多种类别的分类问题，并给出变量 Y 属于每个类别的概率估计。进一步，为了避免 Softmax 模型对参数选择的敏感性可能导致的模型过拟合，我们引入了集成学习策略，对于不同的单一模型结果进行集成，以降低模型参数依赖的前提下更全面有效的把握和利用历史规律。

机器学习的基本原理：以二元分类的逻辑回归为例

在华泰金工人工智能系列报告之一《人工智能选股框架及经典算法简介》中，我们对机器学习的各类基本方法进行了详细的阐述，具体细节本文不再赘述。为帮助读者更好的把握本文脉络，这里我们仅以逻辑回归为例，对于机器学习的基本原理进行简单的介绍。

逻辑回归是机器学习中最常见的监督学习方法之一，可以看作线性回归在分类变量问题上的推广。在线性回归中，我们建立起了特征变量 X 同连续变化的因变量 Y 之间的线性相关关系。例如，我们希望能用股票的市盈率因子预测收益率，选取沪深 300 成分股 2016 年底的市盈率以及 2017 年一季度涨跌幅。对市盈率 TTM 取倒数，进行中位数去极值和标准化处理，得到 EP 因子，相关数据如下图表所示。可以看到，利用线性回归方法，我们可以找到一条直线 $y = w_0 + w_1x$ ，该直线能够反映出因变量 Y 与特征变量 X 之间的变化趋势。根据下图，市盈率越低，EP 因子越大，那么股票越有可能上涨。

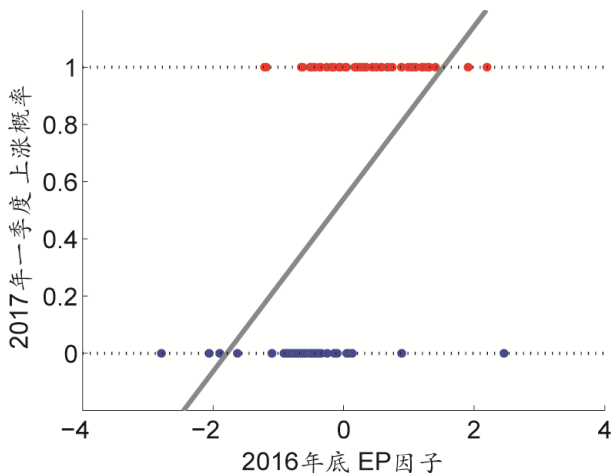
图表4： 市盈率 EP 因子和涨跌幅的线性模型



资料来源：《人工智能选股框架及经典算法简介》，华泰证券研究所

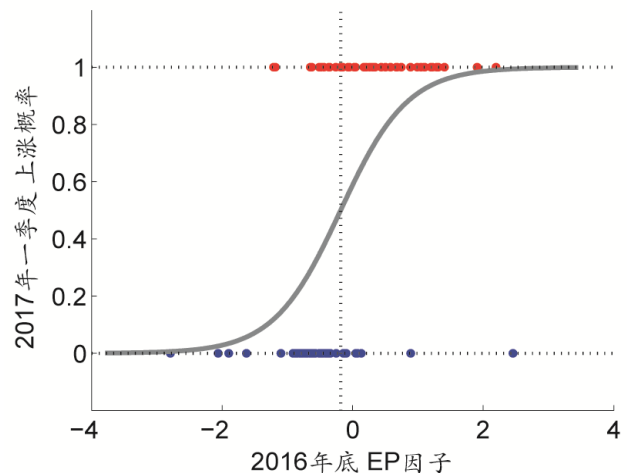
然而，线性回归的基本假设之一为：因变量 Y 连续变化，例如，在上例中股票收益率涨跌幅为连续变化变量。然而，很多时候我们并不需要预测股票下个月具体的涨跌幅，而是希望预测股票下个月会上涨还是下跌。换言之，我们面对的是“分类”问题，而非“回归”问题，即自变量 Y 的取值范围并非某个区间，而是若干个离散的类别变量。特别的，对于二元分类问题，自变量 Y 的取值范围为 $\{0,1\}$ 。同样上文利用市盈率预测股票涨跌为例，这里选取沪深 300 成分股 2017 年一季度的涨跌幅排名前 50 名和后 50 名的个股，计算 2016 年底的市盈率 EP 因子。将涨幅前 50 的个股定义为类别 $y=1$ ，跌幅前 50 的个股定义为类别 $y=0$ 。下文左图展示了线性回归的结果，可以看到，线性回归对于二元类别变量的拟合效果较差：图形中的红点与蓝点代表已发生的事实，红点代表 2017 年一季度涨跌排名前 50 的股票，那么它的估计值应当尽量接近 1；反之，蓝点则应当尽量接近 0，而线性回归，只是“一条线”，他的拟合能力有限。且自变量 Y 的估计值并不具有实际的含义：可能出现大于 1 或者小于 0 的预测值，当出现这样的数值时，我们无法判断股票属于涨跌排名前 50 或涨跌排名后 50 的概率。

图表5： 线性回归模型拟合二分类数据



资料来源：《人工智能选股框架及经典算法简介》，华泰证券研究所

图表6： 逻辑回归模型拟合二分类数据



资料来源：《人工智能选股框架及经典算法简介》，华泰证券研究所

与线性回归不同，逻辑回归能够建立起因变量 Y 属于某一特定类别的概率同特征变量 X 之间的非线性关系，具体模型如下：

$$P(Y = 1|X) = \frac{e^{\theta X}}{1 + e^{\theta X}}$$

$$P(Y = 0|X) = 1 - P(Y = 1|X) = \frac{1}{1 + e^{\theta X}}$$

上文右图展示了对市盈率预测股票涨跌例子的逻辑回归拟合结果，图中灰色曲线为拟合曲线，曲线的参数 θ 根据历史数据估计得到。可以看到，对于给定的特征变量 X （即 EP 因子），因变量 Y 的估计值均在 $[0,1]$ 区间之内，为因变量属于类别 1 的概率估计，该曲线的拟合能力明显好于线性回归：通过参数 θ 的估计，可以使得已发生的事实尽量落入相应的类别数值 1 或 0，且 Y 具有很好的概率含义。

综上，在逻辑回归中，我们根据已知的“特征变量” X 和“标签” Y ，通过历史数据“训练”得到一个反映两者内在非线性关系的模型。如果这种关系规律在未来一段时间内能够延续，那么任意给出一个股票当前时刻的 EP 因子 X ，我们就可以“预测”该股票未来时刻的上涨或者下跌的概率，即得到 $P(Y = 1|X)$ 和 $P(Y = 0|X)$ 的估计值。根据已有的特征和标签历史数据训练模型，使用特征变量在未来时刻的估计值对标签进行预测，是机器学习最核心的两个环节。

周期状态“特征变量”的选择：周期三因子同比序列拟合值与拟合值的增加值

根据上文对于机器学习原理的简单介绍，在确定所用的机器学习模型之前，我们首先需要根据问题选择合适的特征变量 X 和因变量 Y 。由于与资产表现直接相关的变量为资产的收益率，因变量 Y 的选择范围较窄，为资产的收益率指标或者该指标衍生出的连续或离散类别变量，后文会对因变量 Y 的选择进行详细分析。本节重点分析并确定特征变量 X 的选择。

由于我们分析的对象为资产收益率，即价格环比，因此直觉上最理想的方式是直接提取资产环比序列的三周期状态作为特征变量 X 。然而，研究发现由于环比序列本身容易受到短期噪声的冲击，其波动中包含了大量不可解释的部分，通过高斯滤波获取的三周期状态稳定性较差，据此得到的三个周期状态对其自身外推预测的解释力也较低，难以预测资产未来短期表现。因此，本文认为根据资产环比序列得到的资产周期状态的估计值不适宜作为特征变量，具体分析详见附录一。

另一方面，回顾华泰金工深度报告的研究结果，在资产配置策略的构建中，多数使用当期的基于周期三因子同比序列拟合值的增加值（即当期同比序列拟合值减去上期同比序列拟合值）作为资产未来表现的评价指标，增加值越大则认为资产未来表现越好。且根据资产当期同比增加值大小确定资产配置比例的策略取得了较好的回测效果，我们认为同比序列周期状态适宜作为把握资产未来短期收益的“先行指标”。而同比增加值可以作为描述周期状态的特征变量。

在先前的研究中我们在多数情况下仅考虑了当期的同比增加值，然而，本文认为当期同比增加值只是内在逻辑的一种局部反映，不能够从整体角度把握资产周期状态与未来表现间的内在逻辑。

具体而言，利用周期三因子定价模型生成的滞后期和领先期同比拟合值的增加值同样与资产未来短期收益率之间存在相关关系。下图表以沪深 300 指数为例，展示了不同期同比增加值与指数当月(t)，未来第 1 个月($t+1$)和未来第 2 个月($t+2$)收益率之间的相关系数。 t 月领先 l 期同比增加值定义为 $t+l$ 月同比序列的周期三因子拟合值与 $t+l-1$ 月同比序列的周期三因子拟合值的差值：

$$\begin{aligned}\Delta \widehat{TB}_{t+l} &= \widehat{TB}_{t+l|t} - \widehat{TB}_{t+l-1|t} \\ &= \beta_{i,1}(x_{t+l|t}^{42} - x_{t+l-1|t}^{42}) + \beta_{i,2}(x_{t+l|t}^{100} - x_{t+l-1|t}^{100}) + \beta_3(x_{t+l|t}^{200} - x_{t+l-1|t}^{200})\end{aligned}$$

这里， $x_{t+l|t}^{42}$, $x_{t+l|t}^{100}$, $x_{t+l|t}^{200}$ 的含义与前文类似，是根据 t 时刻周期三因子序列外推 l 期的延拓值，

$\beta_1, \beta_2, \beta_3$ 为 t 时刻周期三因子定价方程参数的估计值，本文利用历史 50 个月的数据对于定价方程参数进行估计。

下表给出了沪深 300 指数不同期同比增加值同当月、未来第一个月和未来第二个月收益率相关关系汇总。可以看到，不同领先滞后期的同比增加值序列对于沪深 300 指数当月收益率、未来第 1 个月收益率和未来第 2 个月收益率的解释力度不同，且存在较大差异。例如，对于当月收益率，当期同比增加值与其相关系数最大，达到了 0.1369，也即当期同比增加值能够解释当月收益率波动的 13% 左右，相关系数其次为领先 1 期和领先 2 期同比增加值。对于未来第 1 个月和未来第 2 个月收益率，与之相关性最高的同比增加值指标分别为滞后 2 期和领先 2 期指标。值得关注的是，对于未来每期收益率，除了其对应期同比增加值与之存在较高相关性的同时，仍存在其它期同比增加值同样能较好的对收益率的运行趋势进行解释，因此，本文认为，如若想进一步把握资产收益率未来的走势，需要在模型中同时考虑多期同比增加值。

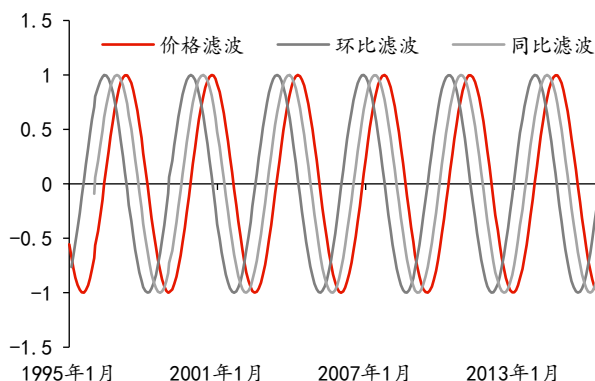
图表7： 沪深 300 不同期同比增加值同当月、未来第一个月和未来第二个月收益率相关关系

	当月收益率	未来第一个月收益率	未来第二个月收益率
ΔTB_t 当期同比增加值	0.1369	0.0230	0.0672
ΔTB_{t-1} 滞后 1 月同比增加值	0.0308	0.0424	0.0392
ΔTB_{t-2} 滞后 2 月同比增加值	0.0348	0.0589	0.0362
ΔTB_{t-3} 滞后 3 月同比增加值	0.0643	0.0518	0.0558
ΔTB_{t-4} 滞后 4 月同比增加值	0.0554	0.0440	0.0075
ΔTB_{t-5} 滞后 5 月同比增加值	0.0425	0.0031	0.0374
ΔTB_{t+1} 领先 1 月同比增加值	0.1255	0.0575	0.0656
ΔTB_{t+2} 领先 2 月同比增加值	0.1036	0.0108	0.0907
ΔTB_{t+3} 领先 3 月同比增加值	0.0912	0.0139	0.0848

资料来源：Wind，华泰证券研究所

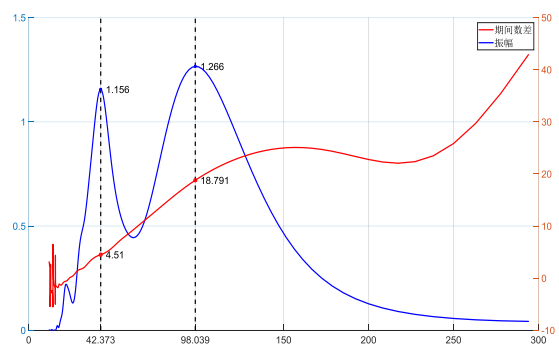
另一方面，上表结果显示，对于沪深 300 指数，不同期同比增加值同当月和未来月收益率之间的相关系数较低，最大相关度远小于 1，仅有 0.1369。这说明，仅仅依赖同比增加值这一个指标无法准确的把握资产未来走势。根据先前研究结果，资产价格衍生序列中同比序列与对数价格序列的各个周期分量之间存在稳定的领先滞后关系。例如，对于基钦周期，同比序列的 42 个月基钦周期分量领先于对数价格序列的该周期分量约 4 至 6 个月；对于朱格拉周期和库兹涅茨周期，同比序列的 100 个月和 200 个月周期分量与对数价格序列相应的周期分量间同样存在稳定的相位关系。

图表8： 上证综指同比、环比、价格序列的 42 个月高斯滤波



资料来源：《市场拐点的判断方法》，华泰证券研究所

图表9： 上证综指同比、价格序列振幅、相位谱



资料来源：《市场拐点的判断方法》，华泰证券研究所

根据华泰金工周期系列研究之《市场拐点的判断方法》，同比序列 42 个月高斯滤波得到的正弦分量的高点的出现时刻总落后于价格序列正弦分量的高点，如上图表中左图给出了上证综指同比序列和价格序列的 42 个月高斯滤波结果。此外，通过交叉谱分析的方式（右图），我们可以精确的得到上证综指的同比序列与价格序列的基钦周期分量间的相位差约

为 4.51 个月。因此，本文认为通过周期三因子定价模型构造的同比序列的领先期和滞后期的拟合值同样能够对预测资产未来的短期表现起到指导作用，是资产周期状态与未来表现间内在逻辑的另一个局部反映。

综上所述，本文判断当期同比增加值仅是周期状态同资产表现间内在逻辑的一种局部反映，其它领先滞后期同比增加值和各期同比序列的拟合值同样能部分的反映资产未来短期的表现。因此，我们初步选择各期同比增加值和同比序列拟合值作为全面反映资产周期状态的特征变量，特征变量的具体选择我们将在后文实证部分进一步分析。

短期收益率预测：概率结果比确定性结果更有效

本节我们主要分析资产表现预测问题中因变量 Y 的选择。由于资产未来收益率同资产表现直接相关，这里我们只需要确定因变量 Y 的具体形式： Y 为连续型变量，例如收益率值，则模型利用特征变量 X 对于 Y 的取值进行估计，给出确定性的预期结果；或 Y 为离散的类别变量，例如不同资产收益率的相对排序，则模型给出给定特征变量 X 的条件下 Y 属于某一特定类别的概率预测结果。我们将论证，在资产收益率分布有偏的前提下，概率结果比确定性结果更为有效。

确定性预测能够对未来可能发生事件的给出确定性预期。具体而言，假设 t 时刻有可能发生若干个事件，确定性预测将综合目前现有信息，对于 t 时刻发生第几个事件作出判断。传统的线性回归方法能够给出确定性预期结果。以未来资产表现趋势判断为例，线性回归利用资产收益率 Y 在外生条件 X 下的条件均值代表了收益率 Y 在未来时刻的期望值，并通过比较不同资产未来时刻的预期值给出资产的未来表现排序，进而给出未来表现最好资产的判断。

概率预测旨在根据现有信息，对未来特定时刻每个事件的发生概率进行估计。具体而言，对于资产未来表现问题，即根据现有外生条件 X ，给出资产收益率 Y 未来大小情况的概率分布。

对于前文基于同比序列周期状态的资产配置策略，其本质是将当期基于周期三因子的同比序列拟合值的增加值作为未来环比序列期望的估计值，并根据资产未来一期环比序列期望估计值的大小进行排序，所给出的资产表现结果本质为确定性预测结果。如果我们认为资产收益率是服从正态分布的，那么环比序列的期望值是资产未来表现的良好衡量指标，即资产收益率期望越大，则下期越有可能表现最好，那么此时确定性预测结果足以准确反映资产未来的预期表现。

然而，实际中资产收益率往往不服从标准的正态分布，其真实分布大多数情况下是非对称的，而在很多情况下，资产收益率分布的不对称性会导致仅仅依靠收益率期望值的大小对资产未来表现进行排序是不合理的，具体的分析详见附录二。

本研究对于 17 个全球主要资产指数的收益率分布的偏度进行了进一步的分析，结果如下表所示。其中，针对每个资产指数本文对于该指数收益率的偏度分布进行了估计，并在图表中表明了该估计分布的上 97.5% 和下 2.5% 分位点对应的区间。若 0 在该区间之外，说明在 95% 的置信度下该指数收益率分布的偏度显著异于 0，即该指数收益率分布存在不对称性。可以看到，进行分析的全部 17 个资产指数中 15 个指数的偏度显著不为 0，其中泛欧总回报指数收益率分布的偏度显著大于 0，而上证综指、标普 500、富时 100 等 14 个指数收益率分布的偏度显著小于 0，仅有美国总回报指数和 CRB 食品现货指数收益率的分布没有表现出明显的不对称性。

图表10： 全球主要资产指数收益率分布偏度分析

	偏度	偏度 2.5%分位点	偏度 97.5%分位点	是否显著
上证综指	-0.604	-0.797	-0.393	是
标普 500	-1.039	-1.364	-0.674	是
富时 100	-0.700	-0.892	-0.496	是
法国 CAC40	-0.643	-0.795	-0.481	是
德国 DAX	-0.870	-1.139	-0.570	是
日经 225	-0.984	-1.321	-0.577	是
恒生指数	-0.738	-1.006	-0.452	是
美国总回报指数	-0.156	-0.431	0.117	否
泛欧总回报指数	1.325	0.820	1.736	是
中国总回报指数	-0.562	-0.856	-0.230	是
CRB 食品现货	-0.181	-0.443	0.083	否
CRB 食用油现货	-0.402	-0.701	-0.066	是
CRB 家禽现货	-0.589	-0.876	-0.264	是
CRB 金属现货	-1.028	-1.520	-0.361	是
CRB 工业现货	-1.437	-2.053	-0.559	是
CRB 综合现货	-1.120	-1.756	-0.245	是
CRB 纺织现货	-0.219	-0.384	-0.046	是

资料来源：Wind, Bloomberg, 华泰证券研究所

因此，我们可以在一定程度上断定，对于绝大多数资产指数，其收益率分布大概率存在不对称的现象，则根据上文分析，基于确定性预期结果的资产表现判断可能出现偏差，概率预期结果比确定性预期结果更适合资产短期收益率预测。

综上所述，本文考虑根据现有信息，对资产未来表现的概率预期结果进行估计，即根据 t 时刻 N 项资产的已知信息，给出 $t+1$ 时刻每项资产收益率取得最好表现的概率估计：

$$P\left(\max_{k=1,2,\dots,N}\{r_k\}=r_i\right), i=1,2,\dots,N$$

机器学习：通过挖掘资产周期规律与其市场表现的内在逻辑预测其收益排序

Softmax 回归（或者叫做多项逻辑回归）是逻辑回归（Logistic）在多分类问题上的推广。在逻辑回归中，其本质是利用逻辑函数建立起外生因变量与类别之间的关联关系，并对具体类别的发生概率进行估计。但由于逻辑函数形式的限制，逻辑回归能够处理的类别个数仅为 2，而在实际应用中，我们常常需要同时分析多类资产的未来表现。**Softmax 回归**将逻辑函数推广为 **Softmax 函数**，它能将一个含任意实数的 K 维实数空间的向量“压缩”到空间 $[0,1]^K$ 中，使每一个元素的范围都在 $[0,1]$ 之间，并且所有元素的和为 1，从而使 **Softmax 回归**能够对多维的分类问题进行建模。

基于 **Softmax 函数**的多维性质，**Softmax 回归**能够处理类别数目多于 2 的情况，且同样能够给出具体类别的发生概率的估计值，在多项线性判别分析，朴素贝叶斯分类器和人工神经网络等的多种基于概率的多分类问题方法中都有着广泛应用。

在本文讨论的资产未来表现预期的问题中，**Softmax 模型**能够建立下一个时刻资产 i 在 N 项资产中表现最好的概率同自变量 g （即基于资产三周期当前状态构造的指标）的相关关系，具体形式如下：

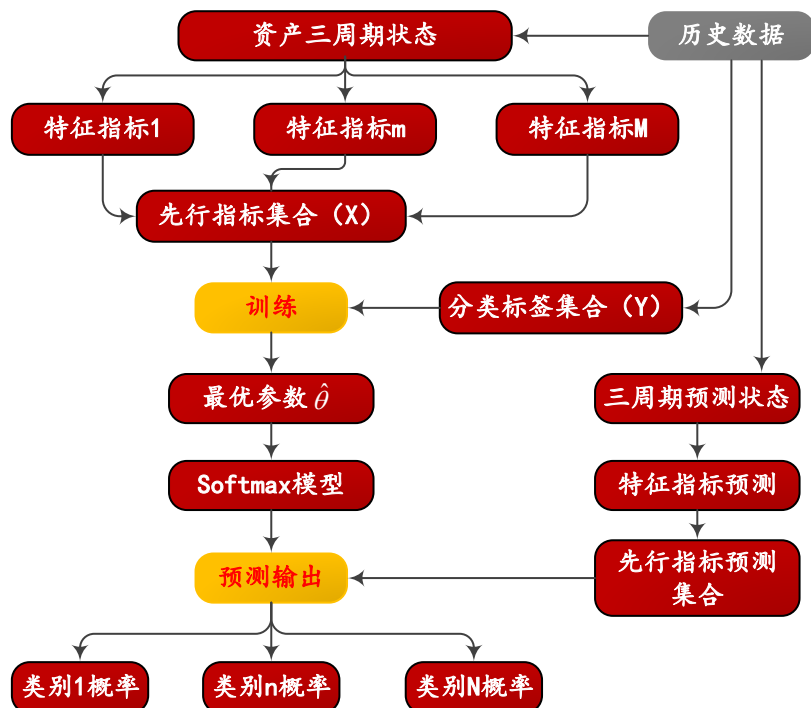
$$\begin{bmatrix} P(y_{t+1}^1=1|\theta) \\ P(y_{t+1}^2=2|\theta) \\ \vdots \\ P(y_{t+1}^N=N|\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^N e^{\theta_j^T g_t}} \begin{bmatrix} e^{\theta_1^T g_t} \\ e^{\theta_2^T g_t} \\ \vdots \\ e^{\theta_N^T g_t} \end{bmatrix}$$

其中 $P(y_{t+1}=k|\theta,t)$ 表示根据 t 时刻的信息，对未来 $t+1$ 时刻因变量属于第 k 类（即 $y_{t+1}=k$ ）概率的估计。**Softmax 方法**正是建立起特征变量 g 同因变量 y 属于某具体类别的概率间的

相关关系。 $\theta = [\theta_1, \theta_2, \dots, \theta_N]$ 为模型待估参数，同线性回归中回归系数 β 的类似，可以通过历史数据进行估计。本文所用 Softmax 模型的具体形式以及估计方式详见附录三。

需要特别说明的是，在 Softmax 回归中，特征变量 $g_t^1, g_t^2, \dots, g_t^N$ 既可以通过将 Softmax 模型同神经网络结合，利用神经网络提取特征的方式得到；也可以根据历史经验积累，人工选择同分类结果具有较强相关关系的先行指标作为特征输入。由于本文回测中采用大类资产的月度数据，因此数据规模有限，很难训练出复杂的特征提取神经网络；同时，根据华泰金工周期系列的研究成果，我们已经验证了很多指标与资产未来的表现存在一定的相关性，例如基于周期三因子模型得到的同比拟合值的增加值、与同比序列领先期和滞后期拟合值等等。因此我们选择在模型中放弃复杂的特征提取层，直接将先行指标作为 Softmax 模型的输入。

图表11： Softmax 回归方法示意图



资料来源:华泰证券研究所

集成学习：避免对参数选择的敏感性，更全面有效的利用历史规律

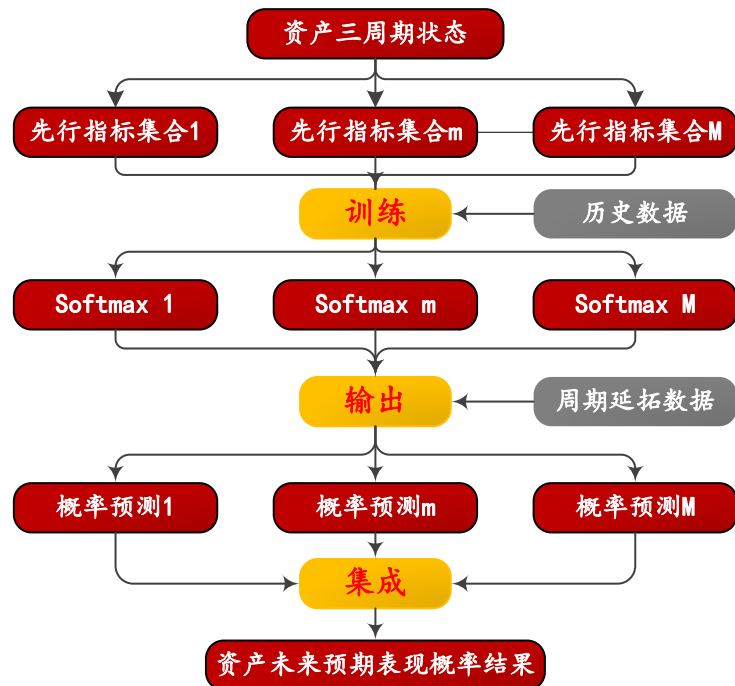
根据上文所述，本文利用 Softmax 模型，建立起了不同资产周期三因子状态同其未来短期表现间的相关关系。然而，根据不同先行指标组合所训练出的 Softmax 模型的效果可能存在差异，换句话说，Softmax 模型对资产未来表现判断的准确性依赖于先行指标组合的选择。后文实证部分的结果可以验证该现象：例如，对全球股债细分资产指数，根据领先 1 期、滞后 1 期和滞后 5 期同比增加值和同比拟合值指标训练出的 Softmax 模型效果最好；但根据领先 1 期、当期和滞后 5 期指标训练出的模型结果最差，其对应的资产组合收益与基准的等权策略相当；相反，对全球大类资产等权指数，根据领先 1 期、当期和滞后 5 期指标训练出的 Softmax 取得了最优的表现。

由于所用指数数据长度较短，因此我们很难将数据分成数据量同时满足指标集合选取和回测精度要求的两部分。如果我们不对数据进行划分，根据资产回测的精度确定该资产最优的先行指标组合，则会造成 Softmax 模型的过拟合。而 Softmax 模型的过拟合将使得其在训练数据集外不能很好的拟合数据，即对未来的数据缺少适应性，使得模型预测结果的置信度下降。

因此，为了避免上述问题，本文引入集成学习的方法，从而减少 Softmax 模型的预测结果对先行指标集合选择的依赖，降低过拟合出现的可能。

本文采用的集成学习方法为 Bootstrap aggregating（简称 Bagging）。随机森林算法（Random Forest）就是一种典型的 Bagging 集成学习方法，该算法正是通过 Bootstrap 方法对原始数据集进行采样，并根据采样集训练并得到彼此存在差异的决策树；最终的结果是利用“少数服从多数”的原则对每个决策树给出的分类结果进行集成。

图表12：集成学习方法流程图



资料来源:华泰证券研究所

本文所用的 Softmax-集成学习方法的流程图如上图所示。首先，对于确定的主要特征指标范围，我们依次选择可能的指标组合构建先行指标集合，并利用历史数据训练不同依赖于先行指标集合的单一 Softmax 模型。在模型集成阶段，本文利用加权平均的方式对每个 Softmax 模型给出的概率预测结果进行汇总，其中，集成中对每个模型结果赋予的权重相同，均为 $1/M$ 。可以看到，该预测结果不依赖于先行指标集合的选择，下文实证部分将对集成学习方法的有效性做进一步分析和验证。

实证分析：周期理论与机器学习的研究方法具有对不同市场的普适性

根据上文所述，本研究拟利用已知的与资产收益率相关的先行指标作为 Softmax 模型的输入，以代替复杂的特征提取网络。通过分析和比较 15 对共 30 个主要特征指标对 Softmax 模型回测准确度的影响，本文将先行指标的范围确定为：当期同比增加值与同比拟合值、滞后 1 期同比增加值与同比拟合值、滞后 5 期同比增加值与同比拟合值和领先 1 期同比增加值与同比拟合值。

进一步，本文利用仿真生成的模拟数据，通过分析 Softmax 模型在仿真数据可预测方差比例和分类准确率的相关关系，验证了 Softmax 模型能够有效的挖掘资产三周期状态同其未来预期表现间的内在逻辑。同时，本文对全球大类资产等权指数、全球股债资产指数与中国大类资产指数分别进行了分析和基于 Softmax 模型回测结果的资产配置，利用市场历史数据的回测结果验证了所提模型在真实市场环境中把握多类资产未来表现的有效性。

实证部分的最后，我们利用周期状态的外推延拓性质，对各类资产未来的周期状态进行预测；并利用预期的三周期状态构建先行指标集合作为 Softmax 模型的输入，根据模型输出的概率结果对各类资产未来的表现进行预期和判断。Softmax 模型结果显示，2018 年 5 月至 12 月，债券资产预期表现占优，需谨防风险资产下行。

预测先行指标的选择：“特征变量”组合的方式与原因分析

华泰金工周期系列研究结果显示，全球主要经济金融指数存在着 42 个月、100 个月和 200 个月左右的周期因子，指数衍生序列的部分波动可由其内在的周期状态解释。于此同时，通过上文的分析，我们发现基于资产周期三因子定价模型计算而得的各期同比增加值对资产未来短期表现均存在一定的解释力度；另一方面，根据华泰金工周期系列报告《市场拐点的判断方法》，资产的同比序列和价格序列的基钦周期分量间存在稳定的领先滞后关系，同比序列的基钦周期分类领先价格序列约 4 至 6 个月，因此，我们认为根据周期三因子得到的领先和滞后期同比序列拟合值同样对资产收益率存在一定解释力度。

综上所述，本文将主要特征指标的范围初步确定为当期同比增加值与同比拟合值、领先 1 期至领先 6 期同比增加值与同比拟合值、滞后 1 期至滞后 8 期同比增加值与同比拟合值 15 对共 30 个指标：

$$(\widehat{TB}_{i,t-8}, \Delta \widehat{TB}_{i,t-8}), \dots, (\widehat{TB}_{i,t-1}, \Delta \widehat{TB}_{i,t-1}), (\widehat{TB}_{i,t}, \Delta \widehat{TB}_{i,t}),$$

$$(\widehat{TB}_{i,t+1}, \Delta \widehat{TB}_{i,t+1}), \dots, (\widehat{TB}_{i,t+6}, \Delta \widehat{TB}_{i,t+6})$$

其中 $\widehat{TB}_{i,t+l}$ 表示领先 l 期的同比序列拟合值，在本研究中，资产 t 时刻的同比拟合值根据同比序列周期三因子定价模型拟合并外推延拓得到，模型中参数根据资产历史 50 个月同比序列值估计； $\Delta \widehat{TB}_{i,t+l} = \widehat{TB}_{i,t+l} - \widehat{TB}_{i,t+l-1}$ 表示领先 l 期的同比增加值，由相邻两期同比拟合值做差计算而得。

对每一对主要特征指标，将其作为 Softmax 模型的输入，并利用动态回测的方式分析该 Softmax 模型的准确率（即每一期预测结果与真实结果间的差异），选择回测准确率最高的若干个主要指标对构成 Softmax 模型的先行指标集合。其中，本文采用的回测准确率评价指标为 Top1Hit, Top2Hit 和 Top3Hit，分别表示回测中预测中表现最好的资产实际排名第一、第二和第三的比例。上述三个指标的取值范围为 0 至 1（100%），取值越大表示模型准确度越高，预测能力越强；对于理想模型，上述三指标的取值均为 1。

具体先行指标集合选取步骤：

- 步骤一：初始化 Softmax 模型，初始化每个主要指标准确率评价；
- 步骤二：时刻 t 对每对主要特征指标 $(\widehat{TB}_{i,t+l}, \Delta \widehat{TB}_{i,t+l})$ 计算其历史 90 期的估计值；
- 步骤三：利用每对主要特征指标的历史 90 期数据训练 Softmax 模型；
- 步骤四：利用每个 Softmax 模型对 t 时刻资产的表现进行预期，并于对比资产的实际表现，

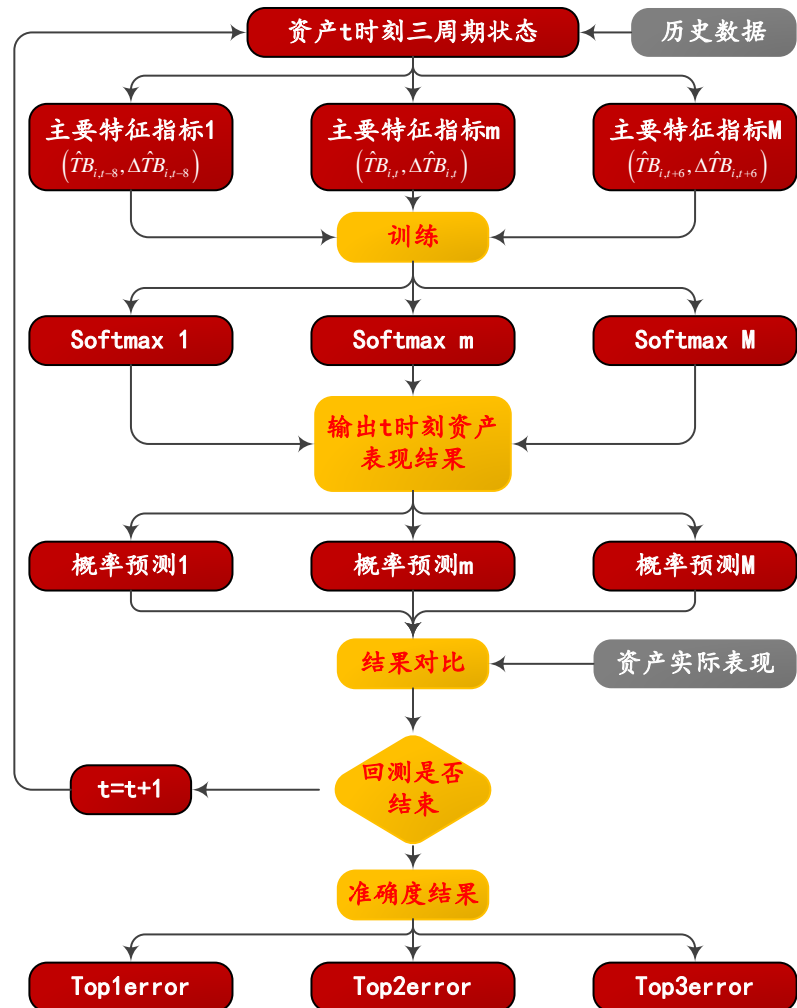
进而更新准确率评价指标 Top1Hit, Top2Hit 和 Top3Hit;

步骤五: 判断回测是否结束, 如未结束, 则令 $t = t + 1$, 返回步骤二; 否则, 进入步骤六;

步骤六: 输出每个主要特征指标的准确率评价指标 Top1Hit, Top2Hit 和 Top3Hit。

先行指标集合选择的流程图如下图所示:

图表13: 先行指标选择流程图



资料来源: 华泰证券研究所

值得一提的是, 为了避免后文大类资产回测和资产配置分析中存在使用未来数据的可能¹, 本文中, 先行指标集合确定中所使用的数据为中国股票市场的 6 大板块指数, 该指数根据申万一级行业数据等权合成。根据不同的主要特征指标在板块数据回测中的准确率表现确定先行指标集合, 进而在全球大类资产等权指数、全球股债细分资产和中国大类资产中进行回测与资产配置表现分析。利用板块数据确定先行指标集合的原因在于如下两个方面: 首先, 我们认为先行指标的选取规律在于资产衍生序列 (同比、环比、对数价格序列) 内在的数理逻辑, 其本质源于资产衍生序列互相之间的相关关系, 而与序列来源于何种资产无关。因此, 我们能够从某类特定资产中挖掘到较优的先行指标集合, 且在其它资产中表现稳定。另一方面, 由于本文所用数据为月度, 数据长度较短, 无法将数据拆分为长度同时满足参数选取和回测的两部分。下文实证结果表明, 根据板块数据选取的先行指标集合在其它三类资产回测中表现同样优异, 为上述分析提供了有力的支持。

¹ 如果先行指标的确定与资产配置回测使用同一组数据, 由于先行指标是根据全体历史数据选取的, 那么回测中即使模型参数由历史数据估计, 单模型的输入仍来源于全体数据, 即存在利用未来数据进行回测的可能

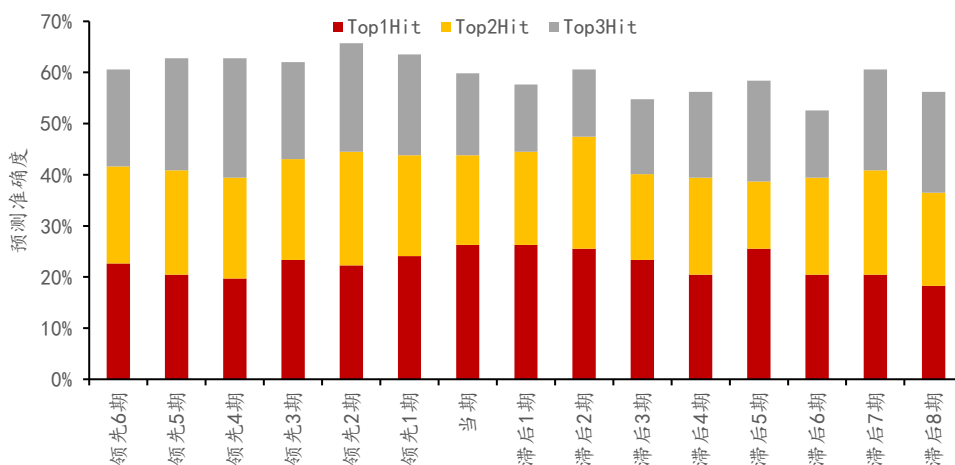
下图展示了基于上述 15 对主要特征指标构建的 Softmax 模型针对板块等权指数的回测准确率表现。可以看到，总体而言领先 1 期、当期、滞后 1 期和滞后 5 期同比增加与同比拟合值的回测准确率最高，Top1Hit 均超过了 24%，高于基准的 16.67%（随机选择 6 项资产中一项，恰好为表现最好资产的概率），Top2Hit 和 Top3Hit 也均在一定程度上高于基准准确率。其中，当期和滞后 1 期主要特征指标的 Top1Hit 最高，达到了 26.28%；领先 1 期特征指标的 Top3Hit 最高，达到 63.50%；滞后 5 期的准确率较为平均，但基于同比序列同价格序列的基钦周期分量间存在稳定的 4.5 个月相位差的结论，本文认为滞后 5 期的同比序列拟合值能够在一定程度上反映资产未来表现，因此将该特征指标纳入先行指标集合中。

图表14：主要特征指标回测准确率汇总

	Top1Hit	Top2Hit	Top3Hit
领先 6 期	22.63%	41.61%	60.58%
领先 5 期	20.44%	40.88%	62.77%
领先 4 期	19.71%	39.42%	62.77%
领先 3 期	23.36%	43.07%	62.04%
领先 2 期	22.28%	44.53%	65.69%
领先 1 期	24.09%	43.80%	63.50%
当期	26.28%	43.80%	59.85%
滞后 1 期	26.28%	44.53%	57.66%
滞后 2 期	25.55%	47.45%	60.58%
滞后 3 期	23.36%	40.15%	54.74%
滞后 4 期	20.44%	39.42%	56.20%
滞后 5 期	25.55%	38.69%	58.39%
滞后 6 期	20.44%	39.42%	52.55%
滞后 7 期	20.44%	40.88%	60.58%
滞后 8 期	18.25%	36.50%	56.20%

资料来源：Wind，华泰证券研究所

图表15：主要特征指标回测准确率柱状图



资料来源：Wind，华泰证券研究所

综上所述，本文将 Softmax 模型的先行指标的范围确定为：当期同比增加与同比拟合值、滞后 1 期同比增加与同比拟合值、滞后 5 期同比增加与同比拟合值和领先 1 期同比增加与同比拟合值。

仿真测试：机器学习在挖掘资产周期状态与未来表现逻辑关系具有有效性

周期三因子定价模型的基础来源于资产服从周期几何布朗运动模型的假设。资产价格的几何布朗运动模型由保罗·萨缪尔森在 1965 年提出，在该模型中，资产价格的变动可以分解为固定趋势的漂移项和固定波动的布朗运动之和。周期几何布朗运动模型是对基础模型的改进，其放松了几何布朗运动模型中漂移趋势固定不变假设，认为该趋势会随时间增长

周期性波动：

$$dS = \left[\mu + \sum_{p=1}^P A_p \omega_p \cos(\omega_p t + \phi_p) \right] S dt + \sigma S dB$$

根据伊藤引理，资产 t 时刻的收益率可以表示为：

$$r_t = \ln \frac{S_t}{S_{t-1}} = \left(\mu - \frac{1}{2} \sigma^2 \right) + \sum_{p=1}^P \int_{t-1}^t A_p \omega_p \cos(\omega_p t + \phi_p) dt + \sigma \varepsilon_t$$

其中 ε_t 服从标准正态分布。可以看到，在周期几何布朗运动模型的假设下，资产 t 时刻的表现可以分解为三个部分：常数项、周期项和随机波动项，且由于常数项对于方差的计算不产生影响，因此资产收益率波动中可预期部分完全取决于资产所处的周期状态。

本文所提出的基于周期三因子模型和 **Softmax** 的资产收益表现预测方法本质是挖掘资产三周期状态同资产未来预期表现间的内在逻辑，如果随着资产收益率方差中可被周期三因子解释部分的比例的提升，**Softmax** 模型预测准确率有显著的提升，则认为 **Softmax** 方法能够有效的挖掘资产周期状态与未来表现间的内在逻辑。

因此，本研究利用仿真方法在不同资产收益率可解释部分比例下生成若干条价格轨迹数据，并测试 **Softmax** 在该仿真数据下的回测准确率表现，进而分析该方法对于挖掘资产周期状态与未来表现间逻辑的有效性。根据周期几何布朗运动模型，可解释方差比例同随机项误差项 σ 的大小反向相关，因此本研究通过调整 σ 的大小来调整可解释方差比例。

图表16：不同随机波动项 σ 对应可解释方差比例

随机波动项	0.001	0.003	0.005	0.01	0.02	0.03
可解释方差比例	98.90%	90.94%	79.02%	49.21%	18.97%	17.29%

资料来源：华泰证券研究所

仿真测试中周期几何布朗运动模型所采用的具体参数如下：

- 资产数目： $N = 7$
- 漂移趋势： $\mu = 0.001$
- 周期数目： $P = 3$
- 周期频率： $\omega_1 = 2\pi/42, \omega_2 = 2\pi/100, \omega_3 = 2\pi/200$
- 周期振幅： $A_1 = A_2 = A_3 = 0.1$
- 周期相位： ϕ_1, ϕ_2, ϕ_3 根据均匀分布随机采样得到
- 随机波动项： $\sigma = 0.001, 0.003, 0.005, 0.01, 0.02, 0.03$

对于每个随机波动项 σ ，随机生成 10 组资产价格集合，每组资产价格集合包括 7 只资产长度为 400 的价格序列，每只资产的价格序列根据周期几何布朗运动模型独立生成，模型参数如上文所示。

基于上文结果，先行指标集合包括 4 对指标：当期同比增加值与同比拟合值、滞后 1 期同比增加值与同比拟合值、滞后 5 期同比增加值与同比拟合值和领先 1 期同比增加值与同比拟合值。本文首先依次选择 3 对指标作为 **Softmax** 模型的输入，从而构建了 4 个基于不同先行指标的模型，利用仿真生成的数据测试不同模型在不同可解释方差比例下的回测准确率结果，得到结果如下表所示：

图表17： 单一 Softmax 模型仿真测试 Top1Hit 准确率汇总

随机项方差	可解释方差比例	Softmax-[-5 -1 0]	Softmax-[-5 -1 1]	Softmax-[-5 0 1]	Softmax-[-1 0 1]
0.001	98.90%	74.61%	58.12%	58.18%	55.64%
0.003	90.94%	66.91%	59.76%	59.39%	58.42%
0.005	79.02%	57.58%	50.48%	50.48%	50.61%
0.01	49.21%	36.85%	34.48%	34.06%	32.85%
0.02	18.97%	21.82%	20.79%	20.79%	20.42%
0.03	17.29%	19.33%	18.85%	17.82%	17.70%

资料来源：华泰证券研究所

图表18： 单一 Softmax 模型仿真测试 Top2Hit 准确率汇总

随机项方差	可解释方差比例	Softmax-[-5 -1 0]	Softmax-[-5 -1 1]	Softmax-[-5 0 1]	Softmax-[-1 0 1]
0.001	98.90%	92.61%	91.27%	91.39%	91.03%
0.003	90.94%	87.58%	83.58%	83.64%	83.82%
0.005	79.02%	81.33%	76.24%	76.42%	77.21%
0.01	49.21%	63.52%	59.94%	60.12%	59.27%
0.02	18.97%	41.03%	39.03%	39.58%	38.97%
0.03	17.29%	38.55%	37.76%	37.27%	37.09%

资料来源：华泰证券研究所

图表19： 单一 Softmax 模型仿真测试 Top3Hit 准确率汇总

随机项方差	可解释方差比例	Softmax-[-5 -1 0]	Softmax-[-5 -1 1]	Softmax-[-5 0 1]	Softmax-[-1 0 1]
0.001	98.90%	95.94%	96.12%	96.18%	96.12%
0.003	90.94%	95.33%	94.18%	94.42%	94.67%
0.005	79.02%	93.39%	90.06%	90.24%	90.55%
0.01	49.21%	80.91%	77.21%	77.58%	77.39%
0.02	18.97%	56.42%	54.73%	54.30%	53.94%
0.03	17.29%	55.15%	54.12%	52.61%	52.55%

资料来源：华泰证券研究所

其中，Softmax-[-5 -1 0]表示输入先行指标为滞后 5 期、滞后 1 期和当期同比增加值和同比拟合值的 Softmax 模型，其余同理。可以看到，对上述 4 个单一 Softmax 模型，随着可解释方差比例的上升，模型回测准确率也随之提升；当可解释方差达到 79.02%时，模型的 Top1Hit 均超过了 50%，Top3Hit 超过了 90%，进一步当可解释方差比例大于 90%时，上述模型的 Top3Hit 均达到 95%以上，表现最好的模型 Softmax-[-5 -1 0]预测结果实际排名第一的百分比也达到约 75%。由于可解释方差比例完全由资产的周期三因子决定，因此单一 Softmax 模型回测准确率验证了模型对挖掘资产周期状态与未来表现间逻辑的有效性。

进一步，本文同样对集成学习模型进行了仿真测试，从而分析集成的 Softmax 模型对于挖掘内在逻辑的有效性。集成学习模型仿真测试结果汇总如下表所示：

图表20： 集成学习模型仿真准确率汇总

随机项方差	可解释方差比例	Top1Hit	Top2Hit	Top3Hit
0.001	98.90%	61.64%	92.12%	95.94%
0.003	90.94%	61.82%	85.58%	95.33%
0.005	79.02%	52.42%	78.55%	91.39%
0.01	49.21%	34.73%	61.39%	78.85%
0.02	18.97%	20.91%	39.33%	55.09%
0.03	17.29%	18.61%	37.88%	53.64%

资料来源：华泰证券研究所

策略回测：基于机器学习预测结果的资产配置策略业绩优势明显

根据上文仿真测试结果，对于仿真数据，Softmax 模型能够有效挖掘资产三周期状态同未来表现间的内在联系，即利用 Softmax 模型能够给出资产未来表现排序的概率预测结果。

本节将就全球股债指数和中国大类资产指数的历史数据，采用基于周期三因子的 Softmax 方法进行回测，进而分析该方法对资产未来收益预测的有效性。

为了更好的展示回测效果，本文设计了基于 Softmax 模型输出结果的资产配置策略，通过比较基于 Softmax 结果的资产配置策略与基准策略净值曲线与风险收益指标的差异，分析和评价 Softmax 模型在预测资产未来短期表现的效果。具体而言，如果基于 Softmax 预测结果的资产组合的主要风险收益指标均优于基准策略，这说明该策略有更高的概率配置未来表现较好的资产，即 Softmax 模型能够有效的把握资产的未来短期表现。

基于 Softmax 模型输出结果的资产配置策略具体步骤如下：

- 动态预测：在每个时刻 t ，利用资产前 90 期历史数据构建 Softmax 模型，并对于 t 时刻的资产表现进行预测；
- 基于动态预测结果的资产配置：基于 t 时刻资产表现的 Softmax 预测结果，对于概率最高的资产下期配置比例为 100%，其余资产配置比例为 0。

本研究设计了单指标策略作为基于 Softmax 模型资产配置策略的对照，单指标策略的具体步骤如下：

- 动态预测：在每个时刻 t ，利用资产前 90 期历史数据构建周期三因子模型，并对 t 时刻资产领先期、当期和滞后期同比进行估计，据此计算各阶同比增加值；
- 基于同比增加值估计得资产配置：基于 t 时刻资产某期同比增加值大小确定资产的配置比例，具体而言，对于同比增加值最大的资产配置比例为 100%，其余资产配置比例为 0。

单指标策略的具体细节详见华泰金工周期系列深度报告《周期三因子定价与资产配置模型》，在此不再赘述。

在下文每一部分回测中，我们首先根据基于机器学习的资产配置策略回测结果，分析其与基准的等权配置策略和上述单指标策略在各项风险收益指标的表现差异，验证 Softmax 模型能够有效的把握周期状态和资产表现的内在逻辑；进而利用集成学习方法对单一 Softmax 模型的预测结果进行整合，分析集成学习-Softmax 模型的回测表现，从而验证上文观点：集成学习能够有效挖掘周期状态与资产表现关联关系的同时，降低模型对先行指标选取的依赖性。

全球主要股债细分资产指数实证结果

首先，我们对全球主要股债细分资产指数利用基于周期三因子的 Softmax 方法进行回测分析。本文所采用的全球主要股债指数包括 10 个细分资产指数，具体如下表所示。其中选取包括上证综指，标普 500 在内的 7 个主要股票市场指数作为股票指数代表，选取美国总回报指数 (U.S. aggregate total return value)，泛欧总回报指数 (Pan-European aggregate total return index value) 和中国总回报指数 (China aggregate Total Return index value) 作为全球债券指数代表。数据的起止时间为 2004 年 1 月至 2018 年 3 月，数据采集频率为月度。

图表21： 全球主要股债细分资产指数选择

类别名称	具体指数	指数代码	起止时间
股票	上证综指	000001.SH	2004-01 至 2018-03
	标普 500	SPX.GI	2004-01 至 2018-03
	富时 100	FTSE.GI	2004-01 至 2018-03
	法国 CAC40	FCHI.GI	2004-01 至 2018-03
	德国 DAX	GDAXI.GI	2004-01 至 2018-03
	日经 225	N225.GI	2004-01 至 2018-03
	恒生指数	HSI.HI	2004-01 至 2018-03
	美国总回报指数		2004-01 至 2018-03
债券	泛欧总回报指数		2004-01 至 2018-03
	中国总回报指数		2004-01 至 2018-03

资料来源：Wind, Bloomberg, 华泰证券研究所

根据上文结果，我们初步选择了滞后 5 期、滞后 1 期、当期和领先 1 期的同比增加值和同比拟合值共 4 对主要特征指标用于构建 Softmax 模型输入的“先行指标”集合。下文依次选择 3 对主要特征指标构建单一 Softmax 模型，共计 4 个 Softmax 模型用于全球主要股债指数资产配置回测。

回测的方式为动态预测，回测区间为 2012 年 5 月至 2018 年 3 月。在每个时刻 t ，利用历史 90 个月的数据训练 Softmax 模型，并根据模型的预测结果调整资产配置比例。单一 Softmax 模型的回测结果如下图表所示：

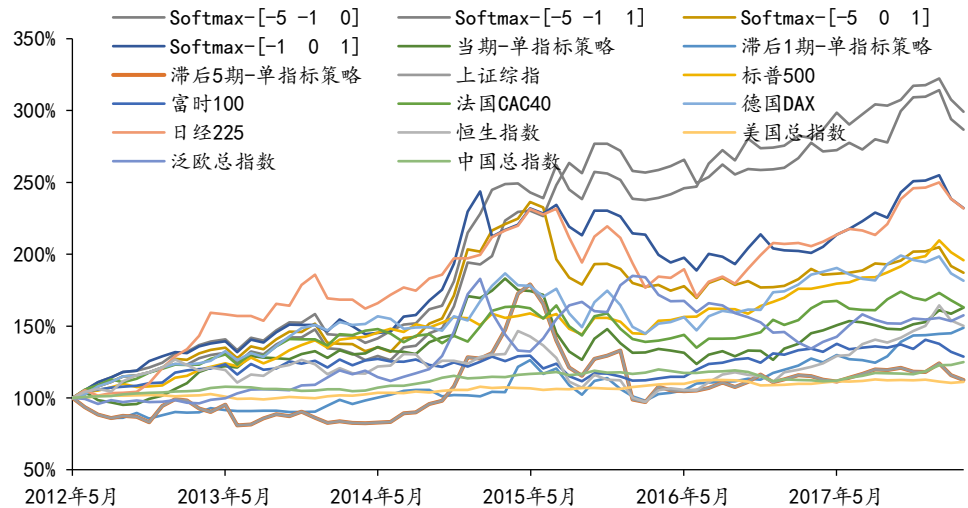
图表22： 单一 Softmax 模型全球股债细分资产配置回测风险收益指标汇总

	总收益率	年化收益率	波动率	最大回撤	夏普比率	信息比率
Softmax-[-5 -1 0]	286.81%	19.80%	0.152	14.52%	1.300	0.228
Softmax-[-5 -1 1]	299.23%	20.67%	0.146	12.62%	1.419	0.292
Softmax-[-5 0 1]	187.13%	11.34%	0.161	28.26%	0.706	0.041
Softmax-[-1 0 1]	232.12%	15.53%	0.162	22.63%	0.956	0.130
当期-单指标策略	162.90%	8.72%	0.161	32.46%	0.542	-0.022
滞后 1 期-单指标策略	148.80%	7.05%	0.138	22.46%	0.510	-0.063
滞后 5 期-单指标策略	112.54%	2.05%	0.240	45.82%	0.085	-0.091
等权配置策略	180.67%	10.67%	0.091	14.12%	1.174	
上证综指	112.54%	2.05%	0.240	45.82%	0.085	-0.091
标普 500	195.83%	12.21%	0.094	9.16%	1.293	0.052
富时 100	128.74%	4.43%	0.101	13.71%	0.439	-0.176
法国 CAC40	163.08%	8.75%	0.128	17.89%	0.681	-0.038
德国 DAX	181.58%	10.77%	0.143	22.49%	0.751	0.019
日经 225	231.84%	15.51%	0.161	26.41%	0.962	0.125
恒生指数	150.06%	7.21%	0.158	33.64%	0.455	-0.057
美国总回报指数	111.09%	1.82%	0.028	3.70%	0.652	-0.265
泛欧总回报指数	157.52%	8.10%	0.168	27.66%	0.483	-0.022
中国总回报指数	124.91%	3.89%	0.039	6.84%	0.992	-0.212

资料来源：Wind, Bloomberg, 华泰证券研究所

其中，Softmax-[-5 -1 0]表示构建该模型的先行指标集合为滞后 5 期、滞后 1 期和当期的同比增加值和同比拟合值 $\{(\bar{T}B_{i,t-5}, \Delta \bar{T}B_{i,t-5}), (\bar{T}B_{i,t-1}, \Delta \bar{T}B_{i,t-1}), (\bar{T}B_{i,t}, \Delta \bar{T}B_{i,t})\}$ 。当期-单指标策略表示依据资产当期同比增加值大小对确定资产配置比例的单指标策略，滞后 1 期、滞后 5 期-单指标策略同理。信息比率计算中的基准参照策略为等权配置策略。

可以看到，基于上述 4 个单一 Softmax 模型结果的资产配置策略的年化收益率均超过了基准的等权配置策略。其中，模型 Softmax-[-5 -1 1]的年化收益率最高，达到了 20.67%，夏普比率超过了基准的等权配置策略且最大回撤与基准策略相当。同时，上述 4 个单一 Softmax 模型的年化收益率、夏普比率和信息比率均超过了各个单指标策略，这说明单一 Softmax 模型在挖掘资产周期状态同未来表现间内在逻辑的有效性。

图表23： 基于不同先行指标的单一 Softmax 预测结果的全球主要股债细分资产配置净值曲线

资料来源：Wind, Bloomberg, 华泰证券研究所

然而，基于上述四个单一 Softmax 模型预测结果的资产配置策略的回测结果仍存在一定差异，例如，模型 Softmax-[-5 0 1]模型回测结果的最大回撤达到了 28.26%，为所有模型中最高，且其年化收益率仅为 11.34%，仅与基准的等权配置策略相当，甚至低于某些单一的细分资产，这说明该 Softmax 模型未能有效的对资产未来表现进行判断。

上述四个单一 Softmax 模型间的唯一差异在于先行指标集合的选取。为了降低 Softmax 模型结果对于先行指标选取的依赖，本研究采用了集成学习的方式，在每个时刻 t ，对于基于不同先行指标集合的单一 Softmax 模型给出的预测结果进行加总，进而输出集成后的预测结果；同时根据集成的资产表现的预测结果进行资产配置。集成学习-Softmax 模型的回测结果如下图表所示：

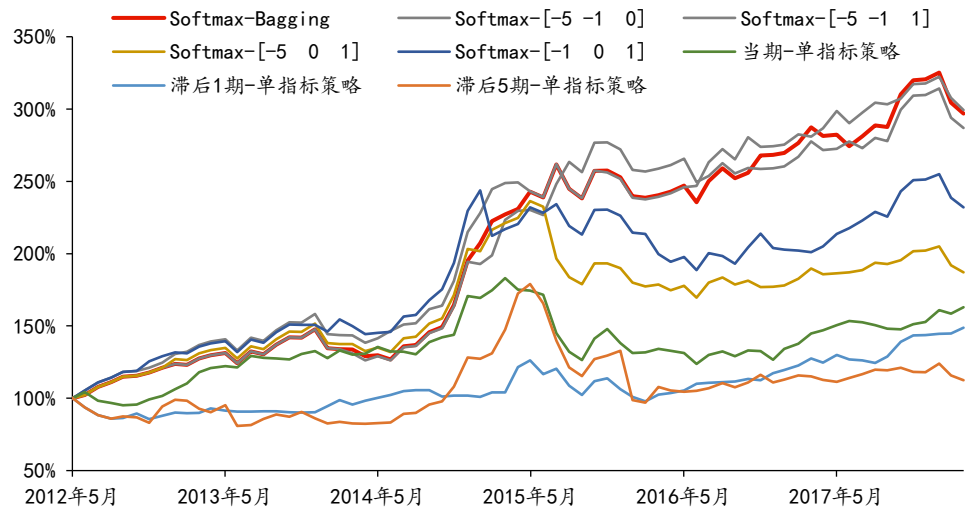
图表24： Softmax-集成学习模型全球股债细分资产配置回测风险收益指标汇总

	总收益率	年化收益率	波动率	最大回撤	夏普比率	信息比率
Softmax-Bagging	296.83%	20.50%	0.151	14.08%	1.355	0.291
Softmax-[-5 -1 0]	286.81%	19.80%	0.152	14.52%	1.300	0.228
Softmax-[-5 -1 1]	299.23%	20.67%	0.146	12.62%	1.419	0.292
Softmax-[-5 0 1]	187.13%	11.34%	0.161	28.26%	0.706	0.041
Softmax-[-1 0 1]	232.12%	15.53%	0.162	22.63%	0.956	0.130
当期-单指标策略	162.90%	8.72%	0.161	32.46%	0.542	-0.022
滞后 1 期-单指标策略	148.80%	7.05%	0.138	22.46%	0.510	-0.063
滞后 5 期-单指标策略	112.54%	2.05%	0.240	45.82%	0.085	-0.091
等权配置策略	180.67%	10.67%	0.091	14.12%	1.174	

资料来源：Wind, Bloomberg, 华泰证券研究所

其中 Softmax-Bagging 表示基于集成学习-Softmax 模型资产配置策略的回测结果。可以看到，基于集成学习模型的资产配置策略能够在保持与基准等权配置策略最大回撤水平相当的前提下提升资产组合的年化收益率和夏普比率。且通过结果的集成，Softmax-Bagging 模型的回测年化收益率、最大回撤和夏普比率均接近单一 Softmax 模型中的最优值，这说明集成学习模型不仅能消除单一模型对于指标选取的依赖，降低过拟合，同样也存在改进模型结果的潜力。

图表25： 基于 Softmax-集成学习预测结果的全球主要股债细分资产配置净值曲线



资料来源：Wind, Bloomberg, 华泰证券研究所

全球大类资产等权指数实证结果

进一步，我们对全球主要大类资产的等权指数利用基于周期三因子的 Softmax 方法进行回测分析。这里，我们将全球主要大类资产分为了股票、债券和大宗商品，并选择了 17 个指数作为上述三个大类资产的代表，具体如下表所示：

图表26： 全球大类资产等权指数选择

类别名称	具体指数	指数代码	起止时间
股票等权指数	上证综指	000001.SH	2004-01 至 2018-03
	标普 500	SPX.GI	2004-01 至 2018-03
	富时 100	FTSE.GI	2004-01 至 2018-03
	法国 CAC40	FCHI.GI	2004-01 至 2018-03
	德国 DAX	GDAXI.GI	2004-01 至 2018-03
	日经 225	N225.GI	2004-01 至 2018-03
	恒生指数	HSI.HI	2004-01 至 2018-03
债券等权指数	美国总回报指数		2004-01 至 2018-03
	泛欧总回报指数		2004-01 至 2018-03
	中国总回报指数		2004-01 至 2018-03
大宗商品等权指数	CRB 食品现货	CRBFD.RB	2004-01 至 2018-03
	CRB 食用油现货	CRBFO.RB	2004-01 至 2018-03
	CRB 家禽现货	CRBLS.RB	2004-01 至 2018-03
	CRB 金属现货	CRBMS.RB	2004-01 至 2018-03
	CRB 工业现货	CRBRI.RB	2004-01 至 2018-03
	CRB 综合现货	CRBSA.RB	2004-01 至 2018-03
	CRB 纺织现货	CRBTS.RB	2004-01 至 2018-03

资料来源：华泰证券研究所

值得一提的是，由于本研究回测所采用的数据频率为月度，因此数据长度非常有限。对于 Softmax 模型，其带估计参数 θ 数量与待分类类别数目 N 呈二次正相关关系，因此，若分类类别数目过大，则根据有限的训练样本长度得到的参数 θ 的估计值稳定性较差，且与真实值之间存在较大偏误²。该偏误会对于模型对资产未来表现预期的准确性产生负面影响。

然而，本文认为上述精确性下降的本质原因在于训练样本数目的不足，并不能作为 Softmax 模型能否有效挖掘周期三因子状态和资产表现内在联系的判断依据。因此，本文在对于全球大类资产指数进行回测时，首先将不同类别的细分资产指数进行等权合成，生成了股票、债券和大宗商品 3 个等权指数，进而利用 Softmax 模型对 3 个等权指数进行回测。其中，

² 与线性回归相似，回归系数估计值与真实值的偏误随样本长度的增加而降低

将上证综指,标普 500,富时 100 等 7 个主要股票市场指数等权合成为股票资产等权指数;将 CRB 食品现货,CRB 食用油现货等 7 个 CRB 指数等权合成为大宗商品资产等权指数;选取美国总回报指数(U.S. aggregate total return value), 泛欧总回报指数(Pan-European aggregate total return index value) 和中国总回报指数(China aggregate Total Return index value) 等权合成债券等权指数。数据的起止时间为 2004 年 1 月至 2018 年 3 月,数据采集频率为月度。经过上述指数合成操作,我们将资产的类别从 17 降低到 3,满足了目前数据长度下 Softmax 模型的训练样本需求。

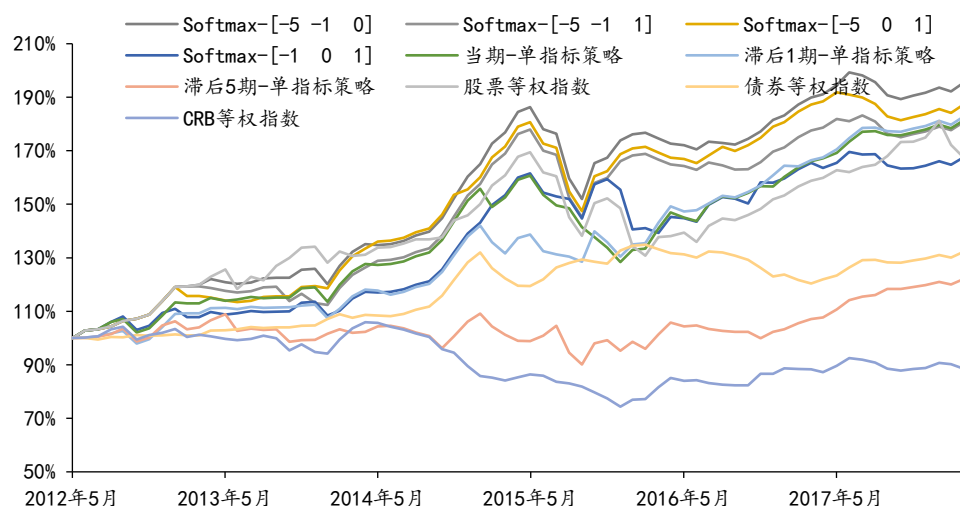
与全球主要股债指数资产配置的回测方式相同,我们选择滞后 5 期、滞后 1 期、当期和领先 1 期的同比增加值和同比拟合值共 4 对主要特征指标用于构建 Softmax 模型输入的“先行指标”集合。回测中依次选择 3 对主要特征指标构建单一 Softmax 模型,共计 4 个 Softmax 模型用于全球主要股债指数资产配置回测。回测的方式为动态预测,回测区间为 2012 年 5 月至 2018 年 3 月。在每个时刻 t,利用历史 90 个月的数据训练 Softmax 模型,并根据模型的预测结果调整资产配置比例。单一 Softmax 模型的回测结果如下图表所示:

图表27: 单一 Softmax 模型全球大类资产等权指数资产配置回测风险收益指标汇总

	总收益率	年化收益率	波动率	最大回撤	夏普比率	信息比率
Softmax-[-5 -1 0]	195.37%	12.17%	0.091	18.39%	1.332	0.219
Softmax-[-5 -1 1]	180.66%	10.67%	0.093	18.39%	1.145	0.124
Softmax-[-5 0 1]	187.23%	11.35%	0.089	18.39%	1.281	0.164
Softmax-[-1 0 1]	167.54%	9.25%	0.097	13.76%	0.956	0.039
当期-单指标策略	181.34%	10.74%	0.088	20.07%	1.222	0.097
滞后 1 期-单指标策略	182.74%	10.89%	0.087	9.46%	1.253	0.117
滞后 5 期-单指标策略	122.02%	3.47%	0.102	17.46%	0.341	-0.157
等权配置策略	161.53%	8.57%	0.061	11.74%	1.394	
股票等权指数	167.55%	9.25%	0.109	22.80%	0.848	0.038
债券等权指数	132.23%	4.91%	0.059	10.75%	0.828	-0.120
CRB 等权指数	88.52%	-2.07%	0.082	29.81%	-0.251	-0.333

资料来源: Wind, Bloomberg, 华泰证券研究所

可以看到,基于上述 4 个单一 Softmax 模型结果的资产配置策略的年化收益率均超过了基准的等权配置策略。其中,模型 Softmax-[-5 -1 0]的年化收益率最高,达到了 12.17%。这同样说明单一 Softmax 模型在挖掘资产周期状态同未来表现间内在逻辑的有效性。然而,与上节全球股债细分指数资产配置回测结果相比,单一 Softmax 模型对于大类资产等权指数的配置效果没有明显优于单指标策略。本研究认为,上述差异的不显著主要在于进行配置的资产种类较少(仅有股票、债券和大宗商品三类),Softmax 模型由于引入概率预期结果的优势在资产种类较少的情况下不能完全体现,因此其对于资产表现判断的准确度与单指标策略差距不大。

图表28： 基于不同先行指标的单一 Softmax 预测结果的全球大类资产等权指数资产配置净值曲线


资料来源: Wind, Bloomberg, 华泰证券研究所

然而, 基于上述四个单一 Softmax 模型预测结果的资产配置策略的回测结果同样仍存在一定差异, 例如, 模型 Softmax $[-1\ 0\ 1]$ 回测效果不佳, 其年化收益率同基准的等权策略相当, 但波动率和最大回撤等风险指标上不如基准的等权策略, 甚至低于某些单一的等权指数。

另一方面, 与全球主要股债细分资产指数回测结果相比, 表现最好的单一 Softmax 模型所依赖的先行指标集合存在差异。具体而言对于全球主要股债细分资产指数, 建立在滞后 5 期、滞后 1 期和领先 1 期同比增加值和同比拟合值的模型 Softmax $[-5, -1, 1]$ 表现最好, 而在资产等权指数的回测中, 模型 Softmax $[-5\ -1\ 0]$ 取得了最为突出的回测效果。因此, 基于 Softmax 模型预测结果的资产配置策略回测效果依赖于先行指标集合的选取, 且对于不同的大类资产指数, 取得最优效果的先行指标集合存在差异。

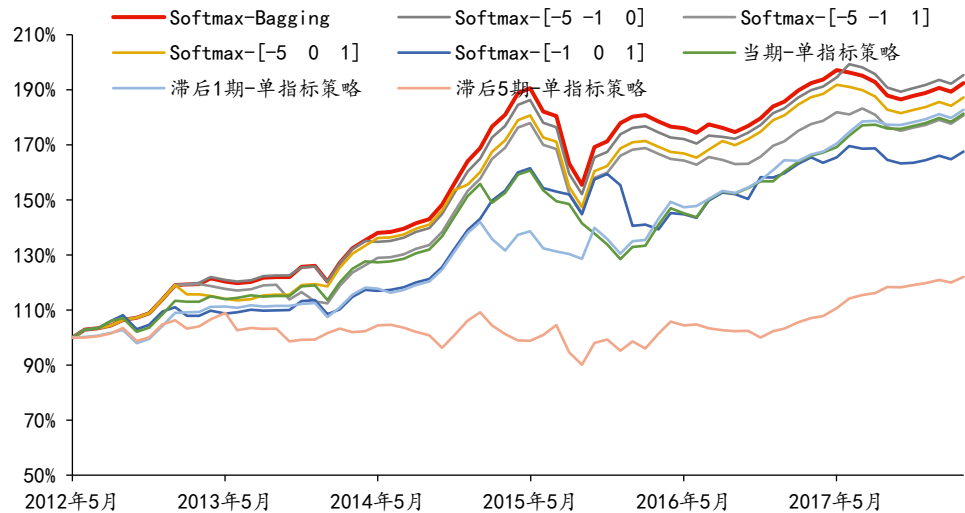
为了降低 Softmax 模型结果对于先行指标选取的依赖, 本研究采用了集成学习的方式, 在每个时刻 t , 对于基于不同先行指标集合的单一 Softmax 模型给出的预测结果进行加总, 进而输出集成后的预测结果; 同时根据集成的资产表现的预测结果进行资产配置。集成学习-Softmax 模型的回测结果如下图表所示:

图表29： Softmax-集成学习模型全球大类资产等权指数配置回测风险收益指标汇总

	总收益率	年化收益率	波动率	最大回撤	夏普比率	信息比率
Softmax-Bagging	192.43%	11.87%	0.092	18.39%	1.293	0.199
Softmax $[-5\ -1\ 0]$	195.37%	12.17%	0.091	18.39%	1.332	0.219
Softmax $[-5\ -1\ 1]$	180.66%	10.67%	0.093	18.39%	1.145	0.124
Softmax $[-5\ 0\ 1]$	187.23%	11.35%	0.089	18.39%	1.281	0.164
Softmax $[-1\ 0\ 1]$	167.54%	9.25%	0.097	13.76%	0.956	0.039
当期-单指标策略	181.34%	10.74%	0.088	20.07%	1.222	0.097
滞后 1 期-单指标策略	182.74%	10.89%	0.087	9.46%	1.253	0.117
滞后 5 期-单指标策略	122.02%	3.47%	0.102	17.46%	0.341	-0.157
等权配置策略	161.53%	8.57%	0.061	11.74%	1.394	\

资料来源: Wind, Bloomberg, 华泰证券研究所

其中 Softmax-Bagging 表示基于集成学习-Softmax 模型资产配置策略的回测结果。可以看到, 集成学习模型在全球大类资产等权指数配置的回测中同样能够超过等权配置策略的年化收益率。

图表30： 基于 Softmax-集成学习预测结果的全球大类资产等权指数配置净值曲线

资料来源: Wind, Bloomberg, 华泰证券研究所

中国主要大类资产指数实证结果

最后，我们将回测分析关注的焦点转移回中国市场，对中国的主要资产指数利用基于周期三因子的 Softmax 方法进行回测分析。本文所采用的中国主要大类资产指数包括股票、债券和大宗商品共三大类 8 个细分指数，具体如下图表所示。其中，选取沪深 300，中证 500 和中证 1000 分别作为大中小盘股票资产的代表；选取中债综合财富指数代表中国债券市场指数；选取南华工业品指数，南华农产品指数，南华金属指数和南华能化指数作为中国大宗商品市场指数的代表。数据的起止时间为 2005 年 1 月至 2018 年 3 月，数据采集频率为月度。

图表31： 中国主要大类资产指数选择

类别名称	具体指数	指数代码	起止时间
股票	沪深 300	000300.SH	2005-01 至 2018-03
	中证 500	399905.SZ	2005-01 至 2018-03
	中证 1000	000852.SH	2005-01 至 2018-03
	中债综合财富指数	CBA00201.CS	2005-01 至 2018-03
债券	中债综合财富指数	CBA00201.CS	2005-01 至 2018-03
大宗商品	南华工业品指数	NH0200.NHF	2005-01 至 2018-03
	南华农产品指数	NH0300.NHF	2005-01 至 2018-03
	南华金属指数	NH0400.NHF	2005-01 至 2018-03
	南华能化指数	NH0500.NHF	2005-01 至 2018-03

资料来源: Wind, 华泰证券研究所

与全球主要股债指数和全球大类资产等权指数资产配置的回测方式相同，我们选择滞后 5 期、滞后 1 期、当期和领先 1 期的同比增加值和同比拟合值共 4 对主要特征指标用于构建 Softmax 模型输入的“先行指标”集合。回测中依次选择 3 对主要特征指标构建单一 Softmax 模型，共计 4 个 Softmax 模型用于全球主要股债指数资产配置回测。

回测的方式为动态预测，回测区间为 2011 年 9 月至 2018 年 3 月。在每个时刻 t ，利用历史 70 个月的数据训练 Softmax 模型，并根据模型的预测结果调整资产配置比例。单一 Softmax 模型的回测结果如下图表所示：

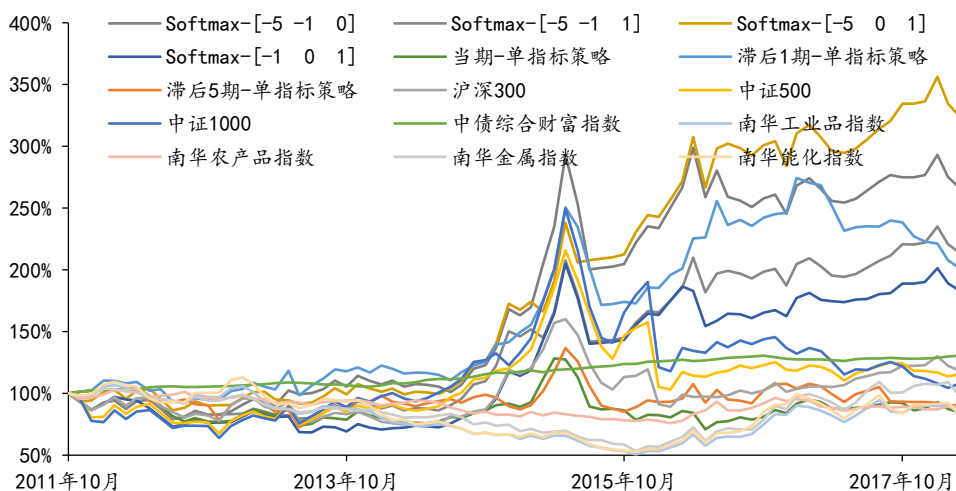
图表32： 单一 Softmax 模型中国主要大类资产指数资产配置回测风险收益指标汇总

	总收益率	年化收益率	波动率	最大回撤	夏普比率	信息比率
Softmax-[-5 -1 0]	266.49%	16.50%	0.264	31.69%	0.625	0.223
Softmax-[-5 -1 1]	213.80%	12.57%	0.247	31.69%	0.510	0.146
Softmax-[-5 0 1]	324.07%	20.11%	0.231	16.76%	0.871	0.261
Softmax-[-1 0 1]	183.13%	9.89%	0.257	31.69%	0.385	0.099
当期-单指标策略	84.65%	-2.56%	0.215	44.71%	-0.119	-0.162
滞后 1 期-单指标策略	201.29%	11.52%	0.233	31.50%	0.495	0.120
滞后 5 期-单指标策略	91.09%	-1.44%	0.237	37.73%	-0.061	-0.114
等权配置策略	149.81%	6.50%	0.162	25.86%	0.400	\
沪深 300	118.42%	2.67%	0.249	44.38%	0.107	-0.033
中证 500	115.45%	2.26%	0.289	52.27%	0.078	-0.015
中证 1000	108.15%	1.23%	0.332	58.14%	0.037	-0.008
中债综合财富指数	130.73%	4.26%	0.023	3.23%	1.823	-0.060
南华工业品指数	82.92%	-2.88%	0.183	52.55%	-0.157	-0.163
南华农产品指数	89.10%	-1.78%	0.096	27.01%	-0.185	-0.161
南华金属指数	98.18%	-0.29%	0.187	48.65%	-0.015	-0.104
南华能化指数	83.07%	-2.85%	0.203	53.14%	-0.140	-0.142

资料来源：Wind，华泰证券研究所

可以看到，对于上述 8 个中国主要大类资产指数，基于上述 4 个单一 Softmax 模型结果的资产配置策略的回测年化收益率全部超过了基准的等权配置策略，其中，模型 Softmax-[-5 0 1] 的年化收益率最高，达到了 20.11%，且其最大回撤同样为所有资产配置策略中最低，仅为 16.76%；与此同时，该策略的夏普比率同样优于基准策略。于此同时，在进行配置的资产种类较多的情况下，基于单一 Softmax 模型的资产配置策略的各项风险收益指标均超过了单指标策略，显示出单一 Softmax 模型在挖掘资产周期状态同未来表现间内在逻辑的有效性。

图表33： 基于不同先行指标的单一 Softmax 预测结果的中国主要大类资产指数资产配置净值曲线



资料来源：Wind，华泰证券研究所

然而，基于上述四个单一 Softmax 模型预测结果的资产配置策略的回测结果仍存在一定差异，例如，模型 Softmax-[-5 -1 0]，Softmax-[-5 -1 1] 和 Softmax-[-1 0 1] 虽然取得了超过基准策略的年化收益率，但其最大回撤均达到了 31.69%，超过了基准策略，为所有模型中最高；其中，模型 Softmax-[-1 0 1] 的夏普比率为上述 4 个基于 Softmax 预测结果的资产策略中最低，其结果劣于基准的等权配置策略，说明该 Softmax 模型未能有效的对资产未来表现进行判断。

另一方面，与全球大类资产等权指数回测结果相比，表现最好的单一 Softmax 模型所依赖的先行指标集合存在差异。具体而言，对于全球大类资产等权指数，建立在滞后 5 期、滞

后 1 期和当期同比增加值和同比拟合值的模型 Softmax-[-5 -1 0] 表现最好，而在中国主要大类资产的回测中，模型 Softmax-[-5 0 1] 取得了最为突出的回测效果。因此，我们同样的到了结论：基于 Softmax 模型预测结果的资产配置策略回测效果依赖于先行指标集合的选取，且对于不同的大类资产指数，取得最优效果的先行指标集合存在差异。

为了降低 Softmax 模型结果对于先行指标选取的依赖，本研究同样采用了集成学习的方式，在每个时刻 t ，对于基于不同先行指标集合的单一 Softmax 模型给出的预测结果进行加总，进而输出集成后的预测结果；同时根据集成的资产表现的预测结果进行资产配置。集成学习-Softmax 模型的回测结果如下图表所示：

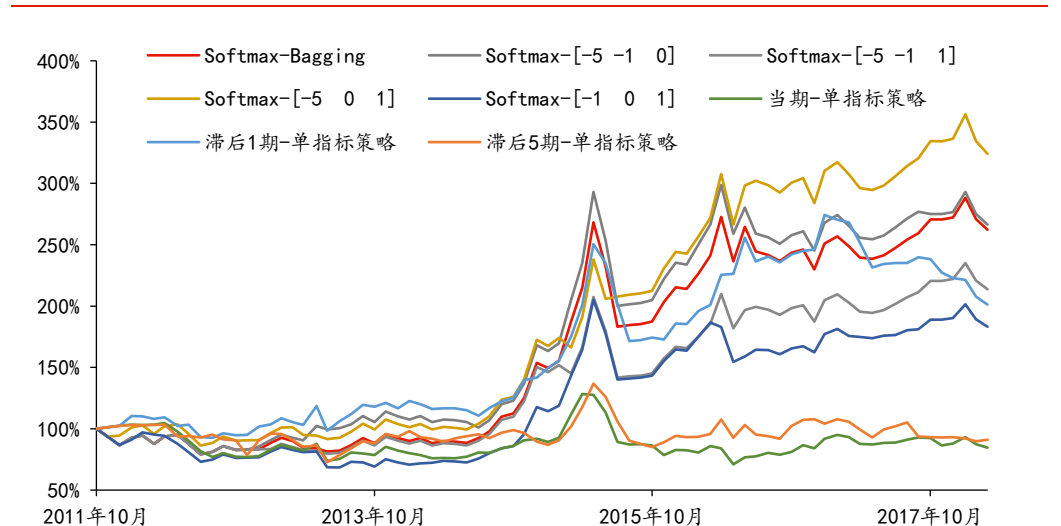
图表34： Softmax-集成学习模型中国主要大类资产指数配置回测风险收益指标汇总

	总收益率	年化收益率	波动率	最大回撤	夏普比率	信息比率
Softmax-Bagging	262.25%	16.21%	0.264	31.69%	0.615	0.219
Softmax-[-5 -1 0]	266.49%	16.50%	0.264	31.69%	0.625	0.223
Softmax-[-5 -1 1]	213.80%	12.57%	0.247	31.69%	0.510	0.146
Softmax-[-5 0 1]	324.07%	20.11%	0.231	16.76%	0.871	0.261
Softmax-[-1 0 1]	183.13%	9.89%	0.257	31.69%	0.385	0.099
当期-单指标策略	84.65%	-2.56%	0.215	44.71%	-0.119	-0.162
滞后 1 期-单指标策略	201.29%	11.52%	0.233	31.50%	0.495	0.120
滞后 5 期-单指标策略	91.09%	-1.44%	0.237	37.73%	-0.061	-0.114
等权配置策略	149.81%	6.50%	0.162	25.86%	0.400	\

资料来源：Wind，华泰证券研究所

其中 Softmax-Bagging 表示基于集成学习-Softmax 模型资产配置策略的回测结果。可以看到，集成学习模型在中国大类资产等权指数配置的回测中同样能够在保持与基准等权配置策略最大回撤水平相当的条件提升资产组合的年化收益率和夏普比率。

图表35： 基于 Softmax-集成学习预测结果的中国主要大类资产指数配置净值曲线



资料来源：Wind，华泰证券研究所

2018 下半年的资产表现排序的预测结论

根据前文所述，我们利用机器学习中的 Softmax 方法，深度挖掘周期状态与资产收益之间的内在联系。同时通过仿真数据和真实资产数据的回测分析，验证了 Softmax 方法把握该内在联系的有效性。本节将利用周期的外推延拓性质，对于资产未来的周期状态进行预测；并将资产周期状态的预期结果作为 Softmax 模型的输入，根据模型输出的概率结果对于资产未来表现进行判断。

根据周期三因子定价模型，资产每个时刻的周期状态可用正弦序列表示，其相位状态是关于时间的线性函数。因此，对于任意时刻 T ，如果我们已知资产的周期频率 ω 和相位信息 φ_T ，

我们可以得到未来任意时刻 $T+n$ 周期状态相位信息的估计值：

$$\hat{\varphi}_{T+n} = \varphi_T + \frac{2\pi}{\omega} \times n$$

该性质称为周期因子的外推延拓性质。利用延拓后的周期相位信息，我们可以得到未来任意时刻资产周期状态的估计值（在 T 时刻已知的信息下），进而可以计算上文提到的各类先行指标的估计值。将上述估计值作为当期 Softmax 模型的输入，则可得到 Softmax 模型对于资产未来任意时刻表现的预期。

具体而言，在时刻 T ，基于周期三因子的外推延拓性质，我们可以得到任意大类资产周期三因子状态在未来 $T+n$ 时刻的预期值，进而得到评价 $T+n$ 时刻每类资产预期收益的先行指标；基于时刻 T 之前的历史数据，我们可以对于不同的先行指标输入的 Softmax 模型；将先行指标作为训练完毕的 Softmax 模型的输入，可以计算未来每个时点各项资产取得最好表现的概率估计值，依据该估计值对于资产未来的表现进行预测和排序。

利用上述方法，下文依次对于全球大类资产等权指数、全球主要股债细分资产指数和中国主要细分资产指数 2018 年 5 月至 2018 年 12 月的表现进行预测。下图表展示了全球大类资产等权指数 2018 年 5 月至 12 月的预期表现汇总：

图表36： 2018 年 5 月至 12 月全球大类资产等权指数预期表现评价汇总（数据越大预期表现越优）

	股票等权指数	债券等权指数	CRB 等权指数
2018-05	0.247	0.712	0.040
2018-06	0.206	0.754	0.039
2018-07	0.158	0.801	0.041
2018-08	0.109	0.845	0.046
2018-09	0.068	0.878	0.054
2018-10	0.039	0.897	0.064
2018-11	0.021	0.901	0.078
2018-12	0.011	0.895	0.095

资料来源：Wind, Bloomberg, 华泰证券研究所

2018 年 5 月至 12 月全球大类资产等权指数预期表现结果显示：代表债券大类资产的债券等权指数表现最强，且随着时间的推移，债券指数表现超过股票和大宗商品等权指数的概率越高，说明在未来一段时间债券资产存在较大的投资机会。大宗商品等权指数的预期概率变化范围不大；上半年股票资产仍存在一定的投资机会，但随后其评分显著下降，18 年 9 月后预期表现不如大宗商品，这与先前报告“大宗商品未来或将高位震荡，风险主要来源于股票市场”的观点不谋而合。

进一步，本文根据全球主要股债细分资产指数的历史数据，对于 10 只股票和债券资产的代表指数 2018 年 5 月至 12 月的预期表现进行分析。可以看到，仅从全球股债市场看，未来 8 个月中债券大类资产的优势更加明显，具体而言，代表中债资产的中国总回报指数综合表现最好，其表现最优的概率最高达到了 0.597，为所有细分资产中最高；泛欧总回报在 18 年后期有较好表现，其取得最优表现的概率在 18 年 9 月至 11 月一度小幅超过了中国总回报指数。股票细分资产中，上证综指在 18 年 5 月至 6 月存在一定投资机会，随后该指数的概率估计值显著下降；其余指数表现最优的概率值均未超过 10%。

图表37: 2018 年 5 月至 12 月全球主要股债细分资产指数预期表现评价汇总 (数据越大预期表现越优)

	法国							美国总回	泛欧总回	中国总回
	上证综指	标普 500	富时 100	CAC40	德国 DAX	日经 225	恒生指数	报指数	报指数	报指数
2018-05	0.164	0.009	0.001	0.017	0.053	0.057	0.010	0.048	0.057	0.585
2018-06	0.140	0.007	0.000	0.039	0.059	0.076	0.005	0.049	0.107	0.519
2018-07	0.091	0.005	0.000	0.073	0.051	0.077	0.003	0.039	0.204	0.458
2018-08	0.045	0.003	0.000	0.099	0.034	0.056	0.002	0.023	0.340	0.397
2018-09	0.018	0.002	0.000	0.095	0.019	0.032	0.002	0.010	0.466	0.356
2018-10	0.007	0.001	0.000	0.068	0.010	0.015	0.003	0.004	0.524	0.366
2018-11	0.003	0.001	0.001	0.039	0.006	0.008	0.012	0.002	0.475	0.455
2018-12	0.001	0.000	0.004	0.016	0.003	0.004	0.082	0.001	0.292	0.597

资料来源: Wind, Bloomberg, 华泰证券研究所

最后, 本文利用中国主要细分资产指数的历史数据训练 Softmax 模型, 对于 8 只代表指数 2018 年 5 月至 2018 年 12 月的预期表现进行预测。可以看到, 对于中国主要细分资产, 2018 年下半年债券资产的预期表现依然强势。代表债券资产的中债综合财富指数的评价价值在 2018 年 6 月超过股票资产后, 一直呈上涨趋势。股票资产在 2018 年 5 至 6 月或仍存一定投资价值, 代表大盘股的沪深 300 指数的评价价值在 5、6 月仍然超过 40%, 代表小盘股的中证 1000 指数的评价价值在 5 月也接近 10%; 随后, 股票细分资产指数表现最优的概率显著下降, 预期成为市场风险的主要来源。

图表38: 2018 年 5 月至 12 月中国主要细分资产指数预期表现评价汇总 (数据越大预期表现越优)

	沪深 300	中证 500	中证 1000	中债综合财富指数	南华工业品指数	南华农产品指数	南华金属指数	南华能化指数
2018-05	0.617	0.004	0.072	0.303	0.004	0.000	0.000	0.000
2018-06	0.470	0.005	0.023	0.496	0.005	0.000	0.000	0.000
2018-07	0.061	0.003	0.002	0.931	0.003	0.000	0.000	0.000
2018-08	0.003	0.001	0.000	0.995	0.001	0.000	0.000	0.000
2018-09	0.000	0.001	0.000	0.999	0.001	0.000	0.000	0.000
2018-10	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.000
2018-11	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2018-12	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000

资料来源: Wind, 华泰证券研究所

风险提示

本文基于华泰金工周期系列研究对全球各类经济金融指标长达近百年样本的实证检验结果，确定周期长度。然而市场存在短期波动与政策冲击，就每轮周期而言，暂无法判断具体长度。周期长度只是估计值，可能存在偏差；历史规律存在失效风险。

附录一：采用同比序列作为资产收益预测的数据基础的原因

本部分对全球主要股票市场、债券市场和大宗商品市场指数的环比序列进行分析，验证观点：环比序列受短期噪声冲击影响，周期状态难以准确把握。

资产价格的衍生序列（同比、环比序列等）的各个周期状态可以通过傅里叶变换和高斯滤波的方式进行确定，其结果的精确程度受到时间序列总方差中可解释部分占比大小的影响。具体而言，如果总方差中可被三周期状态解释部分的占比越高，则滤波结果越精确，即越能够获得资产价格衍生序列的准确周期状态；反之，则周期状态的估计偏误越大。

对于环比序列，本文认为由于序列本身容易受到短期噪声的冲击，其波动中包含了大量不可解释的部分，因此通过高斯滤波获取的三周期状态准确度较低，难以据此对于序列的未来走势进行估计和判断。下文将对全球主要股票市场、债券市场和大宗商品市场指数的环比序列进行分析，从而验证上述观点。分析所用指数如下表所示：

图表39： 环比序列分析所用指标汇总

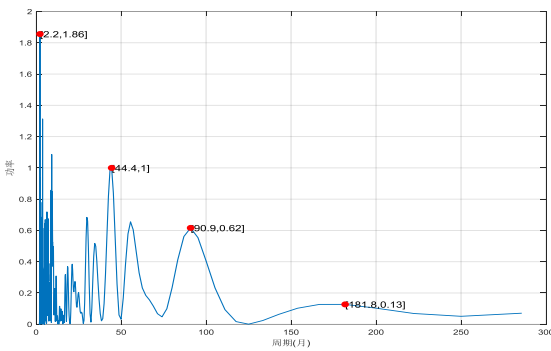
类别名称	具体指数	指数代码	起止时间
股票市场	上证综指	000001.SH	2004-01 至 2018-03
	标普 500	SPX.GI	2004-01 至 2018-03
	富时 100	FTSE.GI	2004-01 至 2018-03
	法国 CAC40	FCHI.GI	2004-01 至 2018-03
	德国 DAX	GDAXI.GI	2004-01 至 2018-03
	日经 225	N225.GI	2004-01 至 2018-03
	恒生指数	HSI.HI	2004-01 至 2018-03
	美国总回报指数		2004-01 至 2018-03
债券市场	泛欧总回报指数		2004-01 至 2018-03
	中国总回报指数		2004-01 至 2018-03
大宗商品市场	CRB 食品现货	CRBFD.RB	2004-01 至 2018-03
	CRB 食用油现货	CRBFO.RB	2004-01 至 2018-03
	CRB 家禽现货	CRBLS.RB	2004-01 至 2018-03
	CRB 金属现货	CRBMS.RB	2004-01 至 2018-03
	CRB 工业现货	CRBRI.RB	2004-01 至 2018-03
	CRB 综合现货	CRBSA.RB	2004-01 至 2018-03
	CRB 纺织现货	CRBTS.RB	2004-01 至 2018-03

资料来源：Wind, Bloomberg, 华泰证券研究所

首先，对于上述全球主要股票市场、债券市场和大宗商品市场共计 17 个指数进行傅里叶变换，分析高频分量的振幅同短周期振幅大小的相对关系。下图展示了上证综指环比序列的功率谱，其中固定短周期（44 个月附近）的振幅为 1，其余频率对应的振幅做相对调整。

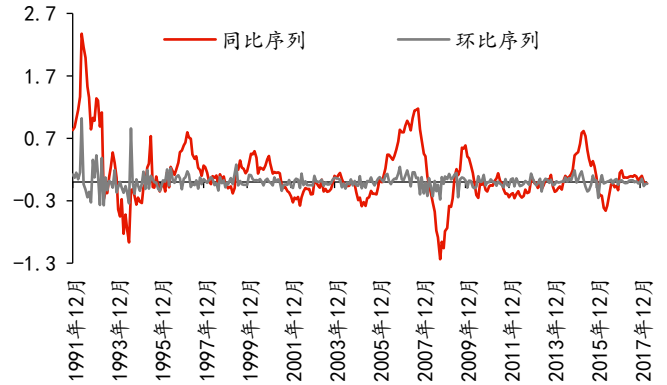
根据下表结果，上证综指环比序列的功率谱函数在 42 个月、100 个月和 200 个月附近存在明显波峰，分别对应三周期中的基钦周期、朱格拉周期和库兹涅茨周期。然而，在功率谱的高频部分（即 x 轴最左端，周期为 0 至 12 个月部分）同样出现了明显的波峰，其中最高点出现在 2.2 个月附近，其功率振幅约为短周期的 1.86 倍；同时，高频部分的功率谱函数起伏变化很大，且各个波峰均超过了 42 个月附近短周期振幅的 0.6 倍，这说明上证综指的环比序列中可以提取到多个高频分量，且每个高频分量均蕴含相对显著的能量，对于上证综指环比的总体方差存在较大贡献。综上所述，短期噪声对于上证综指环比序列影响较大。

图表40： 上证综指环比序列功率谱图（固定短周期功率为 1）



资料来源：Wind，华泰证券研究所

图表41： 上证综指环比、同比序列



资料来源：Wind，华泰证券研究所

下表汇总了上述 17 个全球主要市场指数的功率谱高频部分能量最高处对应的周期和相对振幅（固定短周期振幅为 1）。可以看到，对于绝大多数指数的环比序列，高频部分信号能量最高点的振幅大于短周期分量能量振幅，其中差异最大的为泛欧总回报指数，其高频分量在 10.152 个月处的能量密度为短周期的 8.011 倍。这说明对于大部分指数，其环比序列方差中存在很高比例的不可被周期三因子解释的部分，这部分波动对于指数环比周期状态的估计产生负面影响。

图表42： 全球主要资产指数高频部分功率最高处对应周期和相对振幅汇总

	周期（月）	相对振幅（固定短周期振幅为 1）
上证综指	2.250	1.856
标普 500	5.917	1.144
富时 100	4.329	1.636
法国 CAC40	4.320	1.312
德国 DAX	5.952	1.277
日经 225	4.843	1.360
恒生指数	8.772	1.146
美国总回报指数	2.959	1.703
泛欧总回报指数	10.152	8.011
中国总回报指数	5.525	0.437
CRB 食品现货	3.515	1.650
CRB 食用油现货	11.905	4.208
CRB 家禽现货	11.905	2.718
CRB 金属现货	2.174	0.574
CRB 工业现货	2.268	0.400
CRB 综合现货	11.905	0.582
CRB 纺织现货	11.905	0.806

资料来源：华泰证券研究所

上述周期状态的估计偏误使得线性周期三因子定价模型对于资产环比序列的解释力度降低。周期三因子定价模型可以表示为：

$$y = \alpha + \beta_1 x_{42} + \beta_2 x_{100} + \beta_3 x_{200} + \varepsilon$$

其中 y 表示资产价格的衍生序列，这里为环比序列； x_{42}, x_{100}, x_{200} 为基于资产价格衍生序列 y 提取的周期三因子状态变量； ε 为随机扰动项，为白噪声序列。

图表43： 全球主要市场指数环比序列周三因子回归结果

	短周期回归系数 β_1	中周期回归系数 β_2	长周期回归系数 β_3	可决系数 R^2
上证综指	0.050	0.021	0.003	0.130
标普 500	0.017	0.010	0.003	0.097
富时 100	0.015	0.008	0.001	0.075
法国 CAC40	0.019	0.012	0.003	0.091
德国 DAX	0.020	0.011	0.001	0.072
日经 225	0.019	0.015	0.005	0.091
恒生指数	0.022	0.007	0.002	0.056
美国总回报指数	0.002	0.002	0.001	0.038
泛欧总回报指数	0.011	0.009	0.002	0.053
中国总回报指数	0.004	0.002	0.001	0.104
CRB 食品现货	0.019	0.009	0.000	0.080
CRB 食用油现货	0.029	0.016	-0.003	0.063
CRB 家禽现货	0.021	0.010	0.002	0.060
CRB 金属现货	0.032	0.017	0.005	0.146
CRB 工业现货	0.020	0.008	0.001	0.156
CRB 综合现货	0.018	0.007	0.001	0.137
CRB 纺织现货	0.012	0.006	0.001	0.088

资料来源：华泰证券研究所

周三因子线性回归模型对于环比序列的解释力度可以通过回归的可决系数 R^2 体现，上图展示了上述 17 个全球主要市场指数的周三因子线性回归模型结果，对于上述 17 个指数的环比序列，周三因子线性回归模型的可决系数 R^2 均小于 15%，即环比序列中能够被周期状态解释的波动占总波动的比例均低于 15%，这说明了线性模型对于环比走势的解释力度较差。

综上所述，由于环比序列容易受到短期噪声的影响，其波动中不可解释的高频分量占比较高；较高的不可解释方差占比对于周三因子状态的滤波估计产生了负面影响，使估计值与真实值的偏差增大，这反映在周三因子回归中为模型的可决系数较低，即基于周三因子的线性回归模型不能很好的对资产的月度收益率进行解释，同时对资产未来走势的预测能力不足。因此，需要发掘新的指标和方法，构建周期状态同资产短期表现的相关关系。

图表44： 全球大类资产等权指数选择

类别名称	具体指数	指数代码	起止时间
股票	上证综指	000001.SH	2004-01 至 2018-03
	标普 500	SPX.GI	2004-01 至 2018-03
	富时 100	FTSE.GI	2004-01 至 2018-03
	法国 CAC40	FCHI.GI	2004-01 至 2018-03
	德国 DAX	GDAXI.GI	2004-01 至 2018-03
	日经 225	N225.GI	2004-01 至 2018-03
	恒生指数	HSI.HI	2004-01 至 2018-03
	美国总回报指数		2004-01 至 2018-03
债券	泛欧总回报指数		2004-01 至 2018-03
	中国总回报指数		2004-01 至 2018-03
	CRB 食品现货	CRBFD.RB	2004-01 至 2018-03
大宗商品	CRB 食用油现货	CRBFO.RB	2004-01 至 2018-03
	CRB 家禽现货	CRBLS.RB	2004-01 至 2018-03
	CRB 金属现货	CRBMS.RB	2004-01 至 2018-03
	CRB 工业现货	CRBRI.RB	2004-01 至 2018-03
	CRB 综合现货	CRBSA.RB	2004-01 至 2018-03
	CRB 纺织现货	CRBTS.RB	2004-01 至 2018-03

资料来源：华泰证券研究所

首先，对上述全球主要股票市场、债券市场和大宗商品市场共计 17 个指数进行傅里叶变换，分析高频分量的振幅同短周期振幅大小的相对关系。下表汇总了上述 17 个全球主要市场指数的功率谱高频部分能量最高处对应的周期和相对振幅（固定短周期振幅为 1）。可

以看到，对于绝大多数指数的环比序列，高频部分信号能量最高点的振幅大于短周期分量能量振幅，其中差异最大的为泛欧总回报指数，其高频分量在 10.152 个月处的能量密度为短周期的 8.011 倍。这说明对于大部分指数，其环比序列方差中存在很高比例的不可被周期三因子解释的部分，这部分波动对于指数环比周期状态的估计产生负面影响。

图表45： 全球主要资产指数高频部分功率最高处对应周期和相对振幅汇总

	周期（月）	相对振幅（固定短周期振幅为 1）
上证综指	2.250	1.856
标普 500	5.917	1.144
富时 100	4.329	1.636
法国 CAC40	4.320	1.312
德国 DAX	5.952	1.277
日经 225	4.843	1.360
恒生指数	8.772	1.146
美国总回报指数	2.959	1.703
泛欧总回报指数	10.152	8.011
中国总回报指数	5.525	0.437
CRB 食品现货	3.515	1.650
CRB 食用油现货	11.905	4.208
CRB 家禽现货	11.905	2.718
CRB 金属现货	2.174	0.574
CRB 工业现货	2.268	0.400
CRB 综合现货	11.905	0.582
CRB 纺织现货	11.905	0.806

资料来源：华泰证券研究所

上述周期状态的估计偏误使得线性周期三因子定价模型对于资产环比序列的解释力度降低。周期三因子线性回归模型对于环比序列的解释力度可以通过回归的可决系数 R^2 体现，下图表展示了上述 17 个全球主要市场指数的周期三因子线性回归模型结果，对于上述 17 个指数的环比序列，周期三因子线性回归模型的可决系数 R^2 均小于 15%，即环比序列中能够被周期状态解释的波动占总波动的比例均低于 15%，这说明了线性模型对于环比走势的解释力度较差。

图表46： 全球主要市场指数环比序列周期三因子回归结果

	短周期回归系数 β_1	中周期回归系数 β_2	长周期回归系数 β_3	可决系数 R^2
上证综指	0.050	0.021	0.003	0.130
标普 500	0.017	0.010	0.003	0.097
富时 100	0.015	0.008	0.001	0.075
法国 CAC40	0.019	0.012	0.003	0.091
德国 DAX	0.020	0.011	0.001	0.072
日经 225	0.019	0.015	0.005	0.091
恒生指数	0.022	0.007	0.002	0.056
美国总回报指数	0.002	0.002	0.001	0.038
泛欧总回报指数	0.011	0.009	0.002	0.053
中国总回报指数	0.004	0.002	0.001	0.104
CRB 食品现货	0.019	0.009	0.000	0.080
CRB 食用油现货	0.029	0.016	-0.003	0.063
CRB 家禽现货	0.021	0.010	0.002	0.060
CRB 金属现货	0.032	0.017	0.005	0.146
CRB 工业现货	0.020	0.008	0.001	0.156
CRB 综合现货	0.018	0.007	0.001	0.137
CRB 纺织现货	0.012	0.006	0.001	0.088

资料来源：华泰证券研究所

附录二：资产收益率分布的不对称性致概率预测好于确定性预测

本节我们用一个例子验证资产收益率分布的不对称性会导致基于确定性预期结果进行的资产表现预期结果出现偏误。例如，假设资产 A 的收益率服从均值为 μ 的有偏分布，资产 B 的收益率为常数 μ ，即资产 A 与资产 B 收益率的期望值相同。如果我们仅仅根据资产 A 与 B 收益率的期望判断两类资产的未来表现，则资产 A 未来表现强于资产 B 的概率与 B 强于 A 的概率相等，均为 0.5。

然而，对于下一时刻，资产 A 的表现优于资产 B 的概率可以表示为：

$$P(r_A \geq r_B) = P(r_A \geq \mu)$$

根据条件期望公式，有：

$$\begin{aligned} \mu &= E(r_A | r_A \geq \mu) P(r_A \geq \mu) + E(r_A | r_A < \mu) P(r_A < \mu) \\ &= E(r_A | r_A \geq \mu) P(r_A \geq \mu) + E(r_A | r_A < \mu) (1 - P(r_A \geq \mu)) \\ &= E(r_A | r_A < \mu) + [E(r_A | r_A \geq \mu) - E(r_A | r_A < \mu)] P(r_A \geq \mu) \end{aligned}$$

据此可得资产 A 的表现优于资产 B 的概率为：

$$P(r_A \geq \mu) = \frac{\mu - E(r_A | r_A < \mu)}{E(r_A | r_A \geq \mu) - E(r_A | r_A < \mu)}$$

由于我们假设资产 A 的收益率是有偏的，因此一般 $E(r_A | r_A \geq \mu) - \mu \neq \mu - E(r_A | r_A < \mu)$ ，故资产 A 的表现优于资产 B 的概率一般不为 0.5，而是同资产 A 收益率分布的具体形状相关。但如果我们仅仅依据确定性预期结果（即资产收益率期望）对于资产 A、B 的未来预期表现进行评价，会得到两类资产未来收益表现相当的结论，这与实际情况是不相符的。

附录三：机器学习之 Softmax 模型原理

Softmax 分类器模型是 logistic 回归模型在多分类问题上的推广，在诸如 MNIST 手写数字分类等多类别问题中表现较高的分类效率；同时，该方法较高的效率以及结果含义的明确性，使其成为绝大多数深度学习网络中的底层分类模块。

具体而言，对于 N 个资产指数，在周期三因子模型假设下，其同比序列满足如下线性关系：

$$\widehat{TB}_{i,t+l|t} = \alpha_i + \beta_{i,1}x_{i,t+l|t}^{42} + \beta_{i,2}x_{i,t+l|t}^{100} + \beta_{i,3}x_{i,t+l|t}^{200}$$

其中， $x_{i,t+l|t}^{42}, x_{i,t+l|t}^{100}, x_{i,t+l|t}^{200}$ 分表表示第 i 个资产根据 t 时刻信息得到的 t+l 时刻短中周期状态的估计值。在华泰金工周期系列报告中我们对于周期状态的此类外推延拓性质进行过详细介绍，在此不再赘述。 $\widehat{TB}_{i,t+l|t}$ 表示利用周期状态外推性质，在时刻 t 得到的时刻 t+l 期同比拟合值。

根据各期同比拟合值，我们可以得到资产各期的同比增加值，其计算公式如下：

$$\begin{aligned}\Delta\widehat{TB}_{i,t+l} &= \widehat{TB}_{i,t+l|t} - \widehat{TB}_{i,t+l-1|t} \\ &= \beta_{i,1}(x_{t+l|t}^{42} - x_{t+l-1|t}^{42}) + \beta_{i,2}(x_{t+l|t}^{100} - x_{t+l-1|t}^{100}) + \beta_{i,3}(x_{t+l|t}^{200} - x_{t+l-1|t}^{200})\end{aligned}$$

上述各滞后期、领先期和当期同比增加值与同比拟合值共同构成了主要特征指标集合，本研究在主要特征指标集合中选取合适的指标构成先行指标集作为 Softmax 模型的输入。

令先行指标集为 $I_t^i = \{(\widehat{TB}_{i,t+l_1|t}, \Delta\widehat{TB}_{i,t+l_1}), (\widehat{TB}_{i,t+l_2|t}, \Delta\widehat{TB}_{i,t+l_2}), \dots, (\widehat{TB}_{i,t+l_M|t}, \Delta\widehat{TB}_{i,t+l_M})\}$ 则本文针对 N 个资产指数构建的 Softmax 模型可以描述为：

$$\begin{aligned}& \left[\begin{aligned} P(r_{t+1}^1 = \max_{k=1,2,\dots,N} \{r_{t+1}^k\} | \theta) \\ P(r_{t+1}^2 = \max_{k=1,2,\dots,N} \{r_{t+1}^k\} | \theta) \\ \vdots \\ P(r_{t+1}^N = \max_{k=1,2,\dots,N} \{r_{t+1}^k\} | \theta) \end{aligned} \right] = \frac{1}{\sum_{n=1}^N \sum_{m=1}^M e^{\theta_{nm}^1 \widehat{TB}_{n,t+l_m|t} + \theta_{nm}^2 \Delta\widehat{TB}_{n,t+l_m}}} \\ & \times \left[\begin{aligned} e^{\theta_{11}^1 \widehat{TB}_{1,t+l_1|t} + \theta_{11}^2 \Delta\widehat{TB}_{1,t+l_1}} + \dots + e^{\theta_{1M}^1 \widehat{TB}_{1,t+l_M|t} + \theta_{1M}^2 \Delta\widehat{TB}_{1,t+l_M}} \\ e^{\theta_{21}^1 \widehat{TB}_{2,t+l_1|t} + \theta_{21}^2 \Delta\widehat{TB}_{2,t+l_1}} + \dots + e^{\theta_{2M}^1 \widehat{TB}_{2,t+l_M|t} + \theta_{2M}^2 \Delta\widehat{TB}_{2,t+l_M}} \\ \vdots \\ e^{\theta_{N1}^1 \widehat{TB}_{N,t+l_1|t} + \theta_{N1}^2 \Delta\widehat{TB}_{N,t+l_1}} + \dots + e^{\theta_{NM}^1 \widehat{TB}_{N,t+l_M|t} + \theta_{NM}^2 \Delta\widehat{TB}_{N,t+l_M}} \end{aligned} \right]\end{aligned}$$

其中 r_{t+1}^k 表示第 k 项大类资产在 t+1 时刻的收益率， $r_{t+1}^i = \max\{r_{t+1}^k, k=1,2,\dots,N\}$ 表示第 i 项资产 t+1 时刻收益率在全部 N 项资产排名第一，即该项资产未来预期表现最好。因此，在该定义下，集合 $\{r_{t+1}^i = \max\{r_{t+1}^k, k=1,2,\dots,N\}, i=1,2,\dots,N\}$ 构成了 N 个概率空间中不相交的集合，且上述 N 个集合的并集为全概率空间。因此，上述 N 个集合可以看作 N 个不同的类别。 $\theta = [\theta_{11}^1, \theta_{11}^2, \dots, \theta_{NM}^1, \theta_{NM}^2]$ 为模型待估参数，同线性回归中回归系数 β 的类似，可以通过历史数据进行估计。具体的估计方式为：在训练阶段，对于历史数据样本 $\{(I_t^1, y_t^1), \dots, (I_t^N, y_t^N)\}_{t=1,2,\dots,T}$ ，选择参数 θ 最大化如下似然函数：

$$J(\theta) = -\frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N I(y_t = n) \times \ln \left(\frac{\sum_{m=1}^M e^{\theta_{nm}^1 \widehat{TB}_{n,t+l_m|t} + \theta_{nm}^2 \Delta\widehat{TB}_{n,t+l_m}}}{\sum_{n=1}^N \sum_{m=1}^M e^{\theta_{nm}^1 \widehat{TB}_{n,t+l_m|t} + \theta_{nm}^2 \Delta\widehat{TB}_{n,t+l_m}}} \right)$$

并得到相应的最优参数 $\hat{\theta}$ 。

免责声明

本报告仅供华泰证券股份有限公司（以下简称“本公司”）客户使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司对该等信息的准确性及完整性不作任何保证。本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，本公司可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。本公司不保证本报告所含信息保持在最新状态。本公司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，但本报告所载的观点、结论和建议仅供参考，不构成所述证券的买卖出价或征价。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。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。对依据或者使用本报告所造成的一切后果，本公司及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

本公司及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，也可能为之提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。本公司的资产管理部、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权力。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

本公司具有中国证监会核准的“证券投资咨询”业务资格，经营许可证编号为：91320000704041011J。

全资子公司华泰金融控股（香港）有限公司具有香港证监会核准的“就证券提供意见”业务资格，经营许可证编号为：A0K809

©版权所有 2018 年华泰证券股份有限公司

评级说明

行业评级体系

一报告发布日后的 6 个月内的行业涨跌幅相对同期的沪深 300 指数的涨跌幅为基准；

一投资建议的评级标准

增持行业股票指数超越基准

中性行业股票指数基本与基准持平

减持行业股票指数明显弱于基准

公司评级体系

一报告发布日后的 6 个月内的公司涨跌幅相对同期的沪深 300 指数的涨跌幅为基准；

一投资建议的评级标准

买入股价超越基准 20%以上

增持股价超越基准 5%-20%

中性股价相对基准波动在-5%~5%之间

减持股价弱于基准 5%-20%

卖出股价弱于基准 20%以上

华泰证券研究

南京

南京市建邺区江东中路 228 号华泰证券广场 1 号楼/邮政编码：210019

电话：86 25 83389999/传真：86 25 83387521

电子邮件：ht-rd@htsc.com

深圳

深圳市福田区深南大道 4011 号香港中旅大厦 24 层/邮政编码：518048

电话：86 755 82493932/传真：86 755 82492062

电子邮件：ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同 28 号太平洋保险大厦 A 座 18 层

邮政编码：100032

电话：86 10 63211166/传真：86 10 63211275

电子邮件：ht-rd@htsc.com

上海

上海市浦东新区东方路 18 号保利广场 E 栋 23 楼/邮政编码：200120

电话：86 21 28972098/传真：86 21 28972068

电子邮件：ht-rd@htsc.com