

分析师：

徐寅

xuyinsh@xyzq.com.cn

S0190514070004

西学东渐——海外文献推荐系列之三十一

2019 年 07 月 04 日

报告关键点

投资经理通常会面对成千上万的上市公司和分析师，一个重要的问题是：如何对分析师给出的推荐数据进行建模并基于模型进行投资。本文应用贝叶斯方法对推荐数据进行建模，实证结果表明这种方法构建的投资组合表现十分优秀，同时 Both 类策略即当经纪商的推荐和 IBCC 模型预测结果一致时构建的组合表现是最好的。文章将变分近似与独立贝叶斯分类器组合（IBCC）方法相结合，大幅度提升了计算效率。

相关报告

《西学东渐——海外文献推荐系列之三十》 2019-06-27

投资要点

- 西学东渐，是指从明朝末年到近代，西方学术思想向中国传播的历史过程。西学东渐不仅推动了中国在科学技术和思想文化方面的发展，也有力地促进了社会与政治的大变革。在今天，西学东渐仍有其重要的现实意义。作为 A 股市场上以量化投资为研究方向的卖方金融工程团队，在平日的工作中，常常深感海外相关领域的研究水平之高、内容之新。而这也促使我们通过大量的材料阅读，去粗取精，将认为最有价值的海外文献呈现在您的面前！
- 投资经理通常会面对成千上万的上市公司和分析师，一个重要的问题是：如何对分析师给出的推荐数据进行建模并基于模型进行投资。本文应用贝叶斯方法对推荐数据进行建模，实证结果表明这种方法构建的投资组合表现十分优秀。同时当经纪商的推荐结果与使用 IBCC 获得的预测一致时，投资组合表现最佳。本文还将变分近似与独立贝叶斯分类器组合（IBCC）方法相结合，大幅度提升了计算效率。

团队成员：

风险提示：文献中的结果均由相应作者通过历史数据统计、建模和测算完成，在政策、市场环境发生变化时模型存在失效的风险。

请务必阅读正文之后的信息披露和重要声明



目 录

1、引言	3 -
2、数据	4 -
2.1 输入数据：分析师推荐数据	5 -
2.2 预测标的：推荐后股票收益率	5 -
3、IBCC 模型：先验概率与参数构建	6 -
4、变分近似	8 -
4.1 变分贝叶斯推断	8 -
4.2 IBCC 模型中的变分近似	10 -
4.3 从预测概率到投资决策	12 -
5、实证结果	13 -
6、稳健性检验	16 -
6.1 稳健性检验-企业流动性的影响	16 -
6.2 稳健性检查-交易规则的影响	17 -
6.3 稳健性检查-状态变量阈值的影响	18 -
6.4 稳健性检查-持仓长度的影响	19 -
7、结论	20 -
图表 1、数据集结构	6 -
图表 2、IBCC 模型计算过程	9 -
图表 3、决策规则	13 -
图表 4、多头组合（左）和多空组合（右）的表现	14 -
图表 5、2007-2012 年多头，多空和空头投资组合的业绩统计数据	15 -
图表 6、2007-2012 多头、空头及多头/空头投资组合历年表现(%)	15 -
图表 7、空头投资组合相对于 Brok_Flw_S 基准的表现	15 -
图表 8、不同投资组合的规模	16 -
图表 9、按市值划分后两组股票的推荐数量	16 -
图表 10、股票市场按市值分成两组后的模型结果	16 -
图表 11、两组股票的市值分布	17 -
图表 12、固定 k 选择不同 c 的结果	17 -
图表 13、固定 c 选择不同 k 的结果	18 -
图表 14、2007-2012 不同阈值的多头与空头投资组合结果	19 -
图表 15、2007-2012 年多头组合不同持仓长度的表现	19 -

报告正文

如何对分析师预期数据进行建模？——基于贝叶斯方法的研究

文献来源：

Bew, David and Harvey, Campbell R. and Ledford, Anthony and Radnor, Sam and Sinclair, Andrew, Modeling Analysts' Recommendations via Bayesian Machine Learning, The Journal of Financial Data Science Winter 2019, 1 (1) 75-98; DOI: <https://doi.org/10.3905/jfds.2019.1.1.075>

推荐理由：

投资经理通常会面对成千上万的上市公司和分析师，一个重要的问题是：如何对分析师给出的推荐数据进行建模并基于模型进行投资。本文应用贝叶斯方法对推荐数据进行建模，实证结果表明这种方法构建的投资组合表现十分优秀，同时 Both 类策略即当经纪商的推荐和 IBCC 模型预测结果一致时构建的组合表现是最好的。文章将变分近似与独立贝叶斯分类器组合（IBCC）方法相结合，大幅度提升了计算效率。

我们的思考：

在兴业金融工程团队前期报告《基于集成学习算法的量化选股模型研究》中就曾提到：机器学习应用于投资领域存在着众多痛点。但不可否认的是机器学习方法能够更有效的处理许多之前难以解决的问题。如本文应用 IBCC 方法成功对分析师推荐数据进行建模，为投资经理改善投资流程提供了一定帮助。因此，我们认为，各种方法本身并无优劣之分，关键是看投资管理人如何去有效地运用。

1、引言

2009 年，一项特殊的科技项目“行星分类”正式启动。这个项目设立的目标之一是识别出新的超新星，为此项目组邀请了大量的天文学家参与此项目。这些天文学家被要求依照下述标签为一个行星确定分类：1) 很可能是超新星；2) 可能是超新星；3) 不可能是超新星。同时项目组也根据加州理工学院帕洛玛提出的瞬态因子光谱分析给出了官方分类。一个至关重要的问题是：在某一时刻，可能会有众多天文学家对同一行星进行分类，那么如何将这些分类进行融合呢？一个简单的方式是将分类直接均分，但显然这种方法是不完美的（平均法可能会丢失许多信息）；另一种可能的方式是“少数服从多数”，即按照支持人数最多的分类确定最后的标签。然而，这种方法与平均法存在一个共同的问题：忽略了不同天文学家之间的技能差异。

那么有没有办法建立一个考虑了天文学家技能差异的融合系统呢？特别地，我们希望这个系统是动态的，可以随着时间更新天文学家们的技能水平。由 Ghahramani 和 Kim(2003)提出的独立贝叶斯分类器组合(Independent Bayesian Classifier Combination,下文简称为 IBCC)机器学习方法完美的解决了这个问题。辛普森等人(2013)利用 IBCC 对“行星分类”项目数据进行了分析，实验结果令人惊

请务必阅读正文之后的信息披露和重要声明

- 3 -

叹。他们发现，通过 IBCC 方法可以大幅提升分类正确率。在允许 10% 错误率的情况下，使用 IBCC 方法的正确率从 65% 上升至 97%。

同时我们发现在评估分析师意见时，投资经理面临的问题和行星分类有着惊人的相似之处。与“行星分类”项目类似，投资经理面对着成千上万的对象（上市公司）和成千上万的天文学家（分析师）。“行星分类”项目中的分类机制（即：1、很可能是超新星；2、可能是超新星；3、不可能是超新星）与买进、持有和卖出十分相似。此外，分析师之间的技能水平也存在着差异。因此，鉴于 IBCC 方法在“行星分类”项目中大幅提升了正确率，本文将 IBCC 方法拓展到投资领域，具体用于分析师推荐数据的模型建立中。目前已有许多文献针对分析师推荐数据进行了研究，但我们的文章（据我们所知）是第一篇将 IBCC 方法应用于分析师推荐数据建模问题的文章。

IBCC 在大规模机器学习问题中应用广泛的主要原因是：IBCC 提供了一种可拓展的方法，可以将不同来源或不同组的单一分类器结果进行融合。这个特征也使其成为融合不同分析师推荐数据的理想工具。由于需要融合的数据源众多，如何避免过拟合是一个需要重要的考虑因素。**事实上，由于贝叶斯模型并不存在过多需要点估计的参数，陷入过拟合的可能性会小很多。**在贝叶斯模型中我们应用所有未知数的联合后验分布来描述不确定性。在预测时，需要对这个联合后验分布进行积分，这个过程中我们不需要对参数给出固定的点估计。但**贝叶斯方法一大缺点为计算复杂度较高**。因此，本文采用了变分近似对传统贝叶斯方法进行改进，这种方法在计算上十分高效。另外虽然我们的方法是可以对每个分析师或任意分组级别进行建模的，但目前我们能够得到的报告都以经纪商（即分析师被雇佣的公司）为单位发布，因此本文将分析师推荐数据按经纪商进行分组。

我们发现即使拥有了大量的数据和复杂的机器学习模型，预测股票收益率依然很困难。因此我们的主要目标是给出一个整体的预测模型，其中每个经纪商的权重由经纪商历史推荐数据的长度和质量决定。本文研究结果表明当经纪商的推荐结果与使用 IBCC 获得的预测一致时，投资组合表现最佳。这个规律不仅存在于多头、多空与空头投资组合，在我们所进行的各种稳健性检验中也十分明显。本文结果为基金经理们（包括量化基金经理和主动管理基金经理）提供了一种利用机器学习进行投资决策的框架。

本文架构如下：第二部分叙述了构建模型所需的数据，重点讨论了数据的特性，如数据的稀疏性、相关性等。第三部分详细介绍了 IBCC 模型，并对贝叶斯框架中先验分布和超参数的选择进行探讨。第四部分解释了如何利用变分近似法进行计算，这个方法极大的提升了模型的计算效率。第五、六部分给出了实证结果和一系列稳健性检验。最后第七部分给出了本文结论并提出进一步研究的建议。

2、数据

本文选用的模型属于机器学习中的有监督模型。其中模型输入数据为分析师推荐数据，而预测标的（即输出数据）是在分析师推荐之后该股票（或股票相对于基准）的收益率。我们希望以分析师推荐数据为基础，利用 IBCC 方法对股票收

益率进行预测。

2.1 输入数据：分析师推荐数据

我们采用了 Thomson Reuters I/B/E/S 数据库中分析师关于个股和行业的推荐数据，该数据库涵盖了几乎所有可得的分析家推荐数据。虽然不同分析家的评级方法不同，但整体上来讲分析家的推荐可映射到五个分类：强劲买入、买入、持有、表现不佳和卖出。

分析师推荐数据和我们常用的金融数据有两个很明显的不同。首先，分析师推荐数据并不是以固定的频率观察到的，它的观察值一般是不可预测的离散日期的数据。其次，推荐数据不是定量数据，而是分类数据。这使得它们更适合被应用于分类模型而非回归模型。此外，Thomson Reuters 数据库有以下几个特点：

1) 高维数据：数据库是由成千上万的分析师对于成千上万只股票的推荐数据构成的，因此具有较高的维度。

2) 稀疏性：通常只有少数分析师会在某一天就某只股票给出建议，这就造成了数据十分稀疏。

3) 相关性：我们认为分析师的推荐数据在统计上是相关的，原因如下：

- 横断面相关：分析师接触到的信息通常是相关的。因此即便他们的决策过程是独立的，也能得出相同或相似的结论。
- 时间序列相关：分析师的观点通常是逐渐更新的，并且分析师经常重申他们之前的建议。这就导致了数据存在序列相关性。同时分析师之间的群体行为也可能产生序列相关性(例如，一些分析师率先提出观点，另一些分析师随之提出观点)。

4) 缺乏一致性：虽然 Thomson Reuters 数据库将分析师的推荐映射到了五个分类，但对许多分析师来说常用的分类只有其中的几类。例如一些分析师只给出买入、持有这两个评级。本文将分析师推荐数据分类中的“强劲买入”、“买入”合并标记为买入，“表现不佳”、“卖出”合并标记为卖出。同时将没有推荐数据的值标记为缺失(Missing)。这意味着我们有了四类标签：买入、持有、卖出和缺失，即对于任意的分析师(或经纪商) k ，给出的推荐结果 B_k 有缺失(0)，持有(1)，卖出(2)与买入(3)四种取值。最后，我们注意到推荐数据的分布极不均衡，这可能与分析师对于卖空评级的固有偏见有关。

在传统的贝叶斯模型中，针对这四类数据特征需要分别设立一个概率分布来建模。而我们的 IBCC 方法有下面三个优点：1) IBCC 可以同时解决所有问题，并且保持了较快的计算速度；2) 模型可以根据当前的信息进行动态调整，同时能够进行实时计算；3) IBCC 模型允许容纳之前的信息，以便在需要时，模型可以由观测数据之外的信息来进行调整。

2.2 预测标的：推荐后股票收益率

本文研究的预测标的为分析师推荐后该股票在未来某个时间范围内的收益率。由于 IBCC 模型需要预测标的的数据形式为分类数据，我们首先将收益率(本

请务必阅读正文之后的信息披露和重要声明

- 5 -

文采用超额收益率)转换为分类数据,转换步骤如下:

第一步:计算股票的超额收益。具体计算过程为:假设计算股票收益率的时间范围为 $\Delta\tau$,分析师给出股票推荐数据的日期为 s 。我们计算 s 下一个工作日之后 $\Delta\tau$ 期间股票的超额收益,记为 $r(s, \Delta\tau)$ 。(下文默认取 $\Delta\tau=60$)。

第二步:利用分段函数将收益率转换为分类数据。我们结合指数变动率将股票收益率分为:下跌(Price_Down),波动(Price_Flat)和Price_Up(上涨)三种状态,具体计算方法如式(2-1)所示:

$$R = \begin{cases} 0, & \text{if } r_{(s, \Delta\tau)} \leq -5\% \times RVol \\ 2, & \text{if } r_{(s, \Delta\tau)} \geq 5\% \times RVol \\ 1, & \text{else} \end{cases} \quad (2-1)$$

其中 $RVol$ 表示基准指数波动率,股票收益率状态变量 R 的取值 $r \in \{0, 1, 2\}$,其中 $0, 1, 2$ 分别表示下跌(Price_Down),波动(Price_Flat)和Price_Up(上涨)。需要注意的是,这里定义的股票收益率状态变量取值和分析师的建议是否正确没有关系,它仅取决于分析师推荐后该股票一段时间内收益率的大小。

上文我们成功将股票收益率数据处理成IBCC模型适用的预测标的的形式。下面我们对研究的一些参数进行设定:

- 研究时间为2004.1.1-2013.1.1
- 所选股票主要分布在欧洲部分地区,如奥地利,比利时,捷克共和国,塞浦路斯,丹麦,芬兰,法国,德国,希腊,匈牙利,爱尔兰等泛欧地区,意大利,卢森堡,荷兰,挪威,波兰,葡萄牙,俄罗斯,西班牙,瑞典,瑞士,土耳其和英国。
- 选用的基准为道琼斯欧元斯托克斯指数(Dow Jones Euro Stoxx Index)。

同时我们将分析师按公司雇主(即经纪商)进行分组。需要明确的是,IBCC模型是可以应用于个人分析师级别或经纪商级别的,但本文仅研究经纪商级别。

图表1展示了我们汇总得到的部分数据。数据集总共有105,319行,每一行汇总了347个经纪商的推荐数据,理论上应该有36,545,693个数据。然而实际上,我们只有116,220个数据是非空的,意味着99.7%的数据对应的标签是“Missing”。这表明数据具有极端的稀疏性,也是IBCC方法分析的核心之一。

图表1、数据集结构

Stock ID	Date	Truth	Broker 1	Broker 2	...	Broker N
#1234	July 10, 2008	0	3	0		0
#5678	Feb 7, 2012	0	0	1		2
#5678	July 1, 2012	2	2	0		0
#5678	Mar 14, 2012	1	0	3		1
		⋮	⋮	⋮	⋮	⋮

资料来源: The Journal of Financial Data Science, 兴业证券经济与金融研究院整理

3、IBCC模型: 先验概率与参数构建

IBCC模型是一个全概率模型,使用的数据为分类数据。如图表1所展示的那样,每行包括:1)某日各家经纪商的推荐数据(输入数据,取值为 $\{0, 1, 2, 3\}$,分别代表缺失、持有、卖出与买入);2)股票收益率的状态变量(预测标的,取值

为{0,1,2}，分别代表下跌、波动、上涨)。

首先我们给出股票收益率状态变量 R 的概率质量函数 (PMF):

$$\Pr(R=r|\kappa)=\kappa_r \quad \text{其中 } r \in \{0,1,2\} \quad (3-1)$$

其中 R 为收益率状态变量，取值有 0,1,2 三种情况，分别代表下跌、波动与上涨； $\kappa=\{\kappa_0, \kappa_1, \kappa_2\}$ 代表一个已知的三维向量， $\kappa_0, \kappa_1, \kappa_2$ 表示 R 三种取值出现的概率，且满足 $\kappa_0+\kappa_1+\kappa_2=1$ 。这里我们假设 0,1,2 这三种情况是独立的，且出现其他结果的概率为 0。

接下来我们给出每个分析师在给定不同 R 的取值时所选推荐行为的概率分布。即假设 $B_k | R=0, B_k | R=1, B_k | R=2$ 这三个条件变量的概率质量函数为：

$$\Pr(B_k=b_k|R=r,\pi_r^{(k)})=\pi_{r,b_k}^{(k)} \quad b_k \in \{0,1,2,3\}, k \in \{1,...,N\} \quad (3-2)$$

其中 N 代表分析师总数，每个分析师 k 的推荐数据 B_k 取值为 b_k ， $\pi_r^{(k)}=[\pi_{r0}^{(k)}, \pi_{r1}^{(k)}, \pi_{r2}^{(k)}, \pi_{r3}^{(k)}]$ 为一个四维向量，代表了分析师 k 在给定 $R=r$ ， $r \in \{0,1,2\}$ 的条件下作出四种推荐行为的概率，且满足 $\pi_{r0}^{(k)}+\pi_{r1}^{(k)}+\pi_{r2}^{(k)}+\pi_{r3}^{(k)}=1$ 。以上，我们为每个分析师定义了一个独立的四维分布。

根据式 (3-2) 我们给出第 j 行数据当 $R_j=r_j$ 时的概率分布，如下所示：

$$\begin{aligned} \Pr(R_j=r_j, B_1=b_{1j}, B_2=b_{2j}, ..., B_N=b_{Nj}) \\ = \kappa_{r_j} \pi_{r_j b_{1j}}^{(1)} \pi_{r_j b_{2j}}^{(2)} ... \pi_{r_j b_{Nj}}^{(N)} = \kappa_{r_j} \prod_{k=1}^N \pi_{r_j b_{kj}}^{(k)} \end{aligned} \quad (3-3)$$

其中下标 j 表示各参数在第 j 行数据的取值，如 r_j 表示 R 变量在第 j 行的取值， b_{kj} 为分析师 k 在第 j 行的推荐数据取值， $j \in \{1,...,n\}$ ， n 为数据的总行数。

IBCC 中假设所有数据都是独立的，因此给出所有数据的全概率分布为：

$$\Pr(\mathbf{r}, \mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_N) = \prod_{j=1}^n (\kappa_{r_j} \prod_{k=1}^N \pi_{r_j b_{kj}}^{(k)}) \quad (3-4)$$

其中 $\mathbf{r}=(r_1, ..., r_n)$ 代表了每一行的 R 变量取值组成的向量，同时 $\mathbf{b}_k=(b_{k1}, ..., b_{kn})$ 代表每一行的分析师 k 推荐数据取值组成的向量， $k \in \{1,...,N\}$ 。

目前为止，我们已经给出了将参数 κ 和 $(\pi_0^{(k)}, \pi_1^{(k)}, \pi_2^{(k)})$ 作为常数的股票收益率状态变量和分析师推荐数据取值的概率质量分布(式(3-1)与式(3-2))。通常，我们需要采用最大似然法等方法来估计这些参数。但在贝叶斯方法中，我们将假设这些参数的先验分布，并以此为基础推导后验分布。本文我们采用 Dirichlet 分布作为参数的先验分布。

本文假设参数 $\kappa=\{\kappa_0, \kappa_1, \kappa_2\}$ 满足三维狄利克雷分布 (Dirichlet distribution)，即在定义域：

$$D=\{(\kappa_0, \kappa_1, \kappa_2); 0 \leq \kappa_i \leq 1, \sum_{i=0}^2 \kappa_i = 1\}$$

上的概率密度函数为：

$$\Pr(\mathbf{\kappa}|\mathbf{v}) = C(\mathbf{v}) \prod_{r=0}^2 (\kappa_t^{v_{0r}-1}) \quad (3-5),$$

$$C(\mathbf{v}) = \Gamma(v_{00} + v_{01} + v_{02}) / \{\Gamma(v_{00}) \times \Gamma(v_{01}) \times \Gamma(v_{02})\}$$

其中 $\mathbf{v} = (v_{00}, v_{01}, v_{02})$ 称为超参数 (hyperparameter), $\Gamma(\cdot)$ 为伽马函数。若 $v_{00} = v_{01} = v_{02} = 1$, 则会在定义域 D 上得到一个 $\mathbf{\kappa}$ 的扁平先验 (flat prior)。类似地, 我们假设 $\{\pi_{r0}^{(k)}, \pi_{r1}^{(k)}, \pi_{r2}^{(k)}, \pi_{r3}^{(k)}\}$ 满足四维狄利克雷分布, 同时 $\{\alpha_{0,r0}^{(k)}, \alpha_{0,r1}^{(k)}, \alpha_{0,r2}^{(k)}, \alpha_{0,r3}^{(k)}\}$ 为其超参数。下文, 为了简便, 我们将分析师在给定 R 的取值时推荐数据的信息集表示为:

$$\Pi = [\{\pi_{r0}^{(k)}, \pi_{r1}^{(k)}, \pi_{r2}^{(k)}, \pi_{r3}^{(k)}\} : r=0,1,2; k=1,\dots,N]$$

同时参数集为:

$$\mathbf{A}_0 = [\{\alpha_{0,r0}^{(k)}, \alpha_{0,r1}^{(k)}, \alpha_{0,r2}^{(k)}, \alpha_{0,r3}^{(k)}\} : r=0,1,2; k=1,\dots,N]。$$

在假设了先验分布和似然函数之后, 我们可以构建后验分布如下:

$$\begin{aligned} & \Pr(\mathbf{\kappa}, \Pi, \mathbf{r}, \mathbf{b}_1, \dots, \mathbf{b}_N | \mathbf{A}_0, \mathbf{v}) \\ & \propto \prod_{j=1}^n (\kappa_{r_j} \prod_{k=1}^N \pi_{r_j b_{kj}}^{(k)}) \Pr(\mathbf{\kappa} | \mathbf{v}) \Pr(\Pi | \mathbf{A}_0) \end{aligned} \quad (3-6)$$

这是一个高维的联合概率分布。本文选择定义域上的扁平先验以保证参数的选择对最终结果的影响较小, 更有利于评估我们的框架。

最后需要注意的是: IBCC 没有在类别数据中进行排序。即模型只给出了一组标签 (分析师的推荐) 如何和另一组数据 (股票收益率) 相关联。同时数据的高维度和稀疏性意味着用其他可选的相关模型 (如 copulas) 是不可行的。IBCC 模型假设条件独立从而处理我们特殊的数据, 提供了可拓展且计算上高效率的方法。IBCC 的这一关键特性是其在大型贝叶斯机器学习应用中十分流行的原因之一。

4、变分近似

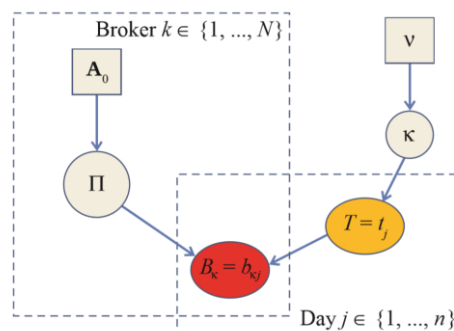
4.1 变分贝叶斯推断

在本节中, 我们将介绍变分贝叶斯推断, 也称为变分贝叶斯 (VB)。借鉴 Bishop(2006, 第 10 章)、Blei, Kucukelbir 和 McAuliffe(2018) 的详细推导并结合 Fox 和 Roberts(2011) 的资料之后, 我们给出将 VB 应用于 IBCC 模型的主要结果。这个模型从理论上来看是比较简单的, 下面我们对其进行数学推导。下文我们使用的所有对数都是自然对数, 即 $\log_e(\cdot)$ 。

设 \mathbf{X} 表示一组观测到的数据, \mathbf{Z} 表示潜在 (即未观察到的) 参数和变量的组合集。 $p(\cdot)$ 表示某一变量的概率分布模型。例如, \mathbf{X} 和 \mathbf{Z} 的联合分布写为 $p(\mathbf{X}, \mathbf{Z})$ 。我们的目标是为后验分布 $p(\mathbf{X}|\mathbf{Z})$ 找到一个较好的近似值 $q(\mathbf{Z})$ 。在我们的 IBCC 方法中, \mathbf{Z} 将代表我们的预测标的的数据, 有三种取值 Price_Up, Price_Down 或

Price_Flat。图表 2 展示了 IBCC 模型的具体计算过程。

图表 2、IBCC 模型计算过程



资料来源：The Journal of Financial Data Science, 兴业证券经济与金融研究院整理

由于 $q(\mathbf{Z})$ 表示概率密度函数则其积分为 1，可得：

$$\log p(\mathbf{X}) = \int q(\mathbf{Z}) \log p(\mathbf{X}) d\mathbf{Z} \quad (4-1)$$

其中 $p(\mathbf{X})$ 被称为模型证据 (model evidence)。此外，由条件概率的定义可得：

$$p(\mathbf{X}) = p(\mathbf{X}, \mathbf{Z}) / P(\mathbf{Z} | \mathbf{X}) \quad (4-2)$$

我们将式 (4-2) 代入式 (4-1)，可得：

$$\begin{aligned} \log p(\mathbf{X}) &= \int q(\mathbf{Z}) \log \left\{ \frac{p(\mathbf{X}, \mathbf{Z})}{P(\mathbf{Z} | \mathbf{X})} \right\} d\mathbf{Z} \\ &= \int q(\mathbf{Z}) \log \left\{ \frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})} \times \frac{p(\mathbf{Z})}{q(\mathbf{Z} | \mathbf{X})} \right\} d\mathbf{Z} \end{aligned} \quad (4-3)$$

将式 (4-3) 简写为：

$$\log p(\mathbf{X}) \Rightarrow L(q) = \int q(\mathbf{Z}) \log \left\{ \frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})} \right\} d\mathbf{Z} + KL(q, p) \quad (4-4)$$

其中 $KL(q, p) = -\int q(\mathbf{Z}) \log \{p(\mathbf{Z} | \mathbf{X}) / q(\mathbf{Z})\} d\mathbf{Z}$ ，称为 $q(\mathbf{Z})$ 和 $p(\mathbf{Z} | \mathbf{X})$ 之间的 KL 距离 (KL-divergence)。 $L(q) = \int q(\mathbf{Z}) \log \{p(\mathbf{X} | \mathbf{Z}) / q(\mathbf{Z})\} d\mathbf{Z}$ ，称为变分自由度 (variational free energy)。

KL 距离具有以下性质：1) 非负性；2) 当且仅当 $q(\mathbf{Z}) = p(\mathbf{Z} | \mathbf{X})$ 时， $KL(q, p) = 0$ 。这意味 $L(q)$ 是 $\log p(\mathbf{X})$ 的下界。同时由于 KL 距离是 $q(\mathbf{Z})$ 的函数，因此可以通过最小化 KL 距离来最大化下界 $L(q)$ 。这被称为一个变分问题。

为了解决上文叙述的变分问题。我们首先用一个受约束但易处理的分布族表示 $q(\mathbf{Z})$ ，然后寻找最大化 $L(q)$ 的族元素。我们将 \mathbf{Z} 划分为 m 组变量，并假设 $q(\mathbf{Z})$ 可以通过：

$$q(\mathbf{Z}) = \prod_{i=1}^m q_i(\mathbf{Z}_i) \quad (4-5)$$

来近似。这个变分近似方法源于物理学，被称为平均场理论。接下来，我们在 $q(\mathbf{Z})$ 所有分布中寻找能够使 $L(q)$ 最大的分布，设这个分布为 $q_i^*(\mathbf{Z}_i)$ 。

我们将式 (4-5) (为了简便，记 $q_i(\mathbf{Z}_i) = q_i$) 代入 $L(q)$ 的表达式中，则可以

得到：

$$L(q) = \int \Pi_i q_i \{ \log p(\mathbf{X}, \mathbf{Z}) - \sum_i \log q_i \} d\mathbf{Z} \quad (4-6)$$

为了更明显的看出 $L(q)$ 与哪些因素相关，我们将式(4-6)改写为：

$$L(q) = \int q_j \{ \int \log p(\mathbf{X}, \mathbf{Z}) \Pi_{i \neq j} q_i d\mathbf{Z}_i \} d\mathbf{Z}_j - \int q_j \log q_j d\mathbf{Z}_j + \text{Constant} \quad (4-7)$$

下面我们首先定义 $\log \tilde{p}(\mathbf{X}, \mathbf{Z}_j) = E_{i \neq j}[\log p(\mathbf{X}, \mathbf{Z})] + c$ ， c 为常数。其中 $E_{i \neq j}[\cdot]$ 表示所有 $i \neq j$ 的 q_i 分布的期望，则可以得到

$$E_{i \neq j}[\log p(\mathbf{X}, \mathbf{Z})] = \int \log p(\mathbf{X}, \mathbf{Z}) \Pi_{i \neq j} q_i d\mathbf{Z}_i \quad (4-8)$$

结合式 (4-7) 与式 (4-8) 可以看出 $L(q)$ 是 $q_j(\mathbf{Z}_j)$ 与 $\tilde{p}(\mathbf{X}, \mathbf{Z}_j)$ 之间 KL 距离的负数，则可以通过使 $q_j(\mathbf{Z}_j) = \tilde{p}(\mathbf{X}, \mathbf{Z}_j)$ 最小化。因此，对每个 $i \neq j$ ，若假设 q_i 为常数，我们可以通过取：

$$\log q_j^*(\mathbf{Z}_j) = \log \tilde{p}(\mathbf{X}, \mathbf{Z}_j) + c \quad (4-9)$$

达到 $L(q)$ 的最大值。这一关键结果为变分法的应用提供了基础。

式 (4-9) 给出了 $L(q)$ 基于式 (4-5) 的假设下的最大化条件。然而，这种方程没有提供显式解，因为估计 $q_j^*(\mathbf{Z}_j)$ 的计算中也需要用到 $q_i^*(\mathbf{Z}_i)$ 的分布。因此我们采用迭代的方式来解这些方程：

- 第一步：设置 $q_i(\mathbf{Z}_i)$ 的初始值， $i \in \{1, \dots, m\}$ ；
- 第二步：对每一个 j ，我们利用最新的 $q_i(\mathbf{Z}_i)$ ($i \neq j$) 的估计值来计算 $E_{i \neq j}[\log p(\mathbf{X}, \mathbf{Z})]$ ，再根据 $E_{i \neq j}[\log p(\mathbf{X}, \mathbf{Z})]$ 的值计算 $q_j(\mathbf{Z}_j)$ ， $j \in \{1, \dots, m\}$ ；
- 第三步：更新 $q_i(\mathbf{Z}_i)$
- 第四步：重复第二、三步，直到收敛。

由于 $L(q)$ 是 $q_i(\mathbf{Z}_i)$ 的凸函数，局部最大值一定存在。此外，在 IBCC 方法中我们选用的分布为指数族，因此这个最大值也是全局最大值。

4.2 IBCC 模型中的变分近似

本文应用的 IBCC 模型和 Kim 和 Ghahramani (2012) 以及 Simpson 等人的模型不同。首先，我们采用的是实时预测，时间一致性要求我们每次预测时仅使用已有的信息来建立模型。第二，除了最近 $\Delta\tau$ 期提出的推荐数据所对应的股票收益率外，我们训练数据中的所有历史股票收益率数据都能被完全观察到，这主要是因为我们的训练数据都源于公开的可用的价格。最后，我们的主要目的是为了预测：

$$\Pr(R=r | B_1=b_1, B_2=b_2, \dots, B_N=b_N) \quad (4-10)$$

而不是后验概率。

虽然可以应用 Simpson 等人的做法，将明确的时间结构拓展进 IBCC 模型中，

请务必阅读正文之后的信息披露和重要声明

但本文选择基于随时间更新的静态模型来进行研究，这种方法相对来说更简单。具体而言，我们在每个评估时间点计算分析师推荐数据和股票收益率取值数据，另外确保在最后的观察值和计算数据之间有 $\Delta \tau$ 的时间。对于用这种方法创建的每个训练数据集，我们计算式 (4-10) 的分布。同时我们所有的研究结果都通过滚动方式获得。

对于每个评估日期，我们都进行了扩展窗口和移动窗口分析。扩展窗口是计算从 2004.1.1 到评估日期的所有数据，而移动窗口是计算每个评估日期过去三年的数据。理论上，评估日期可以选择每一个工作日，但出于实际原因，我们将它设为每季度即 3 月、6 月、9 月和 12 月的第一天。

$i \in (1, \dots, n_r)$ 表示训练数据的行，每改变一次滚动窗口时都重新编号。由于训练数据包括了所有的推荐数据和股票收益率数据，下面我们将式 (4-11) 作为估计式 (4-10) 的起点：

$$\Pr(\kappa, \Pi, r, b_1, b_2, \dots, b_N | \mathbf{A}^*, \mathbf{v}^*) \quad (4-11)$$

这里的符号定义与上文相同，同时 $\mathbf{A}^*, \mathbf{v}^*$ 与前文类似，表示了 Π, κ 的参数集。

式 (4-11) 和式 (3-6) 拥有一样的结构，除了在式 (4-10) 中，股票收益率状态变量的取值 r 不是已知的。

$$\Pr(\kappa, \Pi, r, b_1, \dots, b_N | \mathbf{A}_0, \mathbf{v})$$

$$\propto \prod_{j=1}^n (\kappa_{r_j} \prod_{k=1}^N \pi_{r_j b_{kj}}^{(k)}) \Pr(\kappa | \mathbf{v}) \Pr(\Pi | \mathbf{A}_0) \quad (3-6)$$

进一步，我们将 $\log \Pr(\kappa, \Pi, r, b_1, b_2, \dots, b_N | \mathbf{A}^*, \mathbf{v}^*)$ 表示为下面这种形式：

$$\sum_{s=0}^2 I(r=s) (\log \kappa_s + \sum_{k=1}^N \log \pi_{sb_k}^{(k)}) + \log \Pr(\kappa | \mathbf{v}^*)$$

$$+ \log \Pr(\Pi | \mathbf{A}^*) + \text{Constant} \quad (4-12)$$

在这里，我们引入指标函数 $I(\cdot)$ ，当 $r=s$ 时， $I(r=s)=1$ ；否则， $I(r=s)=0$ 。

接下来我们假设可以对 $q(\mathbf{Z})$ 进行因式分解，即认为：

$$q(r, \kappa, \Pi) = q(r) q(\kappa, \Pi) \quad (4-13)$$

由式 (4-12) 可知 κ 和 Π 独立，则意味着我们可以得到

$$q^*(\kappa, \Pi) = q^*(\kappa) * q^*(\Pi)$$

接下来为了得到 $q^*(r)$ ，我们需要得到：

$$\log q^*(r) = E_{\kappa, \Pi} [\log p(r, \kappa, \Pi) + \text{constant}]$$

结合式 (4-13)，可以将上式转换为式 (4-14)

$$\log q^*(r) = E_{\kappa} \log \kappa_r + \sum_{k=1}^N E \pi_r^{(k)} \log \pi_{rb_k}^{(k)} + \text{Constant} \quad (4-14)$$

同时根据 Dirichlet 分布的性质，我们可以得到：

$$E \kappa \log \kappa_r = \Psi(\nu_r^*) - \Psi\left(\sum_{j=0}^2 \nu_j^*\right)$$

$$E\pi_r^{(k)} \log \pi_{rb_k}^{(k)} = \Psi(\alpha_{rb_k}^{(k*)}) - \Psi\left(\sum_{s=0}^3 \alpha_{rs}^{(k*)}\right)$$

其中 $\Psi(\cdot)$ 为 DiGamma 函数。接下来我们定义：

$$\log \rho_r = \Psi(v_r^*) - \Psi\left(\sum_{j=0}^2 v_j^*\right) + \sum_{k=1}^N (\Psi(\alpha_{rb_k}^{(k*)}) - \Psi\left(\sum_{s=0}^3 \alpha_{rs}^{(k*)}\right)) \quad (4-15)$$

可以得到：

$$q^*(r) = \rho_r / (\rho_0 + \rho_1 + \rho_2),$$

由此我们有了 $\Pr(R=r|B_1=b_1, B_2=b_2, \dots, B_N=b_N)$ 的初始估计。

下面我们基于最新得到的 $q^*(r)$ 对 $q^*(\kappa)$ 和 $q^*(\Pi)$ 进行更新。首先计算 κ ，过程如下：对于任意 X ，记 $I(X)$ 的期望为 $\Pr(X)$ ，则

$$\log q^*(\kappa) = \sum_{s=0}^2 q(r=s) \log \kappa_s + \sum_{s=0}^2 (v_s^* - 1) \log \kappa_s + \text{Constant} \quad (4-16)$$

将式 (4-16) 中的 $\log \kappa_j$ 作为一个整体，我们可以认为 $q^*(\kappa)$ 为 Dirichlet 分布，参数满足 $v_j = v_j^* + q(r=j)$ ，其中 $j \in \{0, 1, 2\}$ 。

根据式 (4-16) 我们可以根据 $q^*(r)$ 更新 $q^*(\kappa)$ ，下面我们开始更新 $q^*(\Pi)$ ，过程如下：由式 (4-11)，我们得到进一步分解：

$$q^*(\Pi) = \prod_{k=1}^N \prod_{s=0}^2 q^*(\pi_s^{(k)})$$

提取 $\pi_s^{(k)}$ 项，同时对 $q^*(r)$ 取期望可以得到

$$q(\pi_s^{(k)}) = \sum_{j=0}^2 q^*(r=j) \sum_{l=1}^N \log \pi_{rb_l}^{(l)} + \sum_{q=0}^3 (\alpha_{sq}^{(k*)} - 1) \log \pi_{sq}^{(k)} + \text{Constant} \quad (4-17)$$

将式 (4-17) 中的 $\log \pi_{rb_l}^{(l)}$ 作为一个整体，我们可以认为 $q(\pi_s^{(k)})$ 为 Dirichlet 分布，参数满足 $\alpha_{jb}^{(k)} = q(r=j) I(b=b_k) + \alpha_{jb}^{(k*)}$ ，其中 $b \in \{0, 1, 2, 3\}$ 同时 $k \in \{1, \dots, N\}$ 。

根据式 (4-16) 与式 (4-17)，我们可以根据已经得到了 $q^*(r)$ 更新 $q^*(\kappa)$ 和 $q^*(\Pi)$ 。接下来，我们利用 $q^*(\kappa)$ 与 $q^*(\Pi)$ 来更新 $q^*(r)$ ，不停迭代，直到收敛。最后得到我们想要的：

$$\Pr(R=r|B_1=b_1, B_2=b_2, \dots, B_N=b_N)$$

的估计，即股票收益率数据取值的 VB 近似。

虽然这里我们推导了单个预测的估计过程，但实际上这个方法可以同时计算多个推荐的真实分布。同时 VB 的迭代过程类似于期望最大化算法 (EM) 的计算过程，但 VB 算法和 EM 算法目的其实完全不同：EM 是通过最大似然函数来获得参数的点估计，但 VB 提供了分布的全局近似。

4.3 从预测概率到投资决策

根据上一节我们得到了股票收益率数据取值的概率分布 (q_0, q_1, q_2) ，其中

请务必阅读正文之后的信息披露和重要声明

q_0, q_1, q_2 分别表示每个经纪商推荐后股票收益率状态变量出现 Price_Down, Price_Flat 和 Price_Up 三种情况的概率。即使已经计算出这些概率，人们仍然需要一个决策规则，即决定采取什么样的行动。

我们主要研究三个离散行动集：Go_Short, No_Trade 和 Go_Long。直接根据 Price_Down, Price_Flat 或 Price_Up 三者的概率进行选择是比较直观的想法。例如，当 $q_0 > \max(q_1, q_2)$ ，则选择 Go_Short，当 $q_2 > \max(q_0, q_1)$ 则选择 Go_Long，否则选择 No_Trade。但这个规则会导致 Go_Long 的操作过多。因为分析师通常会倾向于给出买入建议，因此 Price_Up 标签比 Price_Down 或 Price_Flat 更多。

记 q_t 是条件概率 $\Pr(R=r|B_1=b_1, B_2=b_2, \dots, B_N=b_N)$ 的估计。本文的决策规则为当 q_t 超过了 $R=r$ 的无条件概率，即 κ_t 才采取动作，具体规则如图表 3 所示：

图表 3、决策规则

Decision	Trigger Condition
Go_Short	$q_0/\kappa_0 > c$ and $q_0 > k \max(q_1, q_2)$
No_Trade	otherwise
Go_Long	$q_2/\kappa_2 > c$ and $q_2 > k \max(q_0, q_1)$

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

参数 c 和 k 都会对交易规则造成影响，但它们的影响是不同的。参数 c 影响 R 变量的条件概率和无条件概率的比较。因此，在保持 $k=1$ 固定的同时增加 c 意味着需要推荐数据产生更高的信息值才会采取 Go-Long (Go-Short) 决策。相反，参数 k 的大小影响 R 变量三个取值的条件概率 q_0, q_1 和 q_2 之间的比较，但与无条件概率无关。因此，在维持 $c=1$ 的同时增加 k 会提高决策时采取 Go_Long (Go-Short) 行动所需的阈值。

5、实证结果

本文应用的 IBCC 方法有两个重要特征：

1) **信息池的可选择性**。本文的实证结果都是基于将分析师按经纪商（即他们声明的雇主或所属公司）进行分组而得到的。事实上，IBCC 可以应用于不同的分组，例如可以选择在个体分析师上建立模型。这个特征意味着模型可以在全球行业分类或者任意级别上构建。当然一些更小的级别的模型计算复杂度可能会提升。我们在此报告中不对这一点进行深入研究。

2) **我们应用的 IBCC 模型能够在不引入额外参数的同时将每个股票的多个推荐进行组合**。具体地，我们将推荐数据按 30 天进行组合，即当一个新的推荐数据出现时，我们回顾过去 30 天的窗口内其他经纪商的最新推荐数据，将它们组合在一行数据中。这样做有利于增加训练数据中每一行出现的推荐数。

同时本文在三年的样本周期内构建模型，然后将此模型应用于下一季度的预测中。然后依据预测概率进行决策，本文假设 $c=k=1$ ，之后也会验证这两个参数对于结果的影响。

在评估方面，我们将 IBCC 的模型结果与最初的值（即遵循每个经纪商的推

荐) 进行比较。这个基准也称为 Brok_Flw, 主要用于评价 IBCC 模型的附加价值。

Brok_Flw 基准构造方法如下:

1. 对于每个经纪商的买入(卖出)建议, 创建一个持续 60 个工作日的+1 (-1) 信号。
2. 将上面产生的信号按股票为单位进行加总, 产生结果信号。
3. 将结果信号的上限/下限设置为±10。例如将大于+10 的值设置为+10。
4. 对于只做多头的投资组合, 只包括多头建议; 相反, 只做空头的投资组合, 只包含空头建议。
5. 我们按保持 100 美元的总风险敞口进行投资组合每日再平衡, 换言之:

$$\text{position}_{it} = \text{signal}_{it} / \sum_i |\text{signal}_{it}|$$

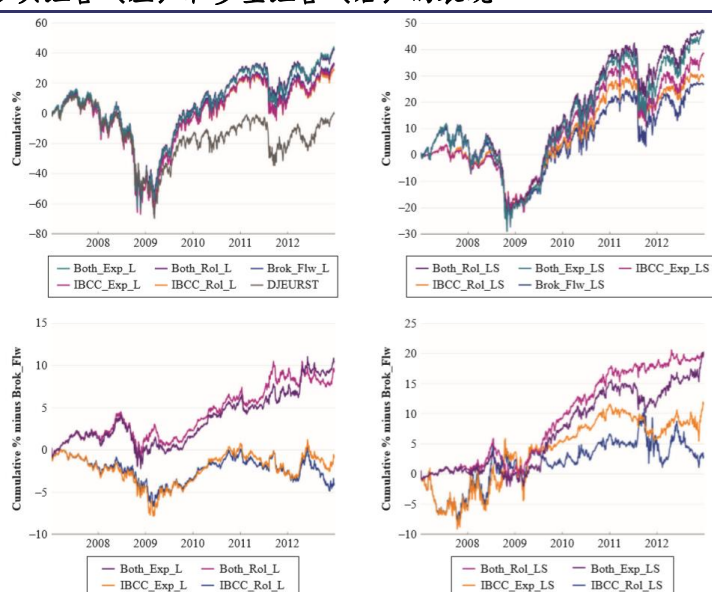
为了之后的结果展示, 我们统一一些命名规则:

- Brok_Flw_LS: 这是前面提到的经纪商跟踪基准。
- IBCC_Rol_LS: 在这里, 我们应用 IBCC 算法, 拟合一个三年滚动窗口, 包括多头和空头头寸。
- IBCC_Exp_LS: 和 IBCC_Rol_LS 类似, 但估计是在扩展窗口上执行的。
- Both_Rol_LS: 表示当单个经纪商的 IBCC 的结果和 Brok_Flw 信号一致时才采取买入(卖出)操作。这可以有效的解决 IBCC 与 Brok_Flw 信号矛盾的情况。同时计算采取三年滚动窗口。
- Both_Exp_LS: 与 Both_Rol_LS 类似, 但估计是在扩展窗口上执行的。

当仅允许多头(空头)时, 使用 L(S) 代替 LS。

图表 4 展示了多头和多空投资组合的自身表现和相对于 Brok_Flw 基准的表现。在全球金融危机期间, 所有多头投资组合都表现不佳, 但从 2009 开始, 其表现优于 DJEURST 指数。多头 IBCC 策略与 Brok_Flw_L 基准表现差距不大, 表现最好的往往是 Both 类(如 Both_Rol_L、Both_Exp_L)策略。

图表 4、多头组合(左)和多空组合(右)的表现



资料来源: The Journal of Financial Data Science, 兴业证券经济与金融研究院整理

请务必阅读正文之后的信息披露和重要声明

所有应用 IBCC 方法构建的多空投资组合表现均优于 Brok_Flw_LS 基准。同样，Both 类的投资组合表现最佳。意味着当 IBCC 模型结果和经纪商推荐数据一致时进行操作的方法有助于帮助投资经理提升管理业绩。

图表 5，图表 6 中提供了分年度的所有多头，多空和空头投资组合的结果。Brok_Flw_S 基准和两种空头 IBCC 策略均为亏损，因此我们不关注他们的直接表现。值得注意的是，Both 类的空头投资组合依旧表现更好，与之前在多头和多空的投資中的结果一致。图表 7 展示了空头投资组合的结果。

图表 5、2007-2012 年多头，多空和空头投资组合的业绩统计数据

Side	Model	Mean	Vol	Alpha	Alpha t-Stat	Beta	Beta t-Stat	Turnover
Long-Only	Brok_Flw_L	5.43	24.18	5.47	2.73	1.01	26.97	5.75
	IBCC_Rol_L	4.77	24.66	4.91	2.09	1.02	23.36	5.74
	IBCC_Exp_L	5.30	24.89	5.39	2.27	1.03	23.63	5.68
	Both_Rol_L	6.99	24.51	7.06	2.75	1.00	20.18	6.13
Short-Only	Both_Exp_L	7.13	24.71	7.28	2.84	1.01	20.47	6.07
	Brok_Flw_S	-0.51	24.96	-0.13	-0.05	-1.03	-29.42	6.38
	IBCC_Rol_S	-3.38	25.24	-3.23	-1.46	-1.05	-34.83	6.15
	IBCC_Exp_S	-3.54	24.85	-3.53	-1.60	-1.03	-34.76	6.18
Long-Short	Both_Rol_S	2.99	25.98	3.45	1.06	-1.03	-21.81	7.12
	Both_Exp_S	2.11	25.71	2.46	0.79	-1.03	-25.39	7.04
	Brok_Flw_LS	4.54	13.92	4.65	2.09	0.52	10.88	6.50
	IBCC_Rol_LS	5.07	11.01	5.30	2.69	0.39	10.63	7.23
	IBCC_Exp_LS	6.50	12.66	6.64	3.35	0.47	13.46	7.06
	Both_Rol_LS	7.99	15.43	8.15	3.18	0.56	11.11	6.32
	Both_Exp_LS	7.88	16.00	8.09	3.12	0.59	11.69	6.29

资料来源：The Journal of Financial Data Science, 兴业证券经济与金融研究院整理

图表 6、2007-2012 多头、空头及多头/空头投资组合历年表现 (%)

Side	Year	Brok_Flw	IBCC_Rol	IBCC_Exp	Both_Rol	Both_Exp	Euro Stoxx
Long-Only	2007	4.37	2.37	2.11	6.07	5.83	7.51
	2008	-47.35	-49.06	-49.38	-47.82	-48.56	-51.09
	2009	44.20	44.45	44.80	45.25	45.47	28.84
	2010	20.29	23.92	24.79	25.00	25.04	5.84
	2011	-9.99	-12.79	-13.56	-8.00	-9.33	-11.91
	2012	20.69	19.42	22.65	20.94	23.81	20.63
Short-Only	2007	5.39	-0.32	-0.99	7.07	8.83	7.51
	2008	46.59	41.07	40.97	41.64	41.26	-51.09
	2009	-42.88	-46.24	-43.83	-39.54	-38.96	28.84
	2010	-12.20	-13.29	-12.73	-7.41	-5.05	5.84
	2011	20.76	18.86	18.04	33.24	25.15	-11.91
	2012	-20.70	-20.14	-22.42	-17.26	-18.74	20.63
Long-Short	2007	5.04	-0.36	-0.99	5.87	5.29	7.51
	2008	-24.52	-16.82	-14.78	-24.42	-24.59	-51.09
	2009	22.66	21.94	24.22	30.99	29.98	28.84
	2010	16.58	21.62	22.81	24.83	24.33	5.84
	2011	-4.83	-6.51	-10.37	-4.03	-7.83	-11.91
	2012	12.00	10.18	17.62	14.12	19.51	20.63

资料来源：The Journal of Financial Data Science, 兴业证券经济与金融研究院整理

图表 7、空头投资组合相对于 Brok_Flw_S 基准的表现



资料来源：The Journal of Financial Data Science, 兴业证券经济与金融研究院整理

6、稳健性检验

6.1 稳健性检验-企业流动性的影响

我们的 IBCC 模型可能存在一个潜在的问题，即倾向于支持那些推荐了规模小、知名度低的股票的经纪商的建议，这会导致规模偏差。图表 8 的结果表明 Brok_Flw_L 投资组合持有的股票规模比 IBCC 多 250 亿美元。

图表 8、不同投资组合的规模

Model	Mega Cap >\$25 Billion	Large Cap \$10 Billion to \$25 Billion	Mid Cap \$2 Billion to \$10 Billion	Small Cap \$250 Million to \$2 Billion	Micro Cap <\$250 Million	Missing Data
Brok_Flw_L	23.1	20.8	48.7	7.2	0.0	0.0
IBCC_Rol_L	17.2	19.9	53.0	9.9	0.0	0.0
IBCC_Exp_L	16.3	19.8	53.8	9.9	0.0	0.0
Both_Rol_L	18.3	19.3	53.2	9.2	0.0	0.0
Both_Exp_L	17.5	19.3	54.0	9.2	0.0	0.0
Brok_Flw_S	15.9	20.6	52.5	10.9	0.0	0.1
IBCC_Rol_S	23.5	19.7	47.7	8.8	0.0	0.0
IBCC_Exp_S	22.8	19.6	48.7	8.6	0.1	0.1
Both_Rol_S	14.8	19.6	53.4	11.7	0.0	0.1
Both_Exp_S	14.2	19.9	54.2	11.2	0.1	0.1

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

为了排除这种影响，我们按市值大小将股票分为两组，并在这两组中进行测试。根据图表 9 展示的两组推荐数据数量，可以看出发现两组数量是差不多的。

图表 9、按市值划分后两组股票的推荐数量

	Universe	Number of Recommendations	As a Percentage
Large Half		58,466	56%
Small Half		45,316	44%

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

图表 10 展示了多头和空头组合的回测表现。图表 11 为两个投资组合的市值分布。根据图表 10、11 我们可以得出结论：IBCC 在按市值划分的两组股票中都能起作用，我们得到的结果不是因为规模偏差。

图表 10、股票市场按市值分成两组后的模型结果

Side	Size	Simulation Name	Return Mean	Vol	Alpha	Alpha t-Stat	Beta	Turnover
Long-Only	Large Half	Brok_Flw_L	5.95	23.93	5.96	3.74	1.01	5.79
		IBCC_Rol_L	6.96	24.57	7.22	3.01	1.01	5.75
		IBCC_Exp_L	7.70	24.50	7.98	3.44	1.01	5.71
		Both_Rol_L	7.77	24.75	7.98	3.14	1.01	6.40
		Both_Exp_L	8.11	24.51	8.38	3.43	1.00	6.32
	Small Half	Brok_Flw_L	4.70	25.60	4.76	1.63	1.03	6.00
		IBCC_Rol_L	4.15	25.98	4.54	1.73	1.06	6.08
		IBCC_Exp_L	2.42	26.08	2.76	1.06	1.07	6.03
		Both_Rol_L	5.89	25.61	6.30	2.13	1.03	6.33
		Both_Exp_L	4.66	25.78	4.98	1.72	1.04	6.27
Short-Only	Large Half	Brok_Flw_S	-2.43	24.81	-2.33	-1.08	-1.03	6.55
		IBCC_Rol_S	-5.69	25.70	-5.30	-2.93	-1.08	6.37
		IBCC_Exp_S	-5.22	25.54	-5.03	-2.46	-1.07	6.41
		Both_Rol_S	-2.53	27.90	-2.13	-0.61	-1.11	7.60
		Both_Exp_S	4.35	28.30	4.59	1.22	-1.12	7.58
	Small Half	Brok_Flw_S	1.98	27.23	2.54	0.77	-1.09	6.51
		IBCC_Rol_S	0.08	26.24	0.23	0.07	-1.05	6.54
		IBCC_Exp_S	-2.68	26.14	-2.13	-0.68	-1.05	6.51
		Both_Rol_S	8.68	27.32	9.49	2.22	-1.02	7.25
		Both_Exp_S	3.23	27.41	4.63	1.02	-1.01	7.16

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

图表 11、两组股票的市值分布

Universe	Mktcap Bucket	Sum Position (%)	Number of Stocks	Return (% p.a.)	Risk (% p.a.)
Large half	Mega Cap (>US\$25 billion)	41.5	75.2	4.2	6.7
	Large Cap (US\$10 billion to US\$25 billion)	33.4	70.6	3.1	6.1
	Mid Cap (US\$2 billion to US\$10 billion)	22.2	53.4	-0.3	6.9
	Small Cap (US\$250 million to US\$2 billion)	2.8	7.8	-1.4	1.9
Small half	Micro Cap (<US\$250 million)	0.0	0.1	0.0	0.1
	Mega Cap (>US\$25 billion)	0.0	0.1	0.0	0.0
	Large Cap (US\$10 billion to US\$25 billion)	5.1	9.9	1.7	0.9
	Mid Cap (US\$2 billion to US\$10 billion)	80.2	149.8	6.8	15.2
	Small Cap (US\$250 million to US\$2 billion)	14.5	30.7	-3.8	8.5
	Micro Cap (<US\$250 million)	0.1	0.2	-0.2	0.3

资料来源：The Journal of Financial Data Science, 兴业证券经济与金融研究院整理

6.2 稳健性检查-交易规则的影响

上文我们假设 $c = k = 1$ ，并以此选择交易行为。本节我们探讨不同 c 和 k 的影响。这里测试了对于多头组合的影响。图表 12 展示了维持 $k=1$ 改变 c 的结果，图表 13 展示了维持 $c=1$ 改变 k 的结果。

当 $k=1$ 改变 c 的结果表明：随着 c 的增加，alpha 和 beta 值都有所增强。维持 $c=1$ 改变 k 时，结果变化的并不显著。但正如预期的那样，我们观察到随着 k 的增加换手率会增加。

图表 12、固定 k 选择不同 c 的结果

Model	c	Mean	Vol	Alpha	Alpha t -Stat	Beta	Beta t -Stat	Turnover
Brok_Flw_L	—	5.43	24.18	5.47	2.73	1.01	26.97	5.75
IBCC_Exp_L	1.0	5.30	24.89	5.39	2.27	1.03	23.63	5.68
	1.1	5.13	25.38	5.43	2.06	1.04	20.67	5.81
	1.2	4.77	26.31	5.42	1.82	1.06	17.78	5.96
	1.3	6.46	26.23	6.49	2.04	1.05	16.58	6.06
	1.4	6.37	26.86	6.15	1.78	1.06	14.70	6.17
IBCC_Rol_L	1.5	5.89	27.34	5.97	1.59	1.06	13.13	6.33
	1.0	4.77	24.66	4.91	2.09	1.02	23.36	5.74
	1.1	4.82	25.14	5.12	1.98	1.03	20.62	5.83
	1.2	4.66	25.60	5.33	1.90	1.04	18.41	5.97
	1.3	5.14	26.04	5.20	1.74	1.05	16.95	5.95
Both_Exp_L	1.4	5.21	26.48	5.18	1.63	1.06	16.03	6.18
	1.5	5.55	27.01	5.56	1.61	1.07	14.56	6.38
	1.0	7.13	24.71	7.28	2.84	1.01	20.47	6.07
	1.1	7.10	25.15	7.59	2.67	1.01	18.17	6.23
	1.2	6.85	26.36	7.22	2.18	1.04	15.14	6.33
Both_Rol_L	1.3	8.73	26.67	8.90	2.50	1.04	14.26	6.49
	1.4	8.33	27.63	8.40	2.14	1.06	12.99	6.56
	1.5	8.52	28.15	8.85	2.10	1.07	11.85	6.73
	1.0	6.99	24.51	7.06	2.75	1.00	20.18	6.13
	1.1	6.98	24.86	7.40	2.63	1.00	17.93	6.28
	1.2	7.12	25.44	7.48	2.37	1.01	15.31	6.38
	1.3	7.36	26.14	7.54	2.18	1.03	13.81	6.37
	1.4	6.34	27.09	6.56	1.74	1.05	12.89	6.58
	1.5	6.97	27.76	7.12	1.76	1.06	11.91	6.78

资料来源：The Journal of Financial Data Science, 兴业证券经济与金融研究院整理

图表 13、固定 c 选择不同 k 的结果

Model	k	Mean	Vol	Alpha	Alpha t-Stat	Beta	Beta t-Stat	Turnover
Brok_Flw_L	—	5.43	24.18	5.47	2.73	1.01	26.97	5.75
IBCC_Exp_L	1.0	5.30	24.89	5.39	2.27	1.03	23.63	5.68
	1.1	5.11	24.96	5.22	2.11	1.03	22.25	5.73
	1.2	5.22	25.20	5.40	2.07	1.03	20.90	5.81
	1.3	5.45	25.78	5.44	1.91	1.04	18.22	5.92
	1.4	5.30	26.25	5.22	1.74	1.06	17.36	6.16
	1.5	5.50	26.51	5.47	1.72	1.06	16.30	6.10
IBCC_Rol_L	1.0	4.77	24.66	4.91	2.09	1.02	23.36	5.74
	1.1	4.88	24.70	5.05	2.07	1.02	22.00	5.83
	1.2	4.83	24.84	5.06	1.97	1.02	20.19	5.85
	1.3	5.24	25.21	5.28	1.94	1.03	18.80	5.97
	1.4	5.23	25.57	5.21	1.83	1.04	17.73	6.12
	1.5	5.37	25.73	5.46	1.83	1.04	16.95	6.05
Both_Exp_L	1.0	7.13	24.71	7.28	2.84	1.01	20.47	6.07
	1.1	7.02	24.78	7.19	2.70	1.01	19.38	6.09
	1.2	7.24	24.89	7.48	2.69	1.01	18.64	6.17
	1.3	7.44	25.65	7.60	2.48	1.03	16.61	6.34
	1.4	7.21	26.30	7.27	2.21	1.04	15.56	6.42
	1.5	7.62	26.67	7.72	2.23	1.05	14.85	6.47
Both_Rol_L	1.0	6.99	24.51	7.06	2.75	1.00	20.18	6.13
	1.1	7.22	24.60	7.32	2.78	1.00	19.44	6.20
	1.2	7.07	24.69	7.29	2.62	1.00	18.06	6.25
	1.3	7.21	24.95	7.40	2.50	1.00	16.94	6.34
	1.4	6.35	25.36	6.48	2.06	1.01	15.54	6.37
	1.5	6.86	25.61	7.03	2.10	1.01	14.65	6.45

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

6.3 稳健性检查-状态变量阈值的影响

上文我们在股票收益率状态变量构建中将阈值设为 5%：

$$R = \begin{cases} 0, & \text{if } r_{(s,\Delta T)} \leq -5\% \times RVol \\ 2, & \text{if } r_{(s,\Delta T)} \geq 5\% \times RVol \\ 1, & \text{else} \end{cases}$$

在这里我们探讨 1%和 10%的阈值变化对我们研究结果的影响。从图表 14 的多头和空头投资组合结果中，可以看出：

- 与之前一样，Both 类策略（多头和空头组合）在所有阈值上均优于相关的 Brok_Flw 基准。
- 对多头投资组合而言，4%-6%为最佳阈值，尤其在 α 的 t 值上表现特别优异。
- 对于空头投资组合而言，稍小的阈值即 2%-3%结果会更好一些。但我们也发现当阈值为 2%-3%时，t 值都不显著。我们认为这可能是由于较小的阈值会导致较大的抽样误差，若样本中存在更多 Price_Down 股票收益率数据取值可能会缓解这种情况。

图表 14、2007-2012 不同阈值的多头与空头投资组合结果

Model	Truth (%)	Mean	Vol	Alpha	Alpha t-Stat	Beta	Beta t-Stat	Turnover
Brok_Flw_L	-	5.43	24.18	5.47	2.73	1.01	26.97	5.75
Both_Exp_L	1	6.07	24.70	6.04	2.55	1.02	21.53	6.00
	2	6.09	24.72	6.24	2.59	1.02	21.34	6.03
	3	6.39	24.65	6.56	2.68	1.01	21.24	6.00
	4	6.57	24.63	6.70	2.68	1.01	20.93	6.03
	5	7.13	24.71	7.28	2.84	1.01	20.47	6.07
	6	7.61	24.86	7.81	2.90	1.01	19.82	6.25
	8	5.94	26.31	6.24	1.86	1.04	15.25	6.56
	10	6.66	26.95	6.74	1.84	1.05	14.93	6.77
	1	6.29	24.39	6.30	2.66	1.00	21.82	6.00
	2	6.10	24.39	6.26	2.62	1.00	21.41	6.03
Both_Rol_L	3	6.10	24.42	6.26	2.59	1.00	21.34	6.00
	4	6.32	24.43	6.46	2.61	1.00	20.69	6.05
	5	6.99	24.51	7.06	2.75	1.00	20.18	6.13
	6	6.97	24.60	7.29	2.75	1.00	19.98	6.29
	8	7.17	26.19	7.95	2.33	1.03	14.78	6.67
	10	7.74	27.00	7.59	2.13	1.06	15.09	6.87
Brok_Flw_S	-	-0.51	24.96	-0.13	-0.05	-1.03	-29.42	6.38
Both_Exp_S	1	2.10	25.76	2.31	0.83	-1.05	-26.19	6.91
	2	3.36	25.56	3.55	1.27	-1.04	-26.51	6.94
	3	3.13	25.45	3.38	1.23	-1.04	-26.27	6.97
	4	2.17	25.53	2.35	0.81	-1.03	-25.31	7.00
	5	2.11	25.71	2.46	0.79	-1.03	-25.39	7.04
	6	4.09	26.51	4.02	1.05	-1.02	-19.30	7.20
	8	0.67	28.84	1.06	0.21	-1.04	-15.77	7.50
	10	-1.15	31.44	-1.28	-0.22	-1.09	-14.66	7.76
	1	2.73	25.68	2.92	1.09	-1.05	-25.52	6.79
	2	2.75	25.43	2.93	1.08	-1.04	-24.57	6.78
Both_Rol_S	3	3.28	25.51	3.49	1.27	-1.04	-24.65	6.84
	4	1.80	25.69	1.98	0.69	-1.04	-24.71	6.99
	5	2.99	25.98	3.45	1.06	-1.03	-21.81	7.12
	6	2.71	26.76	3.07	0.88	-1.05	-23.33	7.23
	8	0.36	27.63	-0.24	-0.06	-1.05	-20.90	7.50
	10	-1.18	29.56	-1.81	-0.36	-1.08	-20.01	7.62

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

6.4 稳健性检查-持仓长度的影响

上文我们取持仓长度为 60 天，本节我们测试不同持仓长度对多头组合的影响。从图表 15 可以看出

- 持有期越短，表现越好，较短的持有期增加了收益。
- Both 类投资组合再次成为表现最好的投资组合。
- IBCC 的表现比 Brok_Flw_L 基准稍逊。

图表 15、2007- 2012 年多头组合不同持仓长度的表现

Model	Holding Period	Mean	Vol	Alpha	Alpha t-Stat	Beta	Beta t-Stat	Turnover
Brok_Flw_L	10	10.04	23.99	10.01	4.55	0.99	29.08	28.20
	20	7.39	24.16	7.27	3.51	1.01	28.17	14.89
	30	6.82	24.33	6.71	3.23	1.02	26.99	10.35
	45	6.05	24.27	6.01	2.98	1.01	26.48	7.31
	60	5.43	24.18	5.47	2.73	1.01	26.97	5.75
	90	5.46	23.94	5.32	2.75	1.00	29.61	4.31
IBCC_Exp_L	10	7.09	24.01	7.08	2.92	0.98	25.66	28.73
	20	5.93	24.81	5.76	2.29	1.02	21.80	15.19
	30	5.15	24.77	5.18	2.13	1.02	21.93	10.55
	45	5.61	24.79	5.76	2.43	1.02	23.20	7.29
	60	5.30	24.89	5.39	2.27	1.03	23.63	5.68
	90	4.95	24.62	5.10	2.23	1.02	26.34	4.20
IBCC_Rol_L	10	8.02	24.01	7.84	3.15	0.98	22.83	28.76
	20	5.75	24.76	5.64	2.22	1.01	21.09	15.24
	30	6.13	24.55	6.20	2.56	1.01	22.49	10.53
	45	5.60	24.46	5.79	2.52	1.01	24.75	7.31
	60	4.77	24.66	4.91	2.09	1.02	23.36	5.74
	90	5.12	24.58	5.28	2.31	1.02	25.71	4.21
Both_Exp_L	10	14.05	23.97	14.21	5.51	0.97	25.38	30.18
	20	9.76	24.75	9.80	3.76	1.01	20.78	15.99
	30	8.42	24.79	8.24	3.18	1.01	20.87	11.03
	45	7.98	24.91	7.93	3.04	1.02	20.73	7.74
	60	7.13	24.71	7.28	2.84	1.01	20.47	6.07
	90	6.63	24.39	6.65	2.66	1.00	22.05	4.71
Both_Rol_L	10	13.38	23.87	13.36	5.00	0.96	22.10	30.28
	20	9.44	24.35	9.62	3.63	0.99	20.48	16.06
	30	9.27	24.66	9.11	3.41	1.00	21.07	11.03
	45	7.71	24.51	7.75	3.07	1.00	21.43	7.73
	60	6.99	24.51	7.06	2.75	1.00	20.18	6.13
	90	6.48	24.27	6.61	2.66	0.99	21.83	4.67

资料来源：The Journal of Financial Data Science，兴业证券经济与金融研究院整理

7、结论

本文叙述了 IBCC 机器学习方法在投资经理的投资流程中的一种应用方式，从实证结果来看效果十分优异。我们将变分近似与 IBCC 方法相结合，用于建立分析师推荐数据的模型，这种方法大幅度提升了计算效率。同时我们发现，Both 类策略即当经纪商的推荐和 IBCC 模型结果一致时采取买入/卖出等操作的效果是最好的。

与其他模型（例如，copula 模型）相比，IBCC 模型的一个重要优点是其可扩展性。文中我们将分析师推荐数据按经纪商（即所属公司）进行分组，但实际上 IBCC 可以用于个别分析师级别或任意级别的数据处理中。同时，我们的模型实现了实时预测。

然而，IBCC 依旧存在着许多缺点。例如 IBCC 所用的数据都是分类数据因此不存在大小的概念。例如股票收益率状态变量的取值 0,1,2，并不能进行大小比较。更重要的是，IBCC 没有将推荐数据和股票收益率数据取值对等，这可能是 Both 类策略表现最好的原因。但为了让模型能够解决奇偶校验效应（parity effect），可能会降低模型的通用性。当然如果有足够多的数据可以解决这个问题，但实际上我们知道数据可得性往往不够。这都是未来有待解决的方向。

参考文献

- 【1】Bernardo, J. M., and A. F. M. Smith. Bayesian Theory. Hoboken: Wiley, 1994.
- 【2】Bishop, C. Pattern Recognition and Machine Learning. New York: Springer, 2006.
- 【3】Blei, D. M., A. Kucukelbir, and J. D. McAuliffe. 2018. "Variational Inference: A Review for Statisticians." arXiv:1601.00670v9.
- 【4】Boyd, S., and L. Vandenberghe. Convex Optimization. Cambridge: Cambridge University Press, 2004.
- 【5】Bradshaw, M. T. 2011. "Analysts' Forecasts: What Do We Know after Decades of Work?" SSRN, June 30, <https://ssrn.com/abstract=1880339>.
- 【6】Brown, L. 1993. "Earnings Forecasting Research: Its Implications for Capital Markets Research." International Journal of Forecasting 9: 295–320.
- 【7】——, ed. 2000. I/B/E/S Research Bibliography. 6th ed. New York: I/B/E/S International Inc., 2000.
- 【8】Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society, Series B (Methodological) 39 (1): 1–38.
- 【9】Fox, C. W., and S. J. Roberts. 2011. "A Tutorial on Variational Bayesian Inference." Artificial Intelligence Review 38 (2): 85–95.
- 【10】Ghahramani, Z., and H. C. Kim. 2003. "Bayesian Classifier Combination." Gatsby Computational Neuroscience Unit technical report no. GCNU-T. London, UK.
- 【11】Givoly, D., and J. Lakonishok. 1984. "Properties of Analysts' Forecasts of Earnings: A Review and Analysis of the Research." Journal of Accounting Literature 3: 117–152.
- 【12】Kim, H. C., and Z. Ghahramani. 2012. "Bayesian Classifier Combination." Proceedings of the 15th AISTATS Conference.
- 【13】Lee, P. M. Bayesian Statistics: An Introduction. Chichester, UK: John Wiley, 2012.
- 【14】Levenberg, A., S. Pulman, K. Moilanen, E. Simpson, and S. Roberts. 2014. "Predicting Economic Indicators from Web Text Using Sentiment Composition." International Journal of

Computer and Communication Engineering 3 (2): 109–115.

【15】Levenberg, A., E. Simpson, S. Roberts, and G. Gottlob. “Economic Prediction Using Heterogeneous Data Streams from the World Wide Web.” In Scalable Decision Making: Uncertainty, Imperfection, Deliberation (SCALE), Proceedings of ECML/PKDD Workshop. New York: Springer, 2013.

风险提示：文献中的结果均由相应作者通过历史数据统计、建模和测算完成，在政策、市场环境发生变化时模型存在失效的风险。

分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

投资评级说明

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级(另有说明的除外)。评级标准为报告发布日后的12个月内公司股价(或行业指数)相对同期相关证券市场代表性指数的涨跌幅。其中：A股市场以上证综指或深圳成指为基准，香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于15%
		审慎增持	相对同期相关证券市场代表性指数涨幅在5%~15%之间
		中性	相对同期相关证券市场代表性指数涨幅在-5%~5%之间
		减持	相对同期相关证券市场代表性指数涨幅小于-5%
		无评级	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级
	行业评级	推荐	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		回避	相对表现弱于同期相关证券市场代表性指数

信息披露

本公司在知晓的范围内履行信息披露义务。客户可登录 www.xyzq.com.cn 内幕交易防控栏内查询静默期安排和关联公司持股情况。

使用本研究报告的风险提示及法律声明

兴业证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告仅供兴业证券股份有限公司(以下简称“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。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本报告所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性，也不保证所包含的信息和建议不会发生任何变更。本公司并不对使用本报告所包含的材料产生的任何直接或间接损失或与此相关的其他任何损失承担任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；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保持在最新状态。同时，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。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。分析中所做的回报预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理部门、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。

本报告并非针对或意图发送予或为任何就发送、发布、可得到或使用此报告而使兴业证券股份有限公司及其关联子公司等违反当地的法律或法规或可致使兴业证券股份有限公司受制于相关法律或法规的任何地区、国家或其他管辖区域的公民或居民，包括但不限于美国及美国公民(1934年美国《证券交易所》第15a-6条例定义为本「主要美国机构投资者」除外)。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给他任其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。

特别声明

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

兴业证券研究

上 海	北 京	深 圳
地址：上海浦东新区长柳路36号兴业证券大厦15层	地址：北京西城区锦什坊街35号北楼601-605	地址：深圳市福田区皇岗路5001号深业上城T2座52楼
邮编：200135	邮编：100033	邮编：518035
邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn