



基于 AI+EI 的长期资产定价模型

王德东 聂挺进 罗戈 查晓磊

摘要 以 ChatGPT 为代表的大模型激起一波新的人工智能 (Artificial Intelligence, AI) 热潮, 资产管理行业也在寻求在各个金融场景中利用 AI 技术来赋能业务的可能。然而金融价格数据信噪比较低, 这就对 AI 技术在该领域的应用提出了新的挑战。与此同时, 人类专家在实际工作中可以很好地胜任多种金融分析任务, 但苦于精力有限无法进一步大幅提升覆盖广度和深度。因此如何有机地整合人类专家经验 (Expert Intelligence, EI) 和 AI 技术就变成了一种有前景的工作思路, 这种思路也被称为 AI+EI 的金融实践。本文聚焦权益资产的长期资产定价问题, 提出基于 AI+EI 的中枢定价模型。实证表明该模型可以充分缓解资产定价当中的多样性难题, 并且相比经典估值计算方法在精度上有大幅提升。

关键词 长期资产定价 估值中枢 AI+EI 资管产品创新

一、引言

长期资产定价在价值投资中长期扮演着核心角色 (Koller et al., 2011)。其中, 现金流折现及其衍生模型在学术领域和投资实践 (Mauboussin et al., 2022) 中都取得了巨大成功。根据研究 (Aggarwal, 2006), 决定公司价值的主要要素有两大类: 一类是公司盈利的质量, 另一类是公司盈利的成长性。在实践中, 卖方的一致预期可作为公司盈利的近似估计, 但即便针对相近的盈利预期, 不同公司的定价也可能相差甚远。这种多样性来源于公司在财务质量、商业模式、竞争格局、终局形态等方面的差异, 对应到 DCF 上就体现为不同的参数

设置。

近年来, 使用机器学习方法对金融数据进行建模并生成投资策略成为投资界热门的研究方向 (张自力等, 2023)。然而, 针对短期量价的工作较多, 聚焦在长期资产定价的量化研究仍较为少见。随着 ChatGPT (Brown et al., 2020) 的兴起, 计算机领域发现基于 Transformer (Vaswani et al., 2017) 的大模型可以很好地处理语言多样性的难题。同时, 基于长期资产定价中的多样性并不显著地多于自然语言处理中的多样性的猜想, 自然联想到将 Transformer 的技术应用到长期资产定价当中。因此, 本文探索利用 Transformer 模型可以掌握多样性知识的特性, 基于 DCF 中

王德东, 上海财经大学信息管理与工程学院; 聂挺进, 华泰证券 (上海) 资产管理有限公司总经理; 罗戈, 华泰证券 (上海) 资产管理有限公司研究中心; 查晓磊, 华泰证券 (上海) 资产管理有限公司权益公募投资部。



的核心要素,即成长、质量因子及宏观因子等,在行业指数上进行定价建模。

在实际中,公司市值在数月内可能变化剧烈,但其内在价值变化缓慢。因此我们通常会用一段时间内的平均市值作为公司合理价值的标签,其也被称作价值中枢。我们会以价值中枢为中心线框定一个价值带。该价值带的上下沿和中心线分别称为上轨、下轨和中轨。一个好的价值带满足两点:①表现市值始终围绕着中枢在上下轨之间波动;②当市值每次显著处于极值时,又刚好和上下轨接触。值得指出的是,即便使用上述标准,从机器学习的角度来讲,该问题标签的质量仍然偏低,这正是该问题的技术难点所在。

相对估值法(Schmidlin, 2014)是一种广泛应用的估值方案。考虑到估值倍数的简明性,在下文中,本文也把公司价值带处理成估值倍数带的形式,简称估值 band,该中枢也称为估值中枢。在实际中,从业者经常使用过去估值的分位数来构建自己的估值 band。该方法虽然能够抓住估值的历史分布情况,但其固定时间窗口、强滞后性等特点也不可避免地陷入在部分时期无法为投资者提供有效参考的困境。在此背景下,如何结合专家先验知识和机器学习方法共同给出公司价值的合理区间,成为实际中迫切的研究课题。本文针对上述问题主要做出了以下贡献。

本文提出经验分布算法,对人类投资者处理历史数据分布的过程进行了模拟,

这一方法相比固定历史时间窗口能更有效地反映历史数据分布情况;同时,本文利用机器学习方法对未来估值数据分布进行预测,结合过去的分布,生成了具备前瞻性、强健性的估值 band,能为投资者提供有价值的赔率参考。

二、相关工作

各大数据提供商常常采用特定的统计算法生成 3~5 条对应的分位数曲线叠加到 PB 等估值时间序列曲线上(以 Wind 为例,附录 1)^①,投资者可以根据估值处于分位数的区间来判断当下估值相较于估值中枢的偏离程度。若当前 PB 向上偏离估值中枢甚至触及上轨,往往意味着该公司或行业相比历史合理估值中枢明显高估,交易赔率明显下降,反之亦然。

如附图 1 所示,针对 $\ln(PB)$ 的时序数据,Wind 的滑动窗口 band 算法能够一定程度上刻画估值高估/低估情况,但在部分时间段上表现出滞后性。此外,固定的历史滑动窗口机制也限制了其刻画数据全貌的能力。理想情况下,通过在滑动窗口中纳入未来的估值数据能够解决 band 的滞后性问题,这就需要拥有能够基于基本面数据对未来的估值数据分布进行预测的能力。同时,固定的历史滑动窗口机制也需要改良,从而建立更符合人类投资者习惯的历史数据退出机制,来改善估值 band 刻画数据整体分布情况的能力。

近年来,Transformer 技术及各种

① 附录 1 至 7、附图 1 至 8 见增强出版,中国知网—《金融市场研究》。



变种在时间序列领域也被广泛应用。Transformer 结构天然适合对长的时间序列依赖进行建模，经过比对（附录 2），我们认为其变种之一 Autoformer（Wu et al., 2021）适合用来解决中枢建模的问题。

三、模型与方法

（一）经验分布估值 band 算法

传统的估值 band 往往采用固定的时间窗口对一段时间内的估值分布进行刻画，但在实际中，投资者往往既要参考中短期的估值分布，又要参考行业在相当长一段时间内的估值分布情况。固定的时间窗口虽然在较小时间窗口能够准确反映中短期的估值分布，但缺乏对长期历史数据分布的刻画；如果时间窗口设置过长，则变相降低了短期数据在整个时间窗口的统计权重。为了解决这个问题，我们发现人之所以会对 band 有一个更好的把握，其本质是对过去数据分布有一个整体性的认识。即，首先是全局性地感知过去历史，其次是历史上越久远的数据越不重要，最后是还要抓住关键性拐点数据。因此类比人类的记忆消退过程，我们借助水库抽样算法（Aggarwal, 2006）思想来模拟人对数据分布的全局性认知，创新性提出了一种计算估值 band 的方法，下文简称为“经验分布 -band”。这种经验分布算法的核心思想是在维护一个固定长度列表的同时，建立一套新数据纳入与旧数据退出的机制，利用水库抽样的思想，将等概率抽样替换为按照特定的映射以不同的权重剔除列表内旧数据，如此一来，在不断加入新数据时，就实现了“时间序

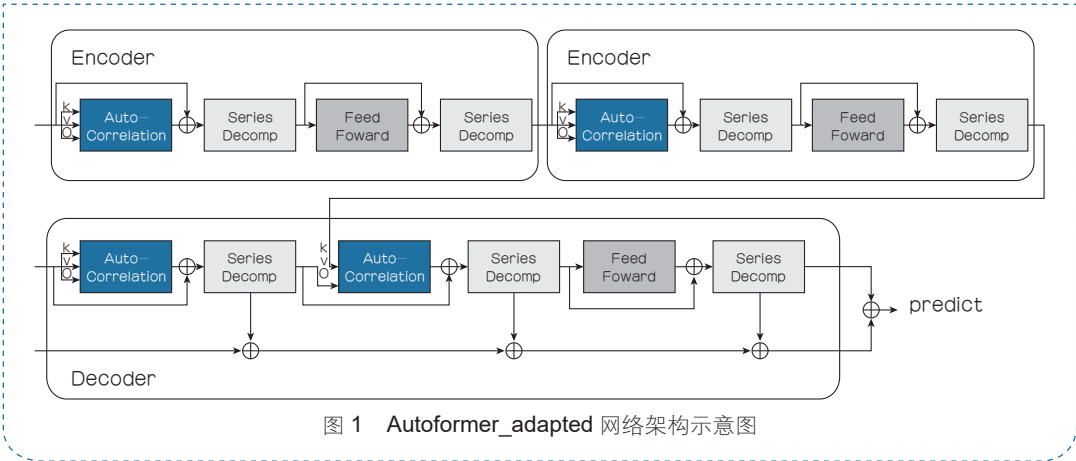
列中的数据离当下时刻越远，对计算分位数起到的潜在影响越小”的目标（附录 3）。

在实际金融市场中，金融数据中有许多一次性的异常情景，比如 2015 年的牛市，如果不加以应对，这些异常情景所产生的数据会导致整个模型结果大幅偏离人类经验。本研究所采用的经验分布 -band 的相比较 Wind-band 可以很好地应对这种异常情景。即，当异常事件发生不久时，由于人们记忆犹新，所以还需要遵守该事件带来的估值锚定效应；但当异常事件已经发生很久了，就需要逐步遗忘这件事，回归长期的金融定价效应上。

（二）基于 Autoformer_adapted 的估值分布预测模型

在长期资产定价问题中，我们选择使用估值 band 来衡量资产定价的效果。同时我们基于财务数据和一致预期构造了一种连续版本的 ROE、NP-growth 等财务指标，其构造思想是先核算出每日等效的盈利，再将过去一年每日的盈利整合在一起，形成一个整体的盈利及相关财务指标（附录 4）。下文中除非特别指出，所有提到的财务数据都是指该版本。

至此，我们基于净资产收益率 ROE、盈利增速 NP-growth、10 年期国债利率等数据，对未来一年的估值分布做出估计。在模型选择上，我们基于 Autoformer 提出一个新模型 Autoformer_adapted，架构如图 1 所示，其在三处做出了重要改进。首先是交换使用序列分解中的周期项和趋势项，使模型在渐进式分解的过程中更关注对趋势项相关性而非周期项相关性的挖掘；其



次是数据预处理上参考 N-Linear (Zeng et al., 2023) 的做法, 用输入数据减去序列的最后一个值, 使模型可以将关注点从数据的相对变化转向变化值大小, 由此减轻数据集不平稳带来的潜在影响; 最后是 loss 函数的选择上使用 MAE-loss, 更加稳健的导数能够避免数据集中部分年份的异常数据波动对训练效果产生过大影响 (附录 5)。

四、实证结果

本节首先对比经验分布 -band 计算方法相比于传统固定窗口 band 计算方法的量化统计优势并进行案例分析, 随后对比纳入预测未来数据后的 band 相比只采用历史数据的 band 的性能提升情况。

具备高参考价值的估值 band 应具备一定的前瞻性, 所以在量化评测的过程中, 本文将通过提前看未来数据 (一年) 的方式构造标签, 分别统计 band 的中轨的偏离情况 (中轨偏离度 = $\text{abs}(\text{中轨_label} - \text{中轨_predict})$, 整体偏离度 = 三根轨道各自偏离度的平均值) 及全部轨道的平均偏离情况。同时, 由

于在 10 年的尺度上 PB 数值可能发生较大变化, 为了更公平地对待不同 PB 数值下的变化幅度, 后续的实验中统一用 $\ln(\text{PB})$ 指代估值, 并基于此计算对应的 band。

(一) 经验分布 PB-band

首先对比两种算法只利用历史数据构造 PB-band 的表现。

1. 案例分析

在附图 2 左图中, 可见 Wind-band (上图) 和经验分布 -band (下图) 在煤炭行业的差距不大, 在只采用历史数据计算的情况下, 二者都能对 PB 时间序列曲线进行一定程度的包络, 并且中轨的估值回归也较为有效。但在较多的行业上, Wind-band 会出现某些异常情况, 如附图 2 右图中, Wind-band 在基础化工行业的 2014 年和 2018 年阶段, 由于数据统计窗口固定, 在此时间区间内的 band 会缩小到很窄的区间。相比之下, 经验分布使用了概率抽样的旧数据退出机制, 使 band 的计算能够一定程度上以较低权重纳入较为久远的历史数据, 构造的 PB-band 更加强健, 利用分



位数的思想生成 band 也避免了估值快速下移阶段 Wind-band 出现喇叭状扩张的问题。

2. 量化评测

附图 3 中展示了偷看未来数据（一年后）的 PB-band 表现情况，可以发现两种算法的 band 中轨都能够更好地反映估值回归，且上下轨能够更好地对估值顶底进行刻画。显然纳入未来数据的 band 能够在估值的赔率层面为投资决策提供非常有效的参考，因此本文以此为标签，统计只利用历史数据计算得到的 band 相比于标签 band 的偏离程度，并以此作为 baseline 进行后续优化，附图 4 给出了 31 个申万一级行业的历史数据 band 偏离程度统计结果。

附图 4 分别展示了 31 个行业（从左到右）Wind-band 中轨、Wind-band 整体、经验分布 -band 中轨、经验分布 -band 整体只采用历史数据的版本相较于标签版本的 PB 偏离程度。可以明显看出，除了石油石化行业之外，几乎全部申万一级行业采用经验分布 -band 的版本偏离幅度均小于 Wind-band，这意味着经验分布算法即便采用历史数据，相对于标签版本的衰退程度依然比 Wind 算法的衰退程度小，具备天然平稳性。

更为重要的是，原有的 Wind-band 上下轨计算结果强健性较差，上下轨偏离情况比中轨要更差，附图 4 中红色柱子高于蓝色柱子也印证了这一点，采用经验分布 -band 算法后，上下轨偏离情况得到了极大的改善，紫色柱子几乎在所有行业矮于绿色柱子，31 个行业 band 平均偏离幅度也从 0.17 下降至 0.12，偏离度下降了 29.4%，band 上下轨的强健性明显提升。

（二）加入模型预测未来数据的 PB-band

在研究中，我们将 2007 年 1 月—2018 年 6 月的数据作为训练集，2018 年 6 月—2023 年 7 月的数据作为测试集。作为对比，用传统的 Transformer 模型、Informer 模型、未经修改的 Autoformer 模型（附录 6）进行了对比。

从预测结果可以发现，基础的 Transformer 和 Informer 虽然在能源等周期性较强的数据集表现较好，但无法对申万一级行业 PB 这类趋势较强而无明显周期波动性的数据进行有效预测，会出现大幅偏离正常值的现象，经验分布算法在纳入模型预测数据后反而性能大幅下降。未经修改的 Autoformer 虽然效果好于前者，但依旧存在预测序列与输入序列之间存在跳变的问题，因此对原始架构进行修改，修改后的模型在预测性能上有较大提升，测试集上的统计偏离结果如表 1 所示。

由表 1 可知，Transformer 等一系列模型对行业估值领域数据集的预测能力较差，在纳入这些模型预测的未来数据后计算得到的经验分布 -band 偏离幅度甚至要大于只采用历史数据的版本。相比之下，Autoformer_adpated 模型能够更好地对行业估值数据进行趋势预测，在 31 个申万一级行业上相比只用历史数据的版本经验分布 -band 中轨偏离幅度从 0.122 减小至 0.073，即，中轨偏离度下降了 40.1%；整体偏离幅度从 0.093 减小至 0.066，即整体偏离度下降了 29.0%，提升效果明显。

纳入模型预测数据后，31 个申万一级行业的具体 PB-band 偏离提升效果如附图 5、



表 1 纳入多种模型数据后申万一级行业经验分布 -band 偏离统计

模型	band 中轨偏离	band 整体偏离
只用历史数据	0.1215	0.0929
Transformer	0.2012	0.3590
Informer	0.1474	0.1580
Autoformer	0.1290	0.1260
Autoformer_adapted	0.0729	0.0659

附图 6 所示，可见加入模型预测未来数据后的新经验分布 -band 在 100% 的行业上优于只采用历史数据计算的的经验分布 -band。

为了进一步说明我们提出方案的特性，我们拿两个典型行业作案例分析。

计算机行业是一个估值变动较大的行业。附图 7 对比了申万计算机纳入预测数据、只用历史数据、偷看未来数据三种模式计算得到的经验分布 -band 情况。可见，纳入模型预测未来数据后，在测试集部分（2018 年 6 月后）上轨的回归速度相较于只用历史数据的版本更具前瞻性且较为平滑，下轨也很少出现被实际估值曲线击穿的现象。这说明针对高波动的行业，我们的估值模型也能给出稳健的估值分布刻画。

相比之下，食品饮料行业估值变动的趋势性相对更加稳定。该行业的估值从 2017 年开始发生系统性抬升，2021 年年年初随着资金流动性达到阶段性顶峰，板块估值也随之见顶。在附图 8 中，一方面在 2019 年的底部，可以发现加入预测数据的 band 下轨对表观估值有更紧密的支撑，即对抬升之后的估值分布有更好的刻画；另一方面对于 2021 年的中轨，只采用历史数据的版本可以明显看出其中轨的趋势逆转点在 2021 年年末，相较于真实的 PB 顶有时长约一年的滞后，而加入了模型预测数据的 band 中轨在上下行期间都明显更具领

先性。这两个案例说明，即便对发生系统性漂移的估值中枢，我们的中枢模型也有较好的刻画力。

综上，本文提出的经验分布 -band+ Autoformer_adapted 的中枢模型相比较 Wind-band 有显著的优势。一方面，通过分位数和水库抽样算法思想，初步解决了 Wind-band 上下轨不强健的缺点；另一方面，通过纳入更高质量的预测数据，最终方案的中枢平均偏离度进一步从 0.12 下降至 0.07，整体偏离度从 0.09 下降至 0.06，缓解了只采用历史数据 band 版本的滞后性问题，并且该模型相比较其他的中枢模型也取得了明显的比较优势（表 1）。同时，基于经典 Transformer 或 Autoformer 方案的估计偏离度并未显著小于不做预测的偏离度，这首先意味着经验分布 -band 本身可以很好地刻画估值中枢，其次也印证了 Autoformer_adapted 中的修改对估值中枢建模是重要的。

五、结论

针对权益资产的长期资产定价问题，本文创新性地提出了一种基于经验分布 -band 和 Autoformer_adapted 的估值定价模型。在申万一级行业指数范围内测试发现，本文提出的 AI+EI 的模型显著地好于基于分位数的估值定价方案，也好于基于经典预测模型的估值定价方案。与此同



时,该模型不仅在平均偏离度上大幅优于基线模型,还在每个一级行业上都优于基线模型。因此该模型在很大程度上缓解了估值定价当中的核心难题之一,即多样性难题。该模型提供了一致性的中枢定价,既

可以作为单行业赔率的信号(附录7),也可以作为跨行业比较赔率的依据。同时该模型也具备直接拓展到所有分析师覆盖个股的可能,这也是下一步的探索方向。[N](#)

学术编辑:韦燕春

参考文献

- [1] 肖欣荣,李怡雯.人工智能如何挑战投资管理?——基于投资决策框架的讨论[J].金融市场研究,2023(03):136-140.
- [2] 张自力,赵学军,陈莹.资本市场复杂网络特性与资产定价新视角[J].金融市场研究,2023(02):54-61.
- [3] Aggarwal C. On biased reservoir sampling in the presence of stream evolution[C]. Proceedings of the 32nd International Conference on Very Large Data Bases, 2006:607-618.
- [4] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33:1877-1901.
- [5] Koller T, Dobbs R, et al. Value: the four cornerstones of corporate finance[M]. New Jersey: John Wiley & Sons, 2011.
- [6] Lim B, Arık S Ö, Loeff N, et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting[J]. International Journal of Forecasting, 2021, 37(4):1748-1764.
- [7] Mauboussin M, Callahan D. The Math of Value and Growth[R]. Morgan Stanley Investment Management, 2022.
- [8] Schmidlin N. The Art of Company Valuation and Financial Statement Analysis[M]. New Jersey: John Wiley & Sons, 2014.
- [9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30:6000-6010.
- [10] Wu H, Xu J, Wang J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[J]. Advances in Neural Information Processing Systems, 2021, 34:22419-22430.
- [11] Zeng A, Chen M, Zhang L, et al. Are transformers effective for time series forecasting[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(9):11121-11128.
- [12] Zhou H, Zhang S, Peng J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12):11106-11115.
- [13] Zhou T, Ma Z, Wen Q, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]. International Conference on Machine Learning, 2022, 162:27268-27286.

Developing a Long-term Asset Pricing Model Based on AI+EI

WANG Dedong¹ NIE Tingjin² LUO Ge³ ZHA Xiaolei⁴

(1. School of Information Management & Engineering, Shanghai University of Finance and Economics; 2. Huatai Securities (Shanghai) Asset Management CO., LTD; 3. Research Center of Huatai Securities (Shanghai) Asset Management CO., LTD; 4. Equity Investment Dept of Huatai Securities (Shanghai) Asset Management CO., LTD)

Abstract With large language models such as ChatGPT sparking a new wave of artificial intelligence (AI), the asset management industry has been actively exploring the possibility of adopting AI technologies in financial scenarios. Financial price data have a notorious low signal-to-noise ratio, which creates new challenges to utilizing AI technology in this sector. Financial experts have accumulated considerable expertise in this area and the integration of expert intelligence (EI) and AI technologies has become a promising direction for future research. This combination has become known as AI+EI financial practice. In this paper, we focus on problems with long-term asset pricing and propose a pioneering asset pricing model by adopting an AI+EI approach. The empirical results show that our model can effectively alleviate the huge diversity problem in asset pricing, and greatly improve the estimation accuracy compared with the empirical method and mainstream baselines.

Keywords Long-term Asset Pricing, Valuation Center, AI+EI, Asset Management Product Innovation

JEL Classification C45 C53 G12