宏观“失聪”与羊群“先觉”？国债风险溢价影响因子的预测有效性

刘方根　纪伟伟　白明浩

摘　要　本文采用2010—2023年中国银行间市场国债收益率数据、宏观金融市场数据及公募基金披露数据，采取单因子回归模型和多因子机器学习模型探究了长期国债风险溢价的影响因子及其有效性测度。研究发现，收益率曲线因子是测度国债风险溢价的基准性因子，因其有效提取了与债券风险溢价密切相关的利率变动、波动率、曲线曲率信息。宏观因子对债券风险溢价的预测能力具有时期依赖性，伴随中国债券市场的深化发展，金融市场参与主体的多元化，投资者的学习效应使债券市场对宏观经济、金融市场信息实时反应在债券定价中，2015年以后宏观因子无法提供风险溢价预测的额外信息。特别地，本文发现除传统的预期风险溢价影响因子外，机构投资者行为是影响债券风险溢价的重要因素。机构羊群效应和正反馈策略加剧了市场单边波动幅度，投资者的偏见或信念对债券风险溢价具有实质性预测能力。本文研究对于国债风险溢价水平的定量测度、观察长期债券市场定价的合理性、监测金融市场利率风险累积情况具有参考意义。

关键词　国债风险溢价　远期利率因子　机器学习模型　机构羊群行为

一、引言

党的二十大报告明确提出“加强和完善现代金融监管，强化金融稳定保障体系，坚决守住不发生系统性风险的底线”。2024年以来，人民银行多次提示债券市场风险，其核心指向市场存在的期限错配与利率风险隐患。国际金融市场近年风险事件提供了深刻镜鉴。2022年美联储开启激进加息周期，显著抬升全球融资成本，加剧市场波动，长期债券利率风险集中暴露。典型案例如美国硅谷银行在低利率时期过度配置长久期国债及MBS，利率骤升导致巨额浮亏并引发挤兑倒闭；英国养老基金及日本农林银行亦因长期国债利率急速上行而蒙受重大损失，甚至诱发局部流动性危机。这些事件印证了金融风险极强的传染性与破坏力，亦推动全球监管框架持续强化宏观审慎管理。

当前，我国长期国债利率处于历史低位，收益率曲线趋平，市场主体对长久期资产的追逐易滋生“羊群效应”，加剧收益率单边下行中的潜在系统性风险。人民银行从宏观审慎视角密切监测市场动态，及时进行风险提示与预期引导，有效阻断金融市场风险累积。深入分析长期债

刘方根，中国建设银行财务会计部高级资深专家、总经理；纪伟伟，中国建设银行金融市场部高级经理；白明浩，中国建设银行金融市场部业务经理。

券风险溢价的定量测度方法及其核心驱动因子，对科学评估市场风险敞口、前瞻性防范系统性利率风险具有重大理论与现实意义。

本文基于2010—2023年中国银行间市场国债收益率曲线、宏观金融数据及公募基金披露数据，通过构建单因子回归模型与多因子机器学习模型，系统探究长期国债风险溢价的影响机制及预测效能。研究发现主要结论如下。一是因子预测效力的动态演进特征。本文发现2015年以前，市场有效性较弱，即使考虑了小样本偏差，宏观变量及机器学习也有助于提升长债风险溢价预测效果，同时收益率曲线信息的预测能力不足。然而，2015年后宏观因子的预测效果显著下降，预测能力无法超越远期利率因子基准模型，投资者获取超额收益的难度大大增加。这种衰减与投资者学习效应有关（Calluzzo et al.，2019），机构投资者会基于有预测能力的宏观变量开展交易，使这些指标因提前定价而失效。二是行为因子的增量解释能力。本文发现机构投资者存在羊群行为，会造成债券市场价格过度反应。据此构建交易拥挤度因子，与宏观因子相比，对当前长期国债风险溢价具有额外的稳定预测能力。路磊（2014）发现股票机构投资者的羊群行为主要来自排名考核等现实压力。现实压力导致机构的羊群行为相对收益率曲线对长债风险溢价具有额外的预测能力。

本文研究的边际贡献体现在，一是揭示市场演进中的因子更替规律，首次实证验证债券风险溢价从“宏观驱动”向“微观结构驱动”的时点（2015），发现投资者学习效应是宏观因子失效的核心机制。二是构建行为金融量化框架，创新开发基于持仓数据的交易拥挤度指标，破解传统模型对市场非理性波动的解释不足。研究表明，当前长债风险溢价预测应构建“利率因子为基、行为因子为补”的双轮驱动模型，这为监测债券市场利率风险提供了重要的量化依据。

二、文献综述与理论分析

（一）文献综述

债券风险溢价作为无风险长期债券相对于短期债券的预期超额收益，其理论内核与预测机制始终是资产定价领域的核心议题。早期纯粹预期假说（PEH）主张远期利率是未来即期利率的无偏估计（Lutz, 1940），但后续实证推翻了风险中性假设，证实远期利率包含时变风险溢价（Ilmanen，2012）。这种溢价受通胀风险补偿（Wachter，2006）、经济周期关联（Cochrane & Piazzesi，2005）及政策不确定性（Bauer & Rudebusch，2020）三重驱动。在测度方法演进中，学界形成三大主流范式：Fama & Bliss（1987）开创的远期利率预测框架证实远期利率差对3~5年期债券超额收益具有预测力；Ludvigson & Serena（2009）通过提取132个宏观变量的主成分，使样本外预测精度提升20%；Kim & Wright（2005）则利用专业机构调查数据分解预期与溢价成分。

学界对宏观因子的实证效力存在根本分歧。支持者强调经济基本面的增量价值，Ludvigson（2009）发现工业生产等宏观因子能补充刻画投资者的风险厌恶程度。反对者则揭示三重局限，Bauer & Hamilton（2018）经自举法检验指出宏观因子的预测能力多源于小样本偏差；Zhang et al.（2022）证明零利率环境下传统利率因

子预测机制失效；吴小伟和丁剑平（2020）发现中国宏观因子预测效果随货币政策框架转变而弱化。机器学习技术为破解争议提供新路径，Bianchi（2020）采用随机森林模型捕捉非线性关系，使预测精度提升37%。

中国债券市场研究呈现显著的“样本选择悖论”。杨宝臣和张涵（2016）基于2007—2014年数据证实远期利率因子预测显著性，但未进行样本外检验；类承曜和陈礼清（2021）采用2010—2019年样本发现宏观经济景气指数具有预测能力，却忽略利率市场化转型期的结构断点；Zhang et al.（2022）的2002—2020年全样本研究则认为宏观因子预测能力源自小样本偏差。这种矛盾源于市场演进与检验方法双重缺陷：一方面，2015年以后，中国利率传导体系完善，因子预测机制发生结构性变化；另一方面，王安兴和余文龙（2012）等研究依赖样本内回归，未详细验证样本外效力。

行为金融理论近年颠覆传统理性预期框架。Bauer & Chernov（2024）发现，投资者并非完全理性，并且长期债券超额收益可预测性的绝大部分来自投资者的非理性情绪而非预期风险溢价。Ben-Rephael et al.（2012）研究发现，投资者一致将债券基金切换为股票基金的行为短期伴随正的股市超额回报，但4~10个月后超额收益转负，资金流动存在“短期正反馈 - 长期反转”特征。这无法由时变的风险溢价解释，而由投资者非理性情绪解释。许年行(2013)、朱菲菲（2019）等研究发现，国内股票市场机构投资者存在忽略个人私有信息而跟从他人投资的羊群行为，导致机构投资者一致买入的股票长期超额收益显著为负。对国内债券市场机构行为的研究较少，类承曜（2024）通过基金季报提取基金久期、杠杆率、信用下沉得分指标，发现策略相似性更高的基金更容易同时出现大规模资产抛售，却未建立行为因子与风险溢价的量化关联。

当前文献主要局限为多数模型未考虑机构行为因子，忽视跨市场传染效应（如2022年理财赎回危机）。未来研究应着力构建基于高频持仓数据的交易拥挤度指标。结合中国金融转型特征与行为金融新范式，构建“宏观 - 微观 - 行为”三维预测框架，将是破解债券风险溢价预测难题的关键突破路径。

（二）理论分析与研究假设

理论上，长期债券风险溢价的可预测性来自两部分，即投资者的非理性及预期风险溢价。早期的研究假设投资者是理性的，金融市场是有效的。金融市场是有效的意味着宏观金融信息能够及时反映在收益率曲线定价中，宏观因子无法额外提升解释能力。投资者是理性的意味着投资者行为无须研究，长期债券风险溢价的可预测性只能来自预期风险溢价。

金融市场有效性假设能够解释现有国内研究关于宏观因子及机器学习在提升债券风险溢价预测能力产生分歧的原因。一方面，采用年度同比数据回归存在重叠样本，容易产生显著结果，这种可预测性可能来自小样本偏差（Bauer & Hamilton，2018）。另一方面，市场的有效定价需要通过投资者交易达成（Calluzzo et al.，2019），这种可预测性也可能来自部分时期市场并非完全有效，收益率曲线无法完全反映全部宏观金融信息。基于此，本文提出假设1。

假设 1：早期国内债券市场曲线定价不充分，宏观因子及机器学习技术能够提升国债风险溢价预测能力。随着国内债券市场的发展，当前收益率曲线已基本反映全部宏观金融信息，宏观因子及机器学习技术的预测效果主要来自小样本偏差。

对于投资者理性假设，本文尝试采用债券公募基金久期中位数刻画机构投资者行为。若债券公募基金久期中位数显著高于历史均值，代表多数机构投资者采用拉长久期的投资策略，即可观察该行为是否造成债券市场过度定价，区分该行为是出于价格发现的理性行为，还是包含声誉风险、从众心理等因素引发的羊群行为。基于此，本文提出假设 2。

假设 2：国内债券机构投资者存在羊群行为，当公募基金普遍拉长债券型基金久期至相对高位时，后续短期超额回报不明显，并在长期伴随显著负向的超额回报。出于业绩排名压力等因素，机构难以为自身的羊群行为定价，导致机构行为因子对国债风险溢价具有额外的预测能力。

三、研究设计与数据

（一）因变量

参考郑振龙 等（2009），本文采用长债超额收益衡量国债风险溢价。具体地，国债风险溢价指 2~10 年期国债持有 1 年收益相对持有 1 年期国债的超额收益。在下式中，$rx^{n}_{t\to t+1}$ 代表 N 年期国债持有 1 年相对 1 年期国债的超额收益，$\overline{rx}_{t\to t+1}$ 代表 2~10 年期国债持有期超额收益的均值。y^{n}_{t} 代表 N 年期即期利率，y^{n-1}_{t+1} 代表 1 年后的 N−1 年期即期利率。

$$\begin{aligned} rx^{n}_{t\to t+1} &= \ln(p^{n-1}_{t+1}) - \ln(p^{n}_{t}) - y^{1}_{t} \\ &= \ln(e^{-(n-1)y^{n-1}_{t+1}}) - \ln(e^{-ny^{n}_{t}}) - y^{1}_{t} \\ &= ny^{n}_{t} - (n-1)y^{n-1}_{t+1} - y^{1}_{t} \qquad n=2,3,\cdots,10 \end{aligned}$$

$$\overline{rx}_{t\to t+1} = \frac{\sum_{n=2}^{10} rx^{n}_{t\to t+1}}{9}$$

本文进一步选取中债长期债券财富指数 ① 相对于中债短期债券财富指数的超额收益作为债券风险溢价的辅助观测指标。中债指数由中央国债登记结算有限责任公司负责编制和发布，被市场广泛用于业绩基准、指数型基金跟踪等。中债长期财富指数和短期财富指数分别记录了待偿期在 10 年以上、10 年以内的利率债券在考虑利息收入、资本利得及利息再投资后的整体价格变动。通过与基期价格比较，中债长期债券财富指数收益率与中债短期债券财富指数收益率的差可以衡量长期债券的超额收益（类承曜和陈礼清，2021）。图 1 记录了两种指标计算得到的国债风险溢价，可以看出两种指标高度相关。2010—2020 年国债风险溢价总体围绕 0 值波动，2020 年以来国债风险溢价均值为正，长期债券回报持续跑赢短期债券，且国债风险溢价波动性有所下降。

（二）自变量

基于 Zhang et al.（2022）等研究的宏观因子，并结合债券市场投资者密切关注的宏观经济周期及金融市场指标，本文构

① 中债财富指数由中央国债登记结算有限责任公司发布，长期、短期财富指数分别记录了待偿期在10年以上、10年以内的记账式国债和政策性银行债包含票息、资本利得在内的全回报。因此，中债长期债券财富指数收益率与中债短期债券财富指数收益率的差可以衡量长期债券的超额收益。

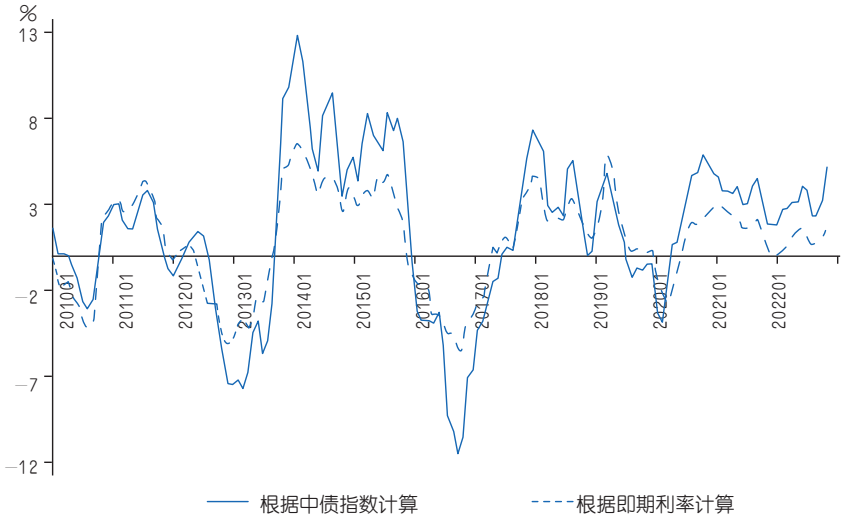

图 1　国债风险溢价历史走势

建六大类共计 26 个自变量。六大类别包括收益率曲线类、市场利差类、资产比价类、机构行为类、宏观数据类和宏观利差类。具体来看，经济增长、通货膨胀和长期债券定价关联紧密，社会融资规模与货币政策密切相关（盛松成，2012）。除宏观总量指标外，国内已有研究从债券波动率、信用利差、隐含税率等角度构建债券市场微观指标（王中和郭栋，2018；杨为敦 等，2020），以对长期债券的投资价值进行评估。本文参考了已有研究成果，并进一步提出公募基金债券久期指标，作为观测机构行为拥挤度的代理指标。

各变量定义如表 1 所示。①

（三）研究设计

本文根据式（1）至式（3）构建预测模型：

$$\overline{rx}_{t\to t+1}=\beta_0+\beta_1 CP_t+\varepsilon_t \quad (1)$$

$$\overline{rx}_{t\to t+1}=\beta_0+\beta_1 CP_t+\beta_i x_{it}+\varepsilon_t \quad (2)$$

$$\overline{rx}_{t\to t+1}=g(CP_t;X_t) \quad (3)$$

其中，$\overline{rx}_{t\to t+1}$ 代表国债风险溢价，CP_t 代表远期利率因子，X_t 代表其他因子。

在（1）式中，本文基于线性回归检验远期利率因子对国债风险溢价的预测效果。在（2）式中，本文在远期利率因子基础上，依次加入其他单因子，检验其对国债风险溢价预测的提升效果。在（3）式中，本文通过机器学习模型对远期利率因子及其他因子进行整合，检验考虑多因子后国债风险溢价预测的提升效果。本文使用的机器学习模型包括主成分分析法、惩罚回归、支持向量机、随机森林、神经网络。

评估模型预测效果的指标包括样本内拟合优度（R2）及样本外预测均方误差（MSE）。样本外预测均方误差（MSE）的计算方法如下：以 2013 年 12 月为例，采用 2010 年 1 月至 2012 年 12 月共计 36 个月数据，使用自变量对国债风险溢价回归，得到回归估计系数，代入 2013 年 12 月远自变量真实值得到 2013 年 12 月国债风险

① 表1~3，见增强出版，中国知网—《金融市场研究》。

溢价预测值。之后采用扩张窗口进行估计，每次在回归中增加 1 个月数据。例如，在 2022 年 12 月，采用 2010 年 1 月至 2021 年 12 月共计 144 个月数据，使用自变量对国债风险溢价回归，得到回归估计系数，代入 2022 年 12 月自变量真实值得到 2022 年 12 月国债风险溢价预测值。将 2013 年 12 月至 2022 年 12 月每月的国债风险溢价预测值 $\hat{y}$ 记录，与真实值 y 进行对比，$MSE=\sum(\hat{y}-y)^2$。

本文的样本外预测区间为 2013 年 12 月至 2022 年 12 月。在此基础上，本文进一步将样本切分为 2013 年 12 月至 2018 年 12 月、2019 年 1 月至 2022 年 12 月子样本，单独观察各自变量在近年来的样本外预测效果。此外，本文详细展示了各预测模型的样本外预测值与真实值时间序列走势，以对各阶段自变量预测效果进行更详细的评估。

（四）数据描述

本文研究使用的数据均来自 Wind，并采用银行间债券市场数据。商业银行根据监管要求自 1997 年全面退出交易所市场，银行间债券市场同步成立。21 世纪以来债券市场进入快速发展轨道。2000 年初，中国债券市场规模仅 2.2 万亿元，债券数量 183 只，至 2024 年 6 月末已达到 164 万亿元，规模位居全球第二，债券数量高达 7.2 万只，年度交易量超过 370 万亿元，其中银行间债券市场托管量和交易量达到全市场的 90%。考虑多数数据可得性，如债券远期利率及 FR007 互换利率等，样本区间为 2010 年 1 月至 2023 年 12 月。对于非平稳时间序列，采用对数差分或差分增强其平稳性，ADF 检验结果表明变量在 5% 左右显著性水平能够拒绝存在单位根的原假设。变量描述性统计如表 2 所示。

四、单因子对债券风险溢价的预测效果

（一）远期利率因子计算方法

本文参考 Cochrane（2005）远期利率因子构建方法，并将期限进一步扩展到 10 年期：

$$\overline{rx}_{t\to t+12}=\alpha_0+\alpha_1 y_t^1+\alpha_2 f_t^3+\alpha_3 f_t^5+\alpha_4 f_t^7+\alpha_5 f_t^{10}$$

$$\text{远期利率因子}=\hat{\alpha}_0+\hat{\alpha}_1 y_t^1+\hat{\alpha}_2 f_t^3+\hat{\alpha}_3 f_t^5+\hat{\alpha}_4 f_t^7+\hat{\alpha}_5 f_t^{10}$$

其中，$\overline{rx}_{t\to t+12}$ 代表持有长债一年的超额收益，y_t^1、f_t^3、f_t^5、f_t^7、f_t^{10} 分别代表当月的 1 年期即期利率和 3 年、5 年、7 年、10 年期远期利率。远期利率因子为根据回归系数计算的远期利率线性加权值。由于本文关注国债风险溢价的样本外预测效果，采用 36 个月的滚动窗口估计远期利率因子。

（二）公募基金久期中位数计算方法

本文基于类承曜（2024）关于策略趋同性的研究基础，创新构建公募基金久期中位数指标（Median_Dur）作为机构行为拥挤度的核心代理变量。基金久期计算采用净值回归法，具体为通过基金净值变化对不同期限的中债净价指数变化进行回归，进而采用回归系数对中债净价指数的久期进行加权来估计基金的久期。具体计算公式如下：

$$\text{Fund}^{\%}=\sum_{i=1}^{6}\omega_i\cdot Index_i^{\%}$$

$$s.t.\sum_{i=1}^{6}\omega_i=1,\ \omega_i\geq 0$$

$$\text{Median\_Dur}=\sum_{i=1}^{6}\omega_i\cdot Index_i^{\text{D}}$$

其中，$\text{Fund}^{\%}$ 代表基金的复权单位净价变动幅度，$Index_i^{\%}$ 代表中债净价指数 i

的涨跌幅度，其中中债净价指数包括1年以内中债净价指数、1~3年中债净价指数、3~5年中债净价指数、5~7年中债净价指数、7~10年中债净价指数、10年以上中债净价指数，采用LASSO回归降低模型的多重共线性。ω_i代表回归系数，$Index_i^D$代表中债净价指数久期。采用ω_i对$Index_i^D$加权得到基金久期Median_Dur。

采取该方法的主要考量是：其一，基于基金公开披露数据计算久期配置离散度（ΔDur_MD），克服传统LSV模型依赖非公开交易数据的局限；其二，与季报披露组合久期的相关系数达0.82（p<0.001），证明测算准确性；其三，通过移动平均基准剥离市场正常波动，精准识别异常拥挤状态（阈值 ±1.5σ）。图2刻画了计算得到的公募基金久期中位数序列[①]。由于国内中长期纯债基金在2014年起初具规模，因此数据的起始点为2014年。采用季报披露的组合久期中位数进行稳健性检验，两者走势高度正相关。可以看出，公募基金久期中位数与未来一年的利率变动[②]负相关。例如，2024年债券市场走出大牛市行情，然而2023年末全市场公募基金久期中位数处于历史低位。

（三）单因子预测能力回归分析

表3记录了收益率曲线及其他因子对国债风险溢价的回归结果。具体来看，表3的第（1）栏记录了远期利率因子对国债风险溢价的回归结果。在表3的第（2）~（10）栏中，回归均保留远期利率因子变量，并在每栏加入1个单因子变量，观察该变量的显著性及预测能力。评估变量预测能力的指标包括R2、MSE及MSE_2。R2代表回归的拟合优度，评估模型的样本内预测能力。MSE代表样本外预测的

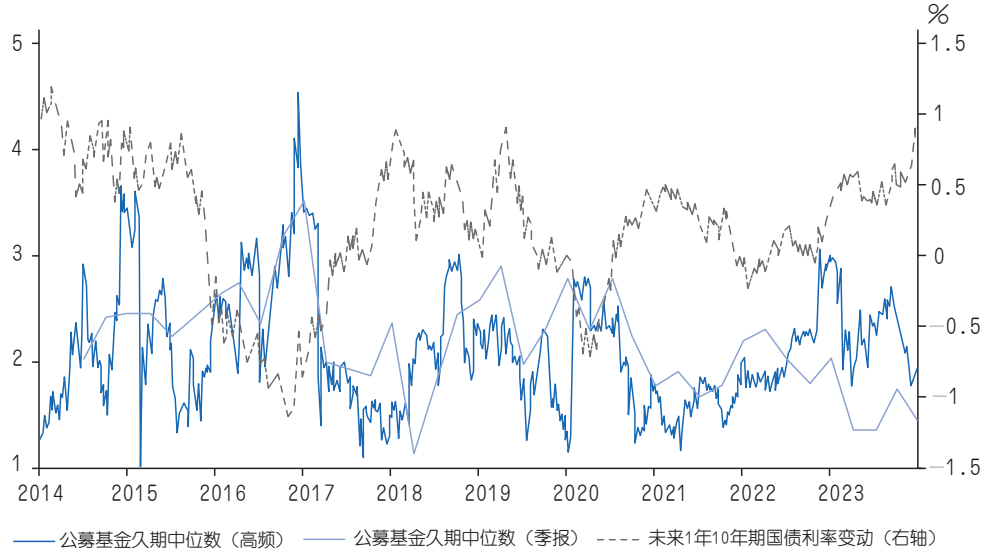

图2　公募基金久期中位数历史走势

① 样本共计60只中长期纯债基金。

② 未来1年利率变动=当前10年期国债利率-1年后10年期国债利率。

均方误差，预测期间为2013—2022年，MSE_2样本外预测期间为2019—2022年，用于评估模型在近年来的预测效果。

表3的回归结果表明，在考虑宏观因子前后，远期利率因子对长债超额回报均具有显著的预测能力。在诸多宏观经济指标中，仅有M1_M2剪刀差、社融同比能够进一步降低债券风险溢价的样本外预测均方误差。相对采用远期利率因子单因子预测，加入M1_M2剪刀差、社融同比分别将样本外预测均方误差由8.41降低至7.78、6.96。

然而，近年来仅有公募基金久期中位数因子能进一步降低样本外预测误差，M1_M2剪刀差、社融同比因子对债券风险溢价的预测能力已消失。相对采用远期利率因子单因子预测，在2019—2022年子样本中，加入公募基金久期中位数因子将样本外预测均方误差由3.63显著降低至1.47，M1_M2剪刀差、社融同比因子反而将样本外预测均方误差由3.63提升至3.70、3.88。这是由于当社融等金融数据领先利率走势成为机构投资者共识后，市场机构开始高度关注相关数据并开发出一系列预测社融的高频模型，导致市场对金融数据提前定价，金融数据的领先性逐步消失。

（四）单因子样本外滚动回归系数检验

为进一步检验不同因子样本外预测效果的稳健性，选定2014—2016年为第一个样本内时期，采用LASSO回归筛选样本期内对未来1年长期债券超额回报预测能力最强的5个宏观因子，计算5个宏观因子对长期债券超额回报回归系数t值绝对值的均值。设定2018年为第一个样本外时期、2019年为第二个样本外时期，计算以上5个宏观因子在样本外时期回归系数t值绝对值的均值。按照同样的方式，选定2017—2019年为第二个样本内时期，2021年、2022年为对应的第一个、第二个样本外时期，记录t值绝对值的均值。

图3 PanelA汇报前5大宏观因子在样本内及样本外的预测表现。结果表明，样本内具有预测能力的宏观因子在样本外时期预测能力迅速失效，且失效速度呈加速趋势。筛选2014—2016年预测能力最强的5个因子，在第一个样本外时期，t值

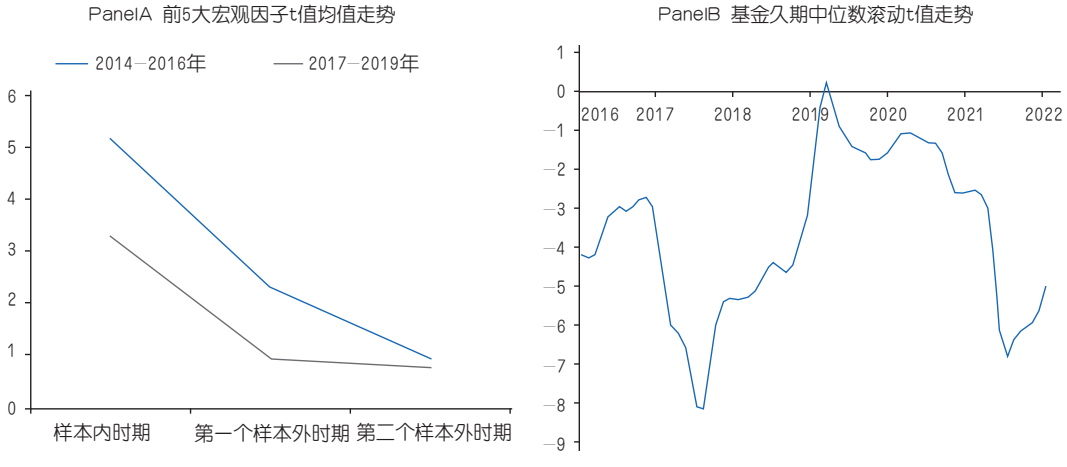

图3 宏观因子与基金久期因子样本外预测能力比较

绝对值均值由样本内时期的 5.17 降低至 2.34，尚保留一定预测能力。在第二个样本外时期，t 值绝对值均值则下降至 0.96，在 10% 显著性水平下不显著，表明 5 个因子在平均意义上已失效。筛选 2017—2019 年预测能力最强的 5 个因子，在第一个样本外时期 t 值绝对值均值已下降至 0.96，无显著的预测能力。以上结果表明投资者会归纳学习过去一段时间有效的预测因子，导致因子因提前定价而失效。并且，伴随投资者研究框架的成熟，因子的失效速度呈加速趋势。

图 3 PanelB 汇报基金久期中位数因子在子样本的预测表现。在 PanelB 中，采用 3 年固定窗宽滚动计算基金久期中位数对未来长期债券超额回报的回归系数 t 值。结果显示，基金久期中位数与未来长期债券超额回报在样本期内均呈现负相关关系，并且，t 值绝对值在 71% 的样本时期高于 5% 显著性水平临界值。基金久期中位数的预测能力并没有因投资者的学习效应而失效，这是由于基金久期中位数本身揭示了机构投资者的羊群行为，机构投资者的羊群行为则因从众心理、声誉等因素的影响难以消除。

五、多因子叠加对债券风险溢价的预测效果

上一节基于单因子分析，结果表明除远期利率因子、公募基金久期中位数因子外，采用其他单一宏观因子难以长期保持对国债风险溢价的稳定预测能力。本节通过机器学习模型进行多因子分析，探讨充分考虑宏观因子的相互关系能否进一步提升样本外预测能力。

（一）基于机器学习的样本内预测

为了对宏观变量进行有效的降维，本节采用传统的主成分分析法（PLS）以及更为复杂的监督式机器学习模型，包括惩罚回归（GLMNET）、支持向量机（SVM）、随机森林（RF）、神经网络（NNET）。为训练模型，本文首先将 2010—2022 年的数据划分为训练集及测试集。具体而言，首个训练集窗口为 2010—2012 年，测试集窗口为 2013 年，并固定训练集及测试集窗宽以此类推。在训练集测试模型，使

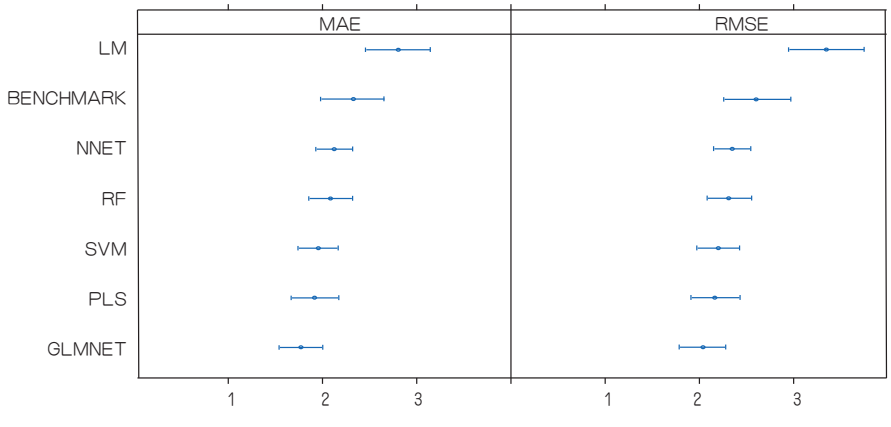

图 4　机器学习模型样本内预测误差均值及 95% 置信区间

模型在所有测试集窗口的平均预测误差最小，记录模型在测试集的平均绝对误差（MAE）及均方根误差（RMSE），结果如图 4 所示。

从图 4 可以看出，相对于包含远期利率因子的基准模型（BENCHMARK），直接将所有因子代入线性回归的简单回归模型（LM）会造成严重的过拟合问题，预测误差明显增大。然而，各类机器学习模型的预测误差均小于基准模型（BENCHMARK），表明采用恰当的预测结构后，考虑全部宏观变量信息有助于提升模型的样本内预测能力。

（二）基于机器学习的样本外预测

本节中将样本分为训练集、测试集、验证集，以检验机器学习模型的样本外预测效果。具体地，选取 2010—2013 年的数据，采用上一节的方法得到样本内预测

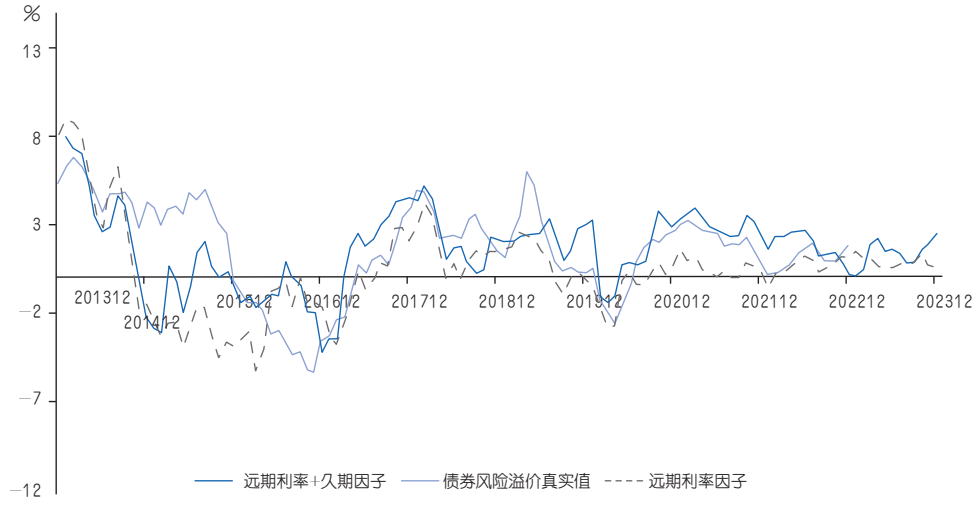

图 5　未来 1 年国债风险溢价预测值及实际值 - 远期利率、久期因子模型

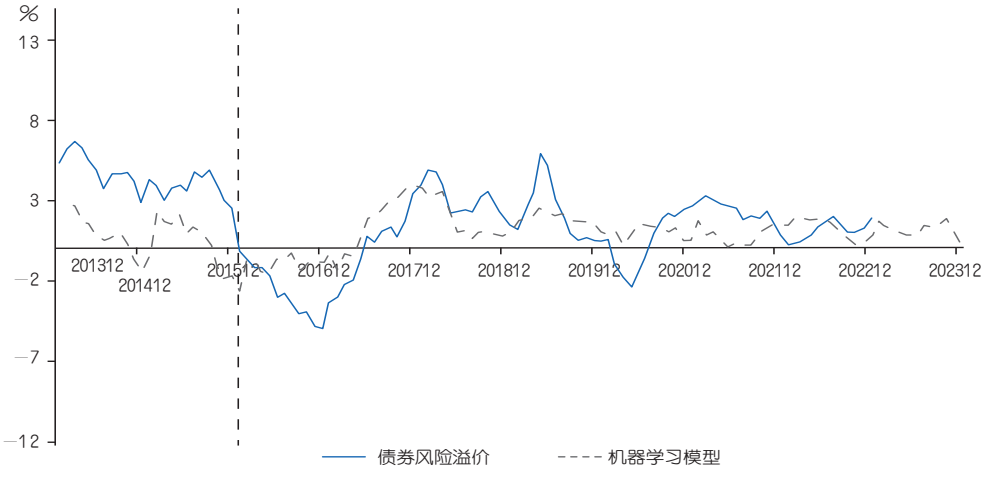

图 6　未来 1 年国债风险溢价预测值及实际值 - 机器学习模型

误差最小的机器学习模型，记录其在 2014 年的样本外预测值。以此类推，最后得到 2014—2022 年各类机器学习模型的样本外预测值，计算各类机器学习模型的样本外预测均值，并与真实值比较。

模型结果表明（图 5~6），2015 年以前债券收益率曲线定价尚不充分、有效信息含量不足，基于远期利率因子的模型样本外预测效果较差，考虑全部宏观因子的机器学习模型占优；2015 年以后采用机器学习模型反而提高了模型的样本外预测误差。具体而言，2015 年以前远期利率单因子模型的样本外预测 MSE 为 25.34，机器学习模型的样本外预测 MSE 为 15.30；2015 年以后远期利率单因子模型的样本外预测 MSE 为 2.67，机器学习模型的样本外预测 MSE 为 3.01。考虑基金久期中位数因子后，样本外预测均方误差则有效下降至 2.06。

以上结果再次印证，2015 年以后我国债券市场收益率曲线定价有效性提升，采用机器学习模型整合宏观因子也无法对长债超额回报提供额外的预测能力。分析背后原因，一方面是彼时我国货币政策框架正从数量型调控向价格型调控转变，利率市场化体系和利率传导机制进一步完善。

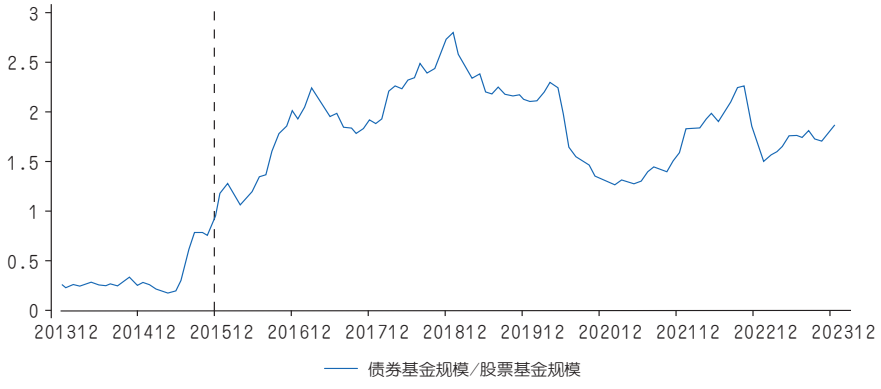

图 7　国内债券基金相对股票基金规模变化

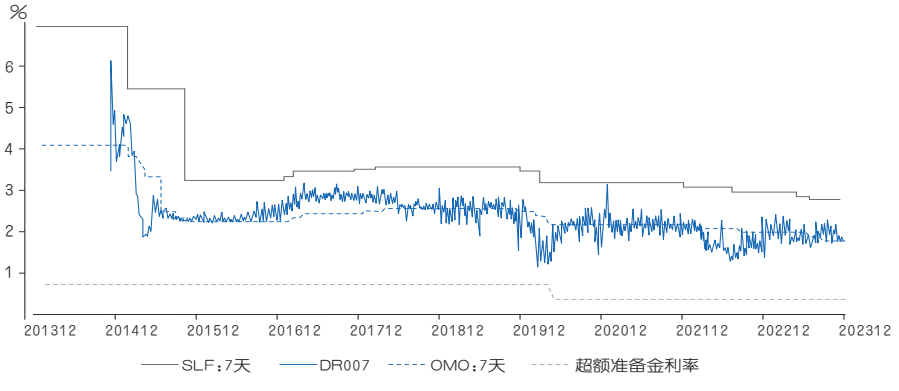

图 8　我国利率走廊构建，市场利率围绕政策利率波动

2015 年四季度《中国货币政策报告》提出“探索利率走廊机制，增强利率调控能力，理顺央行政策利率向金融市场乃至实体经济传导的机制”，标志着人民银行价格型货币调控迈出重要一步，新型货币政策调控框架下，债券市场利率从短端到长端传导更为有效，国债收益率曲线定价有效性提升。另一方面，债券市场深化发展，债券型公募基金、银行理财等资管型机构快速发展，投资群体的多元化、交易活跃度的提升以及利率衍生品市场的发展进一步提升价格发现效率，债券市场定价能够快速反应市场信息（图 7~8）。

六、结论

本文基于 2010—2023 年中国债券市场高频数据集，通过构建动态预测模型与行为金融量化框架，系统检验了预期风险溢价影响因子及机构行为因子对国债风险溢价的预测机制。研究揭示的规律进一步补充了对债券定价理论的理解，也为防范系统性风险提供实务论证，主要结论如下。

（一）远期利率因子的基准性与动态演化

远期利率因子通过深度提取收益率曲线的多维信息，包括利率变动方向、市场波动率水平及曲线曲率特征，为解释长期国债风险溢价的基准性变量，其预测效能随市场制度演进呈现显著提升。传导机制优化方面，2015 年利率市场化改革完成后，收益率曲线从短端向长端的传导效率提升，这直接增强了远期利率因子的样本外预测稳定性。极端环境适应性方面，在市场波动率超过历史 90% 分位数（如 2013 年“钱荒”、2020 年疫情冲击），因子对风险溢价的敏感性提升，反映投资者在不确定性环境中要求更高风险补偿的行为特征。

（二）市场学习效应引发宏观因子效力发生结构性突变

2015 年以前，金融总量数据如社融、M2 等在一定程度上能够预判未来利率走势变化，采用机器学习模型能进一步提高风险溢价的预测效力。2015 年以后，伴随中国利率传导体系完善，债券市场参与主体增多，机构对固定收益业务投研资源加大投入，以公募基金为代表的交易型机构快速发展，债券市场定价有效性提升，宏观经济金融信息能够同步甚至提前反映市场信息，公开的宏观数据对风险溢价不再有额外预测能力，采用机器学习模型对宏观变量进行整合也无法提高样本外预测效果。

（三）机构行为引致的定价扭曲

机构投资者的交易拥挤度对债券风险溢价也具有稳定预测能力。本文发现交易型机构存在一致性预期，在普遍延长投资组合久期至极值后长期债券容易过度定价，导致后期长债超额回报存在反转效应，并且这种现象难以被消除，在业绩排名等现实压力下难有机构能纠正这种异象，同时风险偏好趋同以及正反馈交易策略，使得公募基金行业羊群效应度相对较高，加大市场单边波动。

政策启示与应用路径方面，鉴于国债收益率作为金融体系基准定价锚的基准性地位，建议建立预警监测机制。如建立风险溢价监测指标，将远期利率因子、机构羊群效应纳入评估框架，建立 $\pm 1.5\sigma$ 实时预警系统，并配套开发极端情景压力测试工具（如 200 bps 利率冲击、流动性枯竭情景）。投资者可结合债券风险溢价的测

度及利率风险成本的评估，更好平衡利率风险承担收益与成本，开发基于风险溢价的策略交易模型，指导投资实践，提升投资业绩。

本研究后续可拓展方向，一是开发时变因子权重的马尔可夫区制转换模型，以捕捉资管新规等制度变革对预测机制的冲击；二是建立股债风险溢价联动的 DSGE 一般均衡框架，分析 2022 年理财赎回危机类事件的跨市场传染；三是探索央行预期管理对行为偏差的校正效应，量化市场预期管理至机构行为调整的作用弹性。[N] ①

学术编辑：卢超群

Evaluating the Predictive Power of Underlying Factors on the Treasury Bond Risk Premium

LIU Fanggen[1] JI Weiwei[2] BAI Minghao[2]

(1. Finance and Accounting Department, China Construction Bank;

2. Financial Market Department, China Construction Bank)

Abstract Using Chinese inter-bank treasury bond yield data, macro financial market data and public fund releasing data from 2010 to 2023, this paper uses a single factor regression model and a multi-factor machine learning model to explore the effectiveness of underlying factors on long-term treasury bond risk premium. We find that the predictive power of macro factors on bond risk premium is time-dependent. With the development of the Chinese bond market, the expansion of financial market participants, the learning effect of investors, and information on macroeconomic and financial markets has been reflected in bond pricing in real time. After 2015, macro factors cannot provide additional information on risk premium prediction. Specifically, this article finds that in addition to traditional influential factors, investor behavior is an important factor in bond risk premium. The herd effect and positive feedback strategy in the market exacerbate unilateral volatility, while investor biases have substantial predictive power on bond risk premium. This paper contributes to the quantitative measurement of bond risk premium, the evaluation on bond market pricing rationality, and the monitoring of interest rate risk accumulation.

Keywords Bond Risk Premium, Forward Rate Factor, Machine Learning Model, Institutional Irrational Behavior

JEL Classification G12 G14 G23

① 参考文献[1]~[36]，见增强出版，中国知网—《金融市场研究》。