

## 基于机器学习模型的债券流动性预测

——宏观固收量化研究系列之（五）

东方证券  
ORIENT SECURITIES

## 研究结论

- 流动性是债券市场中非常关键的变量，一般来说，流动性有两个层面的含义。一方面，流动性指的是宏观层面的流动性，衡量的是中央银行以及以商业银行为主的金融机构对资金的供给情况。另一方面，流动性指的是在微观层面上债券变现能力的高低，我们说某个债券流动性好，指的是债券持有者能够及时地按照合意的市场价格卖出该债券。
- 本文把研究重点放在债券微观层面的流动性上。作为中国债券投资者来说，如何准确、合理地评估债券持仓的流动性风险，是当下急需解决的课题。当前市场上大部分关于债券微观流动性的研究，是对债券历史的流动性数据进行评价和打分，并未涉及到对未来不确定性的流动性的评估。而本文侧重于预测与评估债券在未来一段时间的流动性，本质上描述的是债券的流动性风险。
- 当前中国债券市场上对债券流动性的评估主要以定性或者简单打分卡的形式来实现。而本文提供了一种量化的思路，即尝试寻找影响债券未来成交量的潜在特征和合适的统计学模型，对个券的未来成交量做出合理的预测，并对流动性好坏进行打分，从而可以定量地描述个券或者债券组合的流动性情况。
- 本文选取了一系列潜在的影响债券流动性的特征，并测试了不同模型的预测效果，包括打分卡模型、线性模型、Lasso 模型、XGBoost 模型。结果表明，近年来在工业界广泛流行的 XGBoost 模型在预测误差和稳健性上优于其他模型。XGBoost 模型显示，过去一段时间的成交量、存续时间、发行量、剩余期限、到期收益率、Z-Spread、到期收益率的波动率等特征的预测能力较高。
- 为了比较债券在截面上的相对流动性大小，本文基于 XGBoost 模型的预测，构建了流动性评分的规则，对个券的未来流动性进行 0-5 档的映射打分，可以较为直观地展现个券的未来流动性好坏程度。

## 风险提示

- 量化模型失效的风险
- 市场极端环境的冲击

报告发布日期

2021 年 11 月 23 日

证券分析师

朱剑涛

021-63325888\*6077

zhujiantao@orientsec.com.cn

执业证书编号：S0860515060001

联系人

宋之辰

songzhichen@orientsec.com.cn

东方证券股份有限公司经相关主管机关核准具备证券投资咨询业务资格，据此开展发布证券研究报告业务。

东方证券股份有限公司及其关联机构在法律许可的范围内正在或将要与本研究报告所分析的企业发展业务关系。因此，投资者应当考虑到本公司可能存在对报告的客观性产生影响的利益冲突，不应视本证券研究报告为作出投资决策的唯一因素。

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

## 目 录

一、	债券流动性概述 .....	4
二、	个券流动性预测模型 .....	6
2.1	模型介绍 .....	6
2.2	特征选取 .....	8
三、	模型表现 .....	9
3.1	数据 .....	9
3.2	模型表现 .....	10
四、	个券流动性打分 .....	14
	风险提示 .....	21
	参考文献 .....	21

## 图表目录

图 1: 中国债券市场成交量 .....	5
图 2: 中国债券市场成交分布比例(按债券类型) .....	6
图 3: 国债+政金债样本外预测效果(RMSLE) .....	12
图 4: 地方债样本外预测效果(RMSLE) .....	12
图 5: 信用债样本外预测效果(RMSLE) .....	13
图 6: 国债+政金债样本: 特征平均重要性排名前十 .....	14
图 7: 地方债样本: 特征平均重要性排名前十 .....	14
图 8: 信用债样本: 特征平均重要性排名前十 .....	14
图 9: 部分债券的成交量预测与流动性得分历史表现 .....	17
表 1: 国债+政金债样本外预测效果 .....	11
表 2: 地方债样本外预测效果 .....	11
表 3: 信用债样本外预测效果 .....	11
表 4: 各类样本打分规则(对单个截面) .....	15
表 5: 流动性得分为 0 的债券示例(2021/09/24) .....	15
表 6: 流动性得分为 1 的债券示例(2021/09/24) .....	16
表 7: 流动性得分为 4 的债券示例(2021/09/24) .....	16
表 8: 流动性得分为 5 的债券示例(2021/09/24) .....	17

## 一、债券流动性概述

流动性是债券市场中非常关键的变量，一般来说，流动性有两个层面的含义。一方面，流动性指的是宏观层面的流动性，衡量的是中央银行以及以商业银行为主的市场参与者对资金的供给情况。另一方面，流动性衡量的是微观层面的个券变现能力的高低，我们说某个债券流动性好，指的是债券持有者能够及时按照合意的市场价格卖出该债券。本文把研究重点放在微观层面的流动性上，试图通过定量的方法，来刻画个券的流动性。

债券微观流动性的准确定义，在学界和业界并没有一个统一的说法。Harris (1990)比较全面地将资产的流动性定义成四个维度：宽度、深度、即时性和弹性。

- 1) 宽度(width): 在特定的时间内，如果交易买卖双方的价差很小，通常认为该资产的流动性较好。因此往往用买卖价差来衡量宽度的大小，Roll(1984)提出的估计价差、Amihud(2002)提出的非流动性比率等都是从这个维度来刻画债券的流动性；
- 2) 深度(depth): 指的给定价格和时间，量大的交易可以更快的执行。如果某个资产的订单数量多或者成交额大，则可以认为该资产的流动性较好。因此，成交量是一个常用的代理变量，也广泛地被业界采用；
- 3) 即时性(immediacy): 指的是给定数量和价格，买卖双方能够成交的速度。因此往往用平均执行时间来衡量资产成交的即时性；
- 4) 弹性(resiliency): 假定由于大笔订单的成交使得价格上出现较大变化，则偏离均衡水平后恢复均衡价格的速度也是衡量资产流动性的一个维度。

而这些维度，对于流动性程度较低、以场外交易为主的债券市场来说，并非都能找到合理的指标来衡量。Sarig & Warga(1989)认为，在流动性相对较弱的公司债市场，很多债券并不是每天甚至不是每月都有成交，因此即时性维度和弹性维度都很难衡量出债券相对的流动性大小。

出于中国债券市场的交易机制和数据可得性的考虑，本文倾向于从深度的维度来描述个券的流动性大小，即某个债券在未来一段时间内的成交量大小。需要强调的是，当前市场上大部分关于债券微观流动性的研究，是对债券历史的流动性数据进行评价和打分，并未联系到未来不确定性的成交上。事实上，某个债券在未来的流动性高低并非等同于它在过去的流动性高低，一只债券在过去一段时间没有成交量，不代表它在未来的流动性差，或是一只债券在过去的成交量很充足，但未必在未来的流动性好。而本文侧重于预测与评估债券在未来一段时间的流动性，这本质上是对债券未来不确定的流动性描述，与流动性风险紧密相关。

2008 年全球金融危机，整个金融市场都陷入流动性危机的恐慌之中。流动性风险管理，开始逐渐被市场投资者、监管方，还是政策制定者加以重视。而对于债券市场来说，流动性风险更是十分重要的定价因素。2020 年年初的新冠疫情和 11 月发生的“永煤事件”，都对中国债券市场的流动性产生了巨大冲击。

作为债券投资者来说，如何准确、合理地评估投资组合的流动性风险，是当下急需解决的课题。而本文提供了一种量化的思路，即尝试寻找影响债券未来成交量的特征和合适的统计学模型，对个券的未来成交量做出合理的预测，并对流动性进行打分，从而定量地描述个券和债券投资组合的流动性风险。

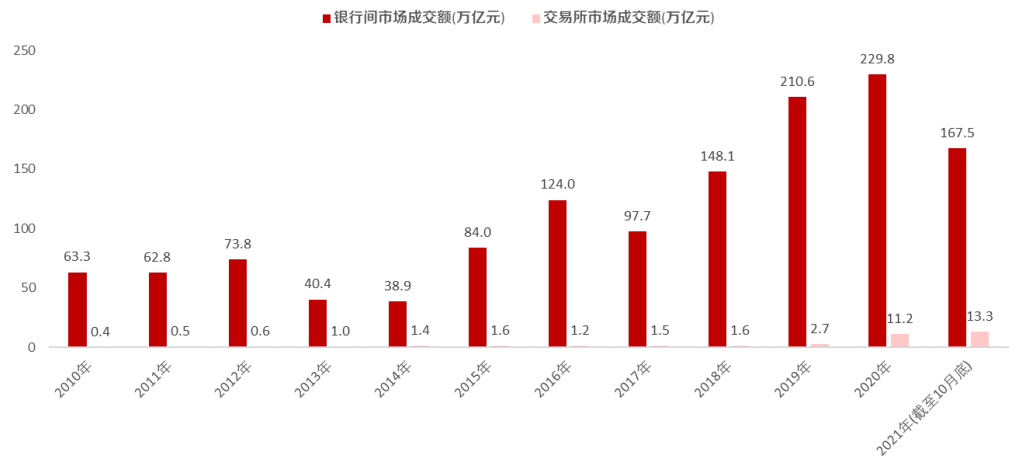
近年来，中国债券市场的规模不断发展壮大，在给实体经济提供直接融资的支持的同时，已经成为全球第二大债券市场。Wind 数据显示，截至 2021 年 10 月末，中国债券市场的债券存量已达 105.7 万亿元(不含同业存单、可转债、可交换债、项目收益票据、ABS、国际机构债、政府支持机构债)，其中利率债存量达 71.5 万亿元，信用债存量达 34.2 万亿元。

从交易总量上看，近几年中国债券市场的成交量也在稳步提升，从 2019 年开始，已经达到 200 万亿的水平，其中银行间市场成交占绝对主导，主要交易方式是经纪商询价。

从结构上看，利率债的成交量占比达到 80%-90%，远高于信用债(这里的信用债口径包括中期票据、企业债、公司债、商业银行债、证券公司债)，而在利率债中，国债、政策性银行债占主要比例，地方政府债的成交较少。市场上对不同品种的债券需求也存在差异，对于国债、政金债，有着较多的交易盘参与，而对于地方债和信用债，配置盘的需求更多，主要的投资策略是持有至到期。

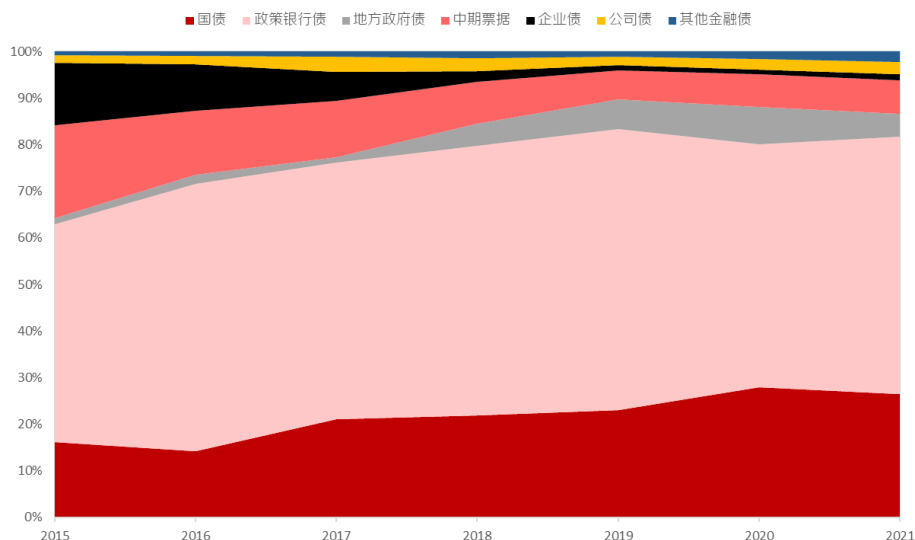
因此，当前的中国债券市场呈现出发债和交易规模逐渐增大，而成交量在不同的债券品种之中，有着结构性分化的特点。因此，在大量债券品种的流动性相对匮乏的市场中，我们更加需要系统地评估个券的未来流动性。

**图 1：中国债券市场成交量**



资料来源：东方证券研究所 & Wind

图 2：中国债券市场成交分布比例(按债券类型)



资料来源：东方证券研究所 & Wind

## 二、 个券流动性预测模型

### 2.1 模型介绍

在介绍预测模型之前，我们首先对变量和模型做一些定义。假设债券  $i$  在一段时间  $t$  的日均成交量为  $Vol_{i,t}$ ，根据本文之前所说，我们预测的是债券  $i$  在下一期的日均成交量  $Vol_{i,t+1}$ ：

$$Vol_{i,t+1} = E_t[Vol_{i,t+1}] + \epsilon_{i,t+1}$$

我们的目标是找到最优的模型  $g^*$  和特征  $X_{i,t}$ ：

$$E_t[Vol_{i,t+1}] = g^*(X_{i,t})$$

使得样本外平均误差(RMSE)最小，或者样本外的  $r^2$  最大。这里的模型  $g$ ，可以是简单的随机游走，也可以是线性模型或者更复杂的机器学习模型。

我们预测的因变量是债券的成交量水平值大小，而样本中成交量数据分布差异较大，如果我们使用传统的 RMSE(Root Mean Squared Error)指标来评估模型的好坏，那么结果会被大成交量的预测误差所主导。为了减轻此现象，本文在拟合参数和调整超参数的过程中，本文采用 RMSLE 指标(Root Mean Squared Logarithmic Error)来取代传统的 RMSE 指标：

$$RMSLE = \sqrt{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \log(1 + Vol_{i,t+1}) - \log(1 + g(X_{i,t})) \right)^2}$$

本文尝试了 4 种不同的统计学模型，分别为

### 1) 打分卡模型

选取  $K$  个重要的特征  $(X^1, X^2, \dots, X^K)$ ，如债券的存续时间、近期日均成交量、债券余额、到期收益率、近期成交天数等，在不同的分组下(比如分评级、分到期期限、分债券类型)，进行标准化处理，取等权平均，得到得分  $Score_{i,t}$ ：

$$Score_{i,t} = \frac{1}{K} (X_{i,t}^1 + X_{i,t}^2 + \dots + X_{i,t}^K)$$

并在训练集中，以  $Score_{i,t}$  为自变量，以  $\log(1 + Vol_{i,t+1})$  为因变量，做线性最小二乘回归(OLS)，估计出  $\hat{\beta}_0, \hat{\beta}_1$ ：

$$\log(1 + Vol_{i,t+1}) = \beta_0 + \beta_1 Score_{i,t}$$

则最终对测试集的样本的  $Vol_{i,t+1}$  预测为：

$$Vol_{i,t+1} = \exp(\hat{\beta}_0 + \hat{\beta}_1 Score_{i,t}) - 1$$

### 2) 线性模型

与打分卡模型类似，同样选取  $K$  个重要的特征  $(X^1, X^2, \dots, X^K)$ ，直接用  $\log(1 + Vol_{i,t+1})$  对这些特征做线性最小二乘回归(OLS)：

$$\log(1 + Vol_{i,t+1}) = \beta_0 + \beta_1 X_{i,t}^1 + \beta_2 X_{i,t}^2 + \dots + \beta_K X_{i,t}^K$$

则最终对测试集的样本的  $Vol_{i,t+1}$  预测为：

$$Vol_{i,t+1} = \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{i,t}^1 + \hat{\beta}_2 X_{i,t}^2 + \dots + \hat{\beta}_K X_{i,t}^K) - 1$$

### 3) Lasso 模型

在以上两种方法中，回归系数的估计方法均是最小化训练集的对数均方误差：

$$L(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \log(1 + Vol_{i,t+1}) - \log(1 + g(X_{i,t}, \beta)) \right)^2$$

对于线性回归模型来说，模型的特征数目越大，复杂度就越高，也就是说发生过拟合的风险就越大。

而 Lasso 模型，对特征的数量施加了约束，它通过构造惩罚项的方法，将某些变量的系数进行压缩，并使得某些回归系数变为 0，进而达到降维，也就是特征选取的效果。

该方法的目标损失函数为：

$$L(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \log(1 + Vol_{i,t+1}) - \log(1 + g(X_{i,t}, \beta)) \right)^2 + \lambda \sum_{j=1}^K |\beta_j|$$

其中， $\lambda$ 是外生参数，控制着 Lasso 模型的复杂度， $\lambda$ 越大，对变量多的情况的惩罚力度就越大，因此结果就会越往变量少的方向优化。

#### 4) XGBoost 模型

以上 3 个模型均为线性模型，线性模型的优点是稳定且逻辑简单，但是仍然会面临问题，一旦特征与因变量之间的关系存在非线性的复杂关系，比如特征之间存在交互关系，线性模型的预测效果就会大大降低。

对于潜在的非线性关系，由 Breiman et al.(1984)率先提出的决策树模型，在业界被广泛地运用，它既可以处理分类问题，也可以处理回归预测问题。与线性模型不同，决策树模型基本思路是：从根节点开始，对样本的某一特征进行二元分类，将样本分配到 2 个子结点，如此递归地对样本进行测试并分配，直至到达叶结点，最终可以将所有的样本划分为  $K$  个区域  $(R_1, R_2, \dots, R_K)$ ，则对任意一个测试集的样本，我们都可以将其划入其中一个区域，并得到对因变量的预测，即训练集中该区域的所有样本的因变量均值：

$$g(X_i) = \sum_{k=1}^K c_k I\{X_i \in R_k\}$$
$$\hat{c}_k = \frac{\sum_{i=1}^N y_i I\{x_i \in R_k\}}{\sum_{i=1}^N I\{x_i \in R_k\}}$$

Friedman(2001)提出了 GBDT(Gradient Boosting Decision Tree)的框架，是集成学习的一个重要方向。前述的决策树模型是弱学习器，可能产生过拟合、预测准确度低、方差大等问题，而集成学习通过不同的算法首先产生很多基学习器，并综合所有基学习器的预测结果来得到性能更强大的学习器。GBDT 的主要思想是根据当前模型损失函数的负梯度方向的信息，对新加入的弱学习器进行训练，然后把训练好的弱学习器更新到当前模型中，从而最终产生性能更为强大的强学习器。

Chen & Guestrin (2016)基于 Gradient Boosting 框架，开发出 XGBoost(eXtreme Gradient Boosting)模型，在模型复杂度控制、并行化处理、损失函数形式等方面进一步优化，运算效率得到更大的提高，在很多工程问题和数据挖掘类竞赛上都取得良好的效果。

本文以回归树模型作为基学习器，将所有可能的特征作为输入来训练 XGBoost 模型，并通过验证集来调整学习率等超参数，最终确定最合适的预测模型，对测试集进行预测。

## 2.2 特征选取

根据海内外参考文献和经验判断，本文选取了一系列可能影响债券未来成交量的特征：



- 1) 过去一段时间的成交量和成交天数：本文选取了过去 20 日的日均成交量、过去 40 日至过去 20 日的日均成交量、过去 60 日至过去 40 日的日均成交量、近 20 日的成交天数、近 60 日的成交天数作为衡量过去流动性水平的特征。
- 2) 债券规模：参考 Chen et al.(2007)，债券余额的大小和成交量的大小显著正相关。
- 3) 存续时间：参考 Sarig 和 Warga(1989)，新券比老券的流动性更好，因此新券相对于老券有更低的到期收益率，老券享有流动性溢价。
- 4) 剩余期限：参考 Longstaff et al.(2005)，临近到期的债券的流动性更好。
- 5) 到期收益率：流动性溢价体现在到期收益率上，因此理论上到期收益率越低的券，流动性更好。对于地方债和信用债，我们使用 Z-spread 来代替到期收益率，从而消除了无风险利率的影响。
- 6) 到期收益率的历史波动率：衡量个券近期的价格风险，理论上，越高的波动率，未来流动性越低。
- 7) 是否是活跃券：对于国债和国开债来说，各期限活跃券的流动性显著高于非活跃券。
- 8) 信用债各分类：是否有担保、是否城投主体、外部评级、债券类型等。
- 9) 地方债各省份：对于地方政府债来说，不同的省份存在不同的利差，因此可能对流动性也有一定程度的影响。

## 三、模型表现

### 3.1 数据

本文使用的是 2016 年 1 月-2021 年 10 月的中国债券数据，包括国债、政金债、地方债、中期票据、企业债、公司债、商业银行债，并根据成交量和特征差异，把这些样本区分成三类分别进行训练和预测——国债+政金债、地方债、信用债(包括中期票据、企业债、公司债、商业银行债、证券公司债)。

不同债券类型之间成交活跃度差异较大，并且银行间债券交易从达成成交到最终录入外汇交易中心系统也存在时滞。考虑到这些原因，对于国债+政金债样本，我们预测未来 10 个交易日的日均成交量；而对于成交相对不活跃的地方债和信用债样本，我们预测未来 20 个交易日的日均成交量。所使用的数据频率均为周频。

债券的基本属性、评级、发行量等数据取自 Wind 资讯，收益率数据取自于中债估值，地方债、信用债的 Z-spread 指标基于中债估值和中债国债即期收益率曲线计算，成交量数据取自外汇交易中心(银行间市场)和 Wind 资讯(交易所市场和固定收益平台)。

对于国债+政金债样本，从 2018 年开始，每季度末训练一次模型，使用每季度末往前 36 个月的数据至往前 3 个月的数据作为训练集，使用近 3 个月的数据作为验证集，下个季度的表现作为测试集。

对于地方债和信用债样本，从 2018 年开始，每季度末训练一次模型，使用每季度末往前 24 个月的数据至往前 3 个月的数据作为训练集，使用近 3 个月的数据作为验证集，下个季度的表现作为测试集。

## 3.2 模型表现

我们关注的是样本外(测试集)的预测效果，因此我们使用以下 2 个评价指标(假定所有迭代产生的测试集为 $\mathcal{T}$ )来评价各个模型：

- 1) 样本外 $R^2$

$$R_{oos}^2 = 1 - \frac{\sum_{(i,t) \in \mathcal{T}} (Vol_{i,t+1} - \widehat{Vol}_{i,t+1})^2}{\sum_{(i,t) \in \mathcal{T}} (Vol_{i,t+1} - \overline{Vol}_{i,t+1})^2}$$

- 2) 样本外 RMSLE(Root Mean Squared Logarithmic Error)

$$RMSLE = \sqrt{\frac{1}{Card(\mathcal{T})} \sum_{(i,t) \in \mathcal{T}} (\log(1 + Vol_{i,t+1}) - \log(1 + \widehat{Vol}_{i,t+1}))^2}$$

除了统计所有样本的 RMSLE，我们还以当期近 20 日成交量为标准划分了不同区间，以观察在模型的不同活跃度样本上的预测效果差异。

下表和下图展示了三类样本的样本外总体表现和分时段表现，我们可以得出以下几点结论：

- 1) 三类样本中，国债+政金债的可预测性最强，这可能是由于国债和政金债的成交较为活跃，但债券数量远少于地方债和信用债，特征较容易捕捉。
- 2) 线性模型、Lasso 模型和 XGBoost 模型的预测效果远好于打分卡模型，这符合我们的直觉，打分卡模型本质上是把多种特征以等权的形式降到一维，在这个过程中损失了大量有用的信息；
- 3) Lasso 模型并未对线性模型改善，说明所选的特征数目较为适中，而样本量较大，所以并未给预测带来过拟合的问题；
- 4) 对国债+政金债样本，XGBoost 模型的表现优于其他模型，并且在不同的成交量分层中也稳定好于其他模型。对信用债样本，XGBoost 模型的表现总体优于其他模型，从成交量分层表现看，XGBoost 模型预测在日均成交量 1 亿以上的券时表现较差(这部分券的数量占比不及 1%)，而预测日均成交量 1 亿以下的券中表现较好。对于地方债样本，XGBoost 模型的表现总体弱于线性模型，但从成交量的分层表现看，XGBoost 模型在日均成交量 5 亿以上的券中表现较差(这部分券的数量占比也不及 1%)，而在日均成交量 5 亿以下的券中表现较好；
- 5) 从分时段的表现上看，XGBoost 模型的表现大部分的时间段都稳定好于其他模型，而打分卡模型表现较差，并且波动性高。

**表 1：国债+政金债样本外预测效果**

	打分卡模型	线性模型	Lasso模型	XGBoost模型
RMSLE	0.837	0.504	0.505	0.494
RMSLE(无成交)	0.298	0.220	0.220	0.216
RMSLE(0-1亿)	0.507	0.421	0.420	0.411
RMSLE(1-5亿)	0.814	0.682	0.681	0.671
RMSLE(5-30亿)	1.120	0.684	0.684	0.675
RMSLE(>30亿)	1.858	0.568	0.575	0.535
R-squared(%)	5.3	76.1	73.6	79.2

资料来源：东方证券研究所 &amp; Wind 资讯

**表 2：地方债样本外预测效果**

	打分卡模型	线性模型	Lasso模型	XGBoost模型
RMSLE	0.146	0.114	0.114	0.115
RMSLE(无成交)	0.073	0.068	0.068	0.067
RMSLE(0-1亿)	0.226	0.199	0.197	0.189
RMSLE(1-5亿)	0.837	0.722	0.713	0.717
RMSLE(5-30亿)	1.588	0.887	0.903	0.969
R-squared(%)	10.6	49.2	45.8	48.3

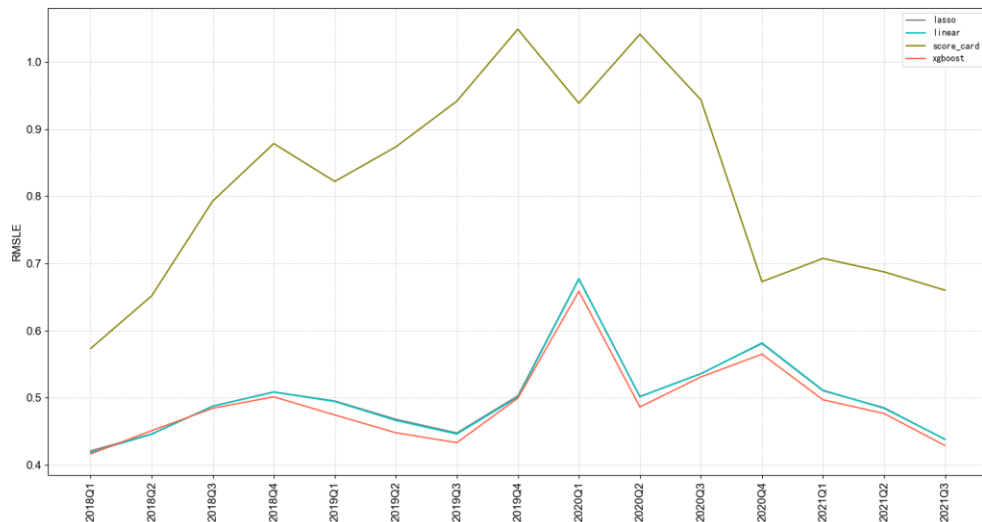
资料来源：东方证券研究所 &amp; Wind 资讯

**表 3：信用债样本外预测效果**

	打分卡模型	线性模型	Lasso模型	XGBoost模型
RMSLE	0.075	0.066	0.067	0.065
RMSLE(无成交)	0.046	0.043	0.043	0.042
RMSLE(0-1亿)	0.087	0.080	0.081	0.078
RMSLE(1-5亿)	0.572	0.412	0.435	0.422
RMSLE(5-30亿)	1.212	0.707	0.772	0.728
R-squared(%)	15.9	39.3	34.0	41.1

资料来源：东方证券研究所 &amp; Wind 资讯

图 3：国债+政金债样本外预测效果(RMSLE)



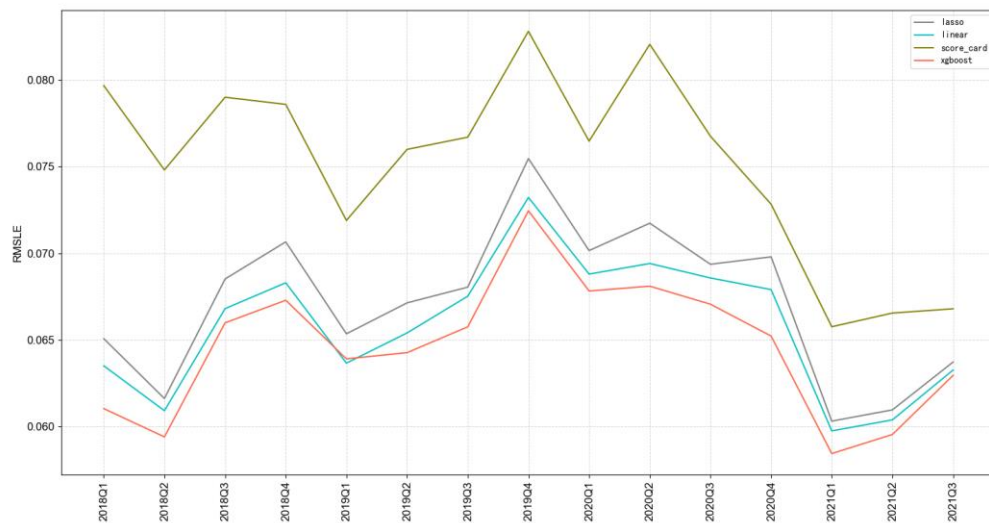
资料来源：东方证券研究所 & Wind

图 4：地方债样本外预测效果(RMSLE)



资料来源：东方证券研究所 & Wind

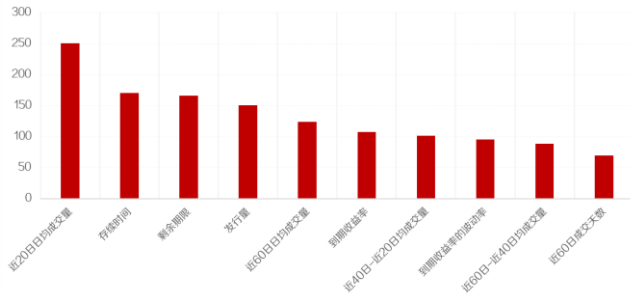
图 5：信用债样本外预测效果(RMSLE)



资料来源：东方证券研究所 & Wind

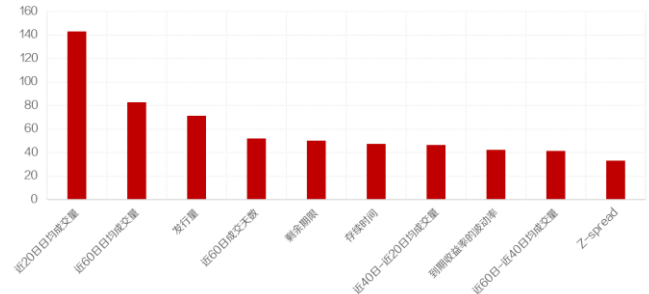
我们统计了每季度最终训练好的模型中各特征的重要性指标——weight，即在一个 XGBoost 模型中的所有决策树中，每个特征总共多少次被用于分割数据。我们对所有训练好的模型的特征重要性取算术平均，得到下图。从 XGBoost 模型的平均特征重要性上看，过去一段时间的成交量、存续时间、发行量、剩余期限、到期收益率、Z-Spread、到期收益率的波动率均有较大的贡献。

图 6：国债+政金债样本：特征平均重要性排名前十



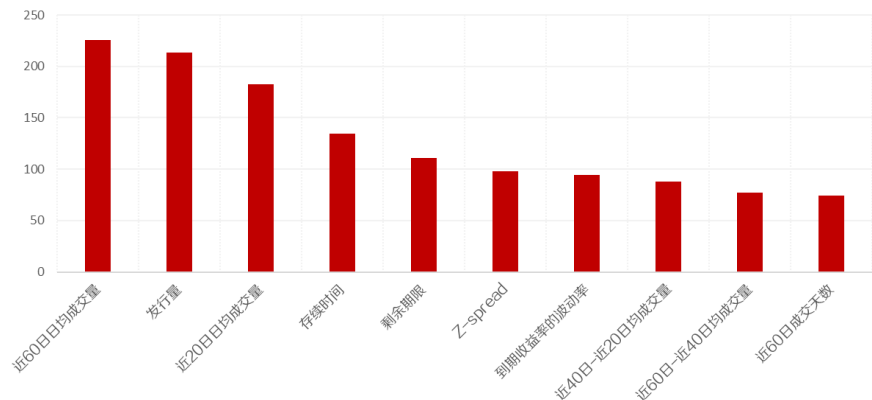
资料来源：东方证券研究所 & Wind

图 7：地方债样本：特征平均重要性排名前十



资料来源：东方证券研究所 & Wind

图 8：信用债样本：特征平均重要性排名前十



资料来源：东方证券研究所 & Wind

## 四、 个券流动性打分

上一部分的研究表明，XGBoost 模型对债券成交量的预测有着较好的效果。而在实际的债券投资中，一方面，投资者可能更关心债券的相对流动性的大小，比如比较几个债券的流动性好坏，或者评估某个组合的整体流动性健康程度。另一方面，债券实际的成交量可能受到宏观环境的影响，比如永煤事件的冲击。出于对这两方面的考虑，本文还基于 XGBoost 模型对未来成交量大小的预测，将每个截面的债券未来成交量映射成 0-5 之间的得分。以便展示债券之间的相对流动性关系。出于对不同品种的债券成交量绝对数值的考虑，我们对不同的样本制定不同的分位数规则，具体如下表所示：

**表 4：各类样本打分规则(对单个截面)**

得分	国债+政金债样本	地方债样本 & 信用债样本
0	全体样本0-40%分位数	预测的日均成交量不超过0.05亿元
1	全体样本40-60%分位数	剩余样本0-70%分位数
2	全体样本60-70%分位数	剩余样本70%-85%分位数
3	全体样本70-80%分位数	剩余样本85%-90%分位数
4	全体样本80-95%分位数	剩余样本90%-95%分位数
5	全体样本95-100%分位数	剩余样本95%-100%分位数

资料来源：东方证券研究所 &amp; Wind 资讯

为了更直观地展示债券的流动性得分，我们以 2021 年 9 月 24 日的截面为例，列出部分债券的 XGBoost 模型预测流动性得分：

**表 5：流动性得分为 0 的债券示例(2021/09/24)**

债券代码	债券名称	债券类型	主体评级	Z-spread(bps)	存续时间(年)	剩余期限(年)	到期收益率(%)	发行量(亿元)	近20日日均成交量(亿元)	未来20日均成交量(亿元)	XGBoost预测值	XGBoost预测流动性评分
110010.IB	11附息国债10	国债	-	-	10.40	9.59	2.87	580.0	0.000	0.000	0.080	0
170027.IB	17附息国债27	国债	-	-	3.75	3.24	2.38	662.2	0.194	0.000	0.304	0
090020.IB	09附息国债20	国债	-	-	12.07	7.92	2.74	260.0	0.054	0.054	0.034	0
140376.IB	14进出76	政策银行债	-	-	6.88	3.11	2.88	50.0	0.000	0.000	0.041	0
110261.IB	11国开61	政策银行债	-	-	9.75	0.24	2.02	20.2	0.201	0.000	0.109	0
010216.IB	01国开16	政策银行债	-	-	19.87	0.12	2.07	100.0	0.125	0.000	0.115	0
1572032.IB	15陕西债32	地方政府债	-	64.6	5.15	4.14	3.25	11.0	0.000	0.000	0.000	0
104536.IB	19四川债32	地方政府债	-	30.6	2.57	7.42	3.15	2.0	0.000	0.000	0.002	0
2005415.IB	20辽宁债13	地方政府债	-	26.6	1.33	28.65	3.65	20.0	0.000	0.000	0.011	0
101800459.IB	18鲁招金MTN001	中期票据	AAA	86.2	3.42	1.58	3.33	10.0	0.000	0.031	0.039	0
101790001.IB	17湘潭高新MTN001	中期票据	AA	504.4	3.80	1.20	7.45	5.0	0.000	0.000	0.021	0
102001118.IB	20商丘发展MTN001	中期票据	AA+	140.5	1.30	1.69	3.89	5.0	0.000	0.030	0.020	0
2026006.IB	20南洋银行02	商业银行债	AAA	111.5	0.76	4.24	3.76	5.0	0.000	0.000	0.006	0
1920062.IB	19宁波银行小微债03	商业银行债	AAA	52.7	1.94	3.06	3.07	10.0	0.000	0.071	0.022	0
1920052.IB	19民泰商行小微债02	商业银行债	AA+	97.9	2.06	0.93	3.31	8.0	0.000	0.050	0.011	0
149378.SZ	21福投02	公司债	AAA	49.0	0.62	2.37	3.01	5.0	0.000	0.000	0.009	0
112836.SZ	18赛格01	公司债	AA	97.8	2.68	0.25	2.91	7.0	0.000	0.000	0.016	0
155702.SH	G19科环	公司债	AA+	77.5	2.00	2.98	3.32	9.0	0.000	0.000	0.015	0

资料来源：东方证券研究所 &amp; Wind 资讯

表 6：流动性得分为 1 的债券示例(2021/09/24)

债券代码	债券名称	债券类型	主体评级	Z-spread(bps)	存续时间(年)	剩余期限(年)	到期收益率(%)	发行量(亿元)	近20日日均成交量(亿元)	未来20日均成交量(亿元)	XGBoost预测值	XGBoost预测流动性评分
170013.IB	17附息国债13	国债	-	-	4.25	2.74	2.39	1086.6	0.840	0.588	1.177	1
170007.IB	17附息国债07	国债	-	-	4.44	0.55	2.17	1100.0	1.141	0.343	2.020	1
180017.IB	18附息国债17	国债	-	-	3.17	26.83	3.49	910.5	1.801	0.358	0.762	1
140222.IB	14国开22	政策银行债	-	-	7.08	2.91	2.79	410.0	0.329	0.359	0.406	1
150310.IB	15进出10	政策银行债	-	-	6.35	0.64	2.42	240.0	1.668	0.414	0.983	1
150305.IB	15进出05	政策银行债	-	-	6.62	3.38	2.90	430.0	1.520	1.376	1.076	1
2105389.IB	21河南债38	地方政府债	-	36.8	0.20	9.79	3.23	2.0	0.056	0.000	0.054	1
147746.IB	18四川07	地方政府债	-	3.8	3.22	1.77	2.53	117.0	0.041	0.282	0.054	1
2105539.IB	21贵州债22	地方政府债	-	24.8	0.15	29.84	3.65	21.9	0.000	0.035	0.062	1
101901355.IB	19冀矿MTN005	中期票据	AAA	123.6	1.94	1.06	3.62	20.0	0.101	0.030	0.074	1
102100057.IB	21镇江城建MTN001	中期票据	AA+	339.9	0.69	2.31	5.92	10.0	0.219	0.015	0.053	1
101901134.IB	19鞍钢MTN005	中期票据	AAA	98.6	2.08	0.91	3.31	15.0	0.101	0.040	0.062	1
2020049.IB	20洛阳银行01	商业银行债	AA+	70.2	1.06	1.94	3.21	30.0	0.121	0.222	0.086	1
2020086.IB	20长沙银行小微债02	商业银行债	AAA	52.4	0.76	2.23	3.05	40.0	0.000	0.380	0.116	1
2020011.IB	20青海银行小微债01	商业银行债	AA+	198.6	1.52	1.48	4.44	30.0	0.119	0.059	0.093	1
188516.SH	21亦庄05	公司债	AAA	89.7	0.12	4.86	3.60	15.0	0.050	0.199	0.058	1
143731.SH	18金隅01	公司债	AAA	52.0	3.17	1.80	3.01	15.0	0.030	0.126	0.062	1
155182.SH	19CHINE02	公司债	AAA	48.3	2.55	0.43	2.70	30.0	0.000	0.066	0.066	1

资料来源：东方证券研究所 & Wind 资讯

表 7：流动性得分为 4 的债券示例(2021/09/24)

债券代码	债券名称	债券类型	主体评级	Z-spread(bps)	存续时间(年)	剩余期限(年)	到期收益率(%)	发行量(亿元)	近20日日均成交量(亿元)	未来20日均成交量(亿元)	XGBoost预测值	XGBoost预测流动性评分
200009.IB	20附息国债09	国债	-	-	1.22	1.77	2.48	2662.8	26.292	17.984	25.367	4
200005.IB	20附息国债05	国债	-	-	1.45	3.54	2.69	3292.9	16.884	18.035	15.460	4
190208.IB	19国开08	政策银行债	-	-	2.23	2.77	2.79	1724.5	10.281	15.924	10.194	4
180211.IB	18国开11	政策银行债	-	-	3.11	1.89	2.67	1808.6	12.050	11.676	10.649	4
160207.IB	16国开07	政策银行债	-	-	5.57	1.42	2.62	1570.0	10.493	7.780	8.607	4
104675.IB	19安徽债13	地方政府债	-	19.2	2.09	2.91	2.73	236.7	0.603	0.552	0.310	4
2105702.IB	21广东债75	地方政府债	-	17.7	0.09	29.90	3.59	101.0	1.084	0.894	0.283	4
2105469.IB	21江苏债14	地方政府债	-	36.7	0.16	9.84	3.23	166.2	0.724	0.703	0.344	4
102101489.IB	21中石化MTN003	中期票据	AAA	43.5	0.12	1.87	2.94	20.0	0.295	0.339	0.175	4
101901211.IB	19东航MTN001	中期票据	AAA	71.2	2.05	0.95	3.05	30.0	0.338	0.423	0.151	4
101900328.IB	19沪港务MTN001	中期票据	AAA	47.8	2.53	0.47	2.73	40.0	0.135	0.291	0.160	4
2128020.IB	21招商银行小微债02	商业银行债	AAA	44.3	0.30	2.70	2.98	200.0	0.875	1.434	1.240	4
1628022.IB	16交行绿色金融债02	商业银行债	AAA	50.6	4.84	0.16	2.46	200.0	1.017	0.681	0.710	4
163216.SH	20世茂G1	公司债	AAA	652.8	1.54	1.44	8.98	20.0	0.243	0.274	0.132	4
188400.SH	21银河G6	证券公司债	AAA	74.5	0.16	4.82	3.45	18.0	0.381	0.388	0.187	4
188170.SH	21中电Y1	公司债	AAA	86.8	0.23	2.76	3.41	30.0	0.041	0.273	0.138	4
2080105.IB	20厦门轨道债01	企业债	AAA	76.2	1.41	3.58	3.34	30.0	0.251	0.200	0.108	4

资料来源：东方证券研究所 & Wind 资讯



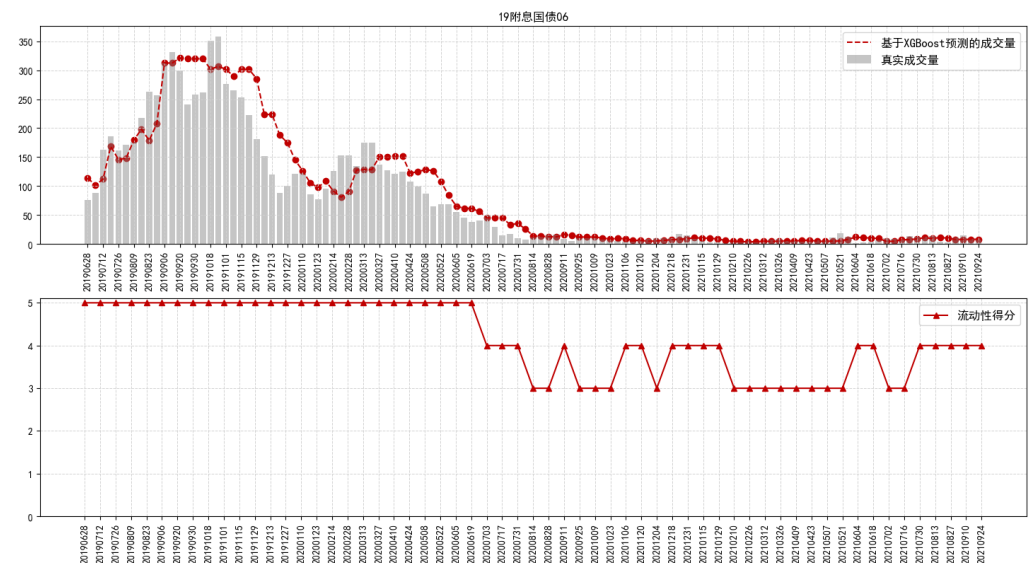
表 8：流动性得分为 5 的债券示例(2021/09/24)

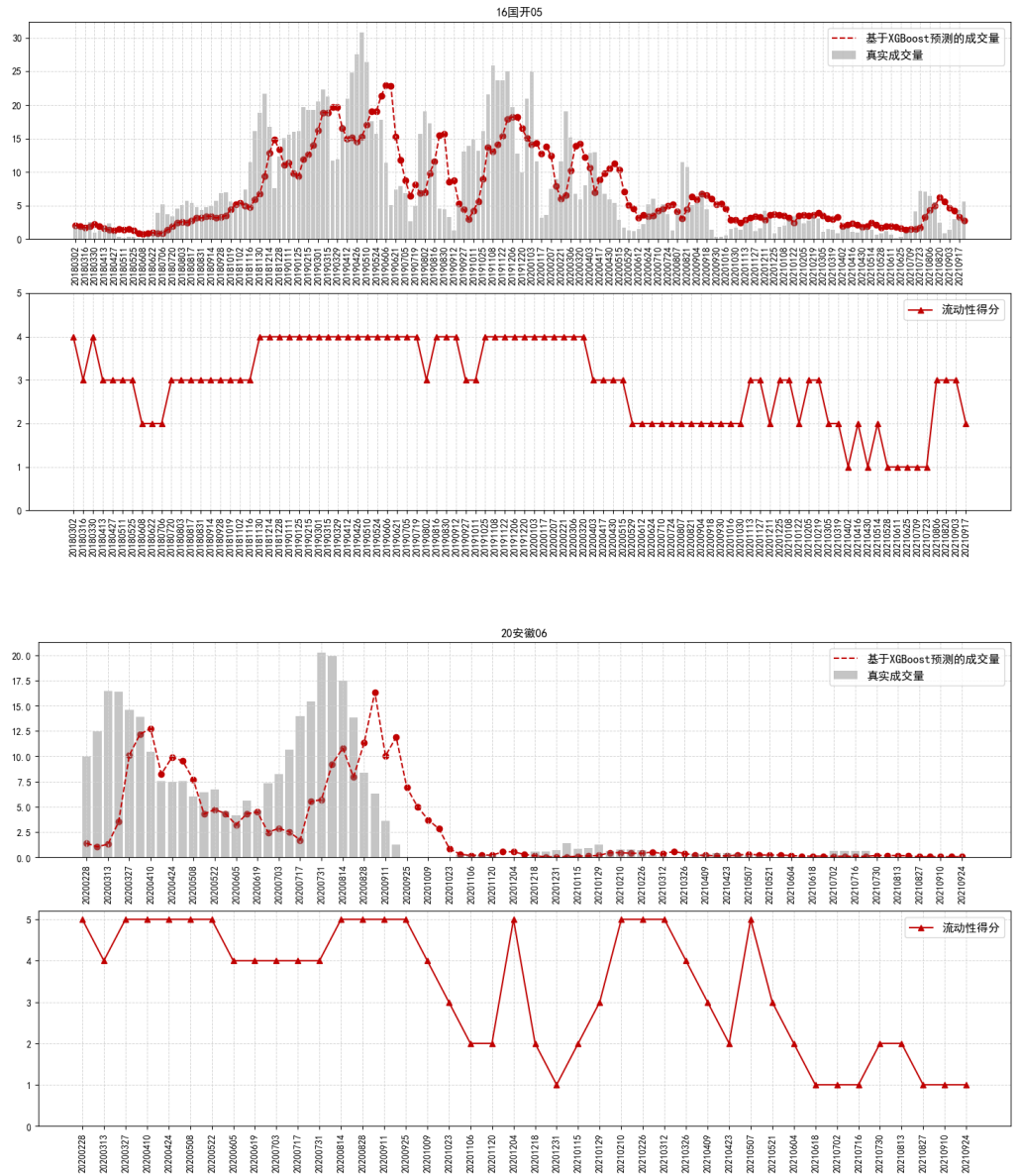
债券代码	债券名称	债券类型	主体评级	Z-spread(bps)	存续时间(年)	剩余期限(年)	到期收益率(%)	发行量(亿元)	近20日日均成交量(亿元)	未来20日均成交量(亿元)	XGBoost预测值	XGBoost预
200006.IB	20附息国债06	国债	-	-	1.33	8.65	2.98	3393.6	53.031	62.934	40.466	5
200016.IB	20附息国债16	国债	-	-	0.84	9.15	2.90	3421.0	52.562	44.816	44.021	5
200215.IB	20国开15	政策银行债	-	-	0.92	9.07	3.30	2736.6	44.399	38.988	42.465	5
210207.IB	21国开07	政策银行债	-	-	0.26	2.73	2.76	828.4	43.297	41.279	43.054	5
147885.IB	18吉林03	地方政府债	-	25.5	3.27	6.72	3.07	70.0	4.125	0.742	1.668	5
2105067.IB	21河北债01	地方政府债	-	26.0	0.51	4.48	2.93	170.6	1.088	0.955	0.527	5
2105408.IB	21陕西债08	地方政府债	-	28.0	0.20	4.79	2.98	62.9	1.379	0.791	0.574	5
102000048.IB	20中金集MTN001	中期票据	AAA	54.2	1.69	1.31	2.97	30.0	0.156	0.438	0.176	5
101900950.IB	19华润MTN005	中期票据	AAA	50.3	2.18	0.82	2.79	50.0	0.474	0.523	0.253	5
102000806.IB	20津城建MTN003	中期票据	AAA	218.8	1.42	1.58	4.66	15.0	0.426	0.404	0.189	5
1928037.IB	19交通银行02	商业银行债	AAA	34.4	1.78	1.22	2.76	400.0	2.443	1.670	1.388	5
1928034.IB	19交通银行01	商业银行债	AAA	34.4	1.82	1.18	2.75	500.0	1.365	1.961	1.699	5
175361.SH	20茅台01	公司债	AAA	17.4	0.87	2.12	2.97	130.0	0.671	0.315	0.549	5
188228.SH	21中企01	公司债	AA+	74.3	0.26	2.72	3.26	30.9	0.086	0.331	0.179	5
143891.SH	18疏浚01	公司债	AAA	72.2	2.88	0.08	2.95	40.0	0.897	1.232	0.343	5
122402.SH	15城建01	公司债	AAA	62.9	5.95	0.82	2.92	58.0	0.917	0.520	0.342	5
188502.SH	21宁铁14	公司债	AAA	60.5	0.12	2.86	3.15	28.0	0.275	0.298	0.173	5
1780416.IB	17西安高新债01	企业债	AAA	149.8	3.73	3.26	4.06	50.0	0.124	0.219	0.155	5

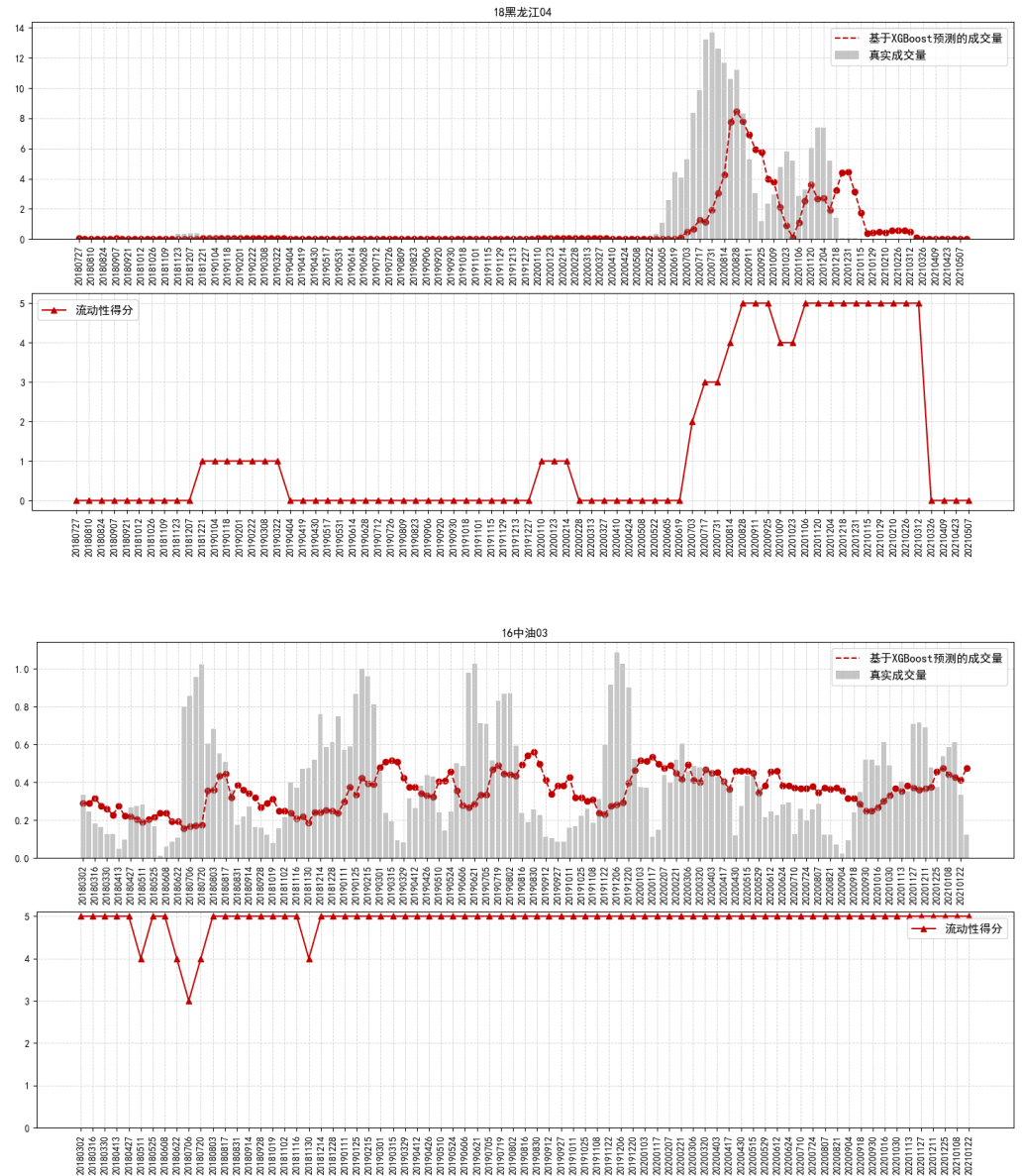
资料来源：东方证券研究所 & Wind 资讯

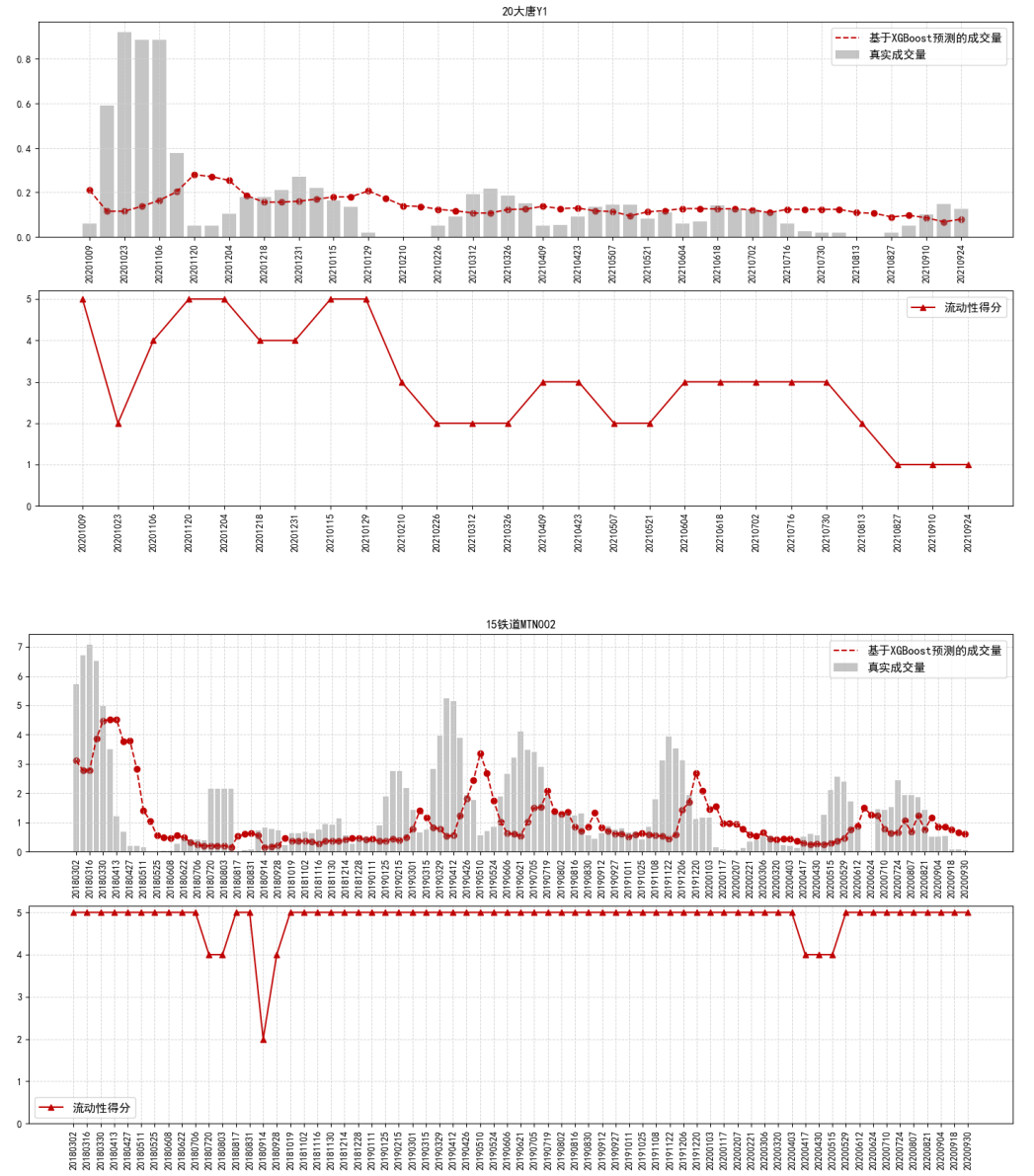
除了展示单个截面上的示例，我们还列出部分债成交量预测和流动性得分的时间序列(周频)作为参考。

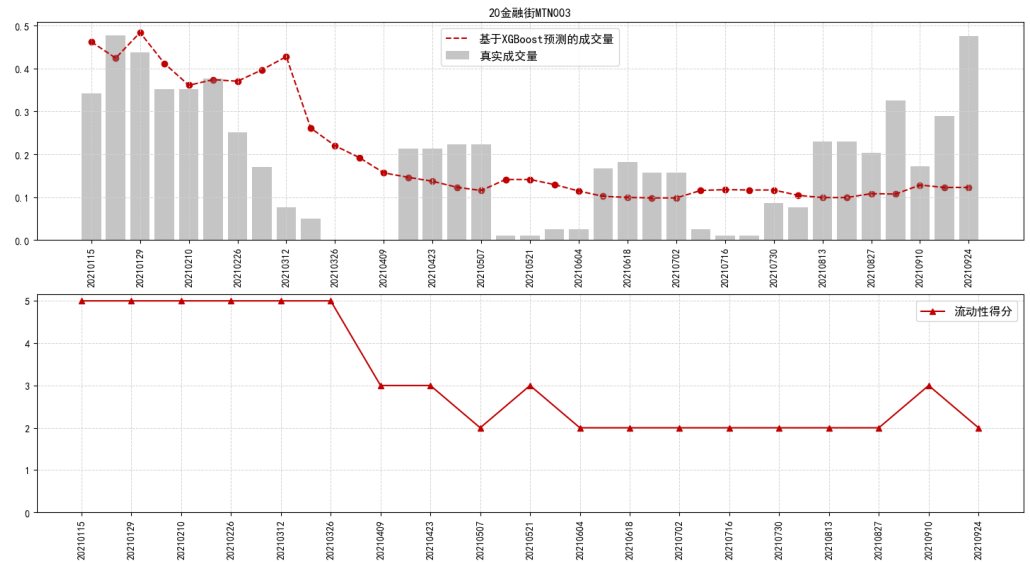
图 9：部分债券的成交量预测与流动性得分历史表现











资料来源：东方证券研究所 & Wind 资讯

完成了对个券的流动性成交量的预测和流动性评分，可以在指数化投资、组合流动性风险管理等问题上有应用前景：

- 1) 指数化投资。不同于蓬勃发展的股票指数化投资，当下中国债券指数仍以覆盖所有样本的总量指数为主，缺乏由成交活跃的债券构成的指数。有了对债券的流动性量化评分，我们可以实现对不同流动性层级的债券样本筛选，因此可以尝试在流动性较好的债券池中构建选债策略，增加了该类投资的可操作性；
- 2) 组合流动性风险管理。对于债券持仓组合来说，完成了对债券流动性预测和评分之后，可以从定量的角度来迅速捕捉持仓债券的流动性风险暴露和组合的流动性整体水平。

## 风险提示

- 量化模型失效的风险
- 市场极端环境的冲击

## 参考文献

- [1] Amihud, Yakov. "Illiquidity and stock returns: cross-section and time-series effects." *Journal of financial markets* 5.1 (2002): 31-56.

- [2] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.
- [3] Friedman, Jerome H. "Greedy function approximation: a gradient boosting machine." *Annals of statistics* (2001): 1189-1232.
- [4] Harris, Lawrence. "Statistical properties of the Roll serial covariance bid/ask spread estimator." *The Journal of Finance* 45.2 (1990): 579-590.
- [5] Roll, Richard. "A simple implicit measure of the effective bid - ask spread in an efficient market." *The Journal of Finance* 39.4 (1984): 1127-1139.
- [6] Sarig, Oded, and Arthur Warga. "Bond price data and bond market liquidity." *Journal of Financial and Quantitative Analysis* 24.3 (1989): 367-378.

## 分析师申明

每位负责撰写本研究报告全部或部分内容的研究分析师在此作以下声明：

分析师在本报告中对所提及的证券或发行人发表的任何建议和观点均准确地反映了其个人对该证券或发行人的看法和判断；分析师薪酬的任何组成部分无论是在过去、现在及将来，均与其在本研究报告中所表述的具体建议或观点无任何直接或间接的关系。

## 投资评级和相关定义

报告发布日后的 12 个月内的公司的涨跌幅相对同期的上证指数/深证成指的涨跌幅为基准；

### 公司投资评级的量化标准

买入：相对强于市场基准指数收益率 15%以上；

增持：相对强于市场基准指数收益率 5% ~ 15%；

中性：相对于市场基准指数收益率在-5% ~ +5%之间波动；

减持：相对弱于市场基准指数收益率在-5%以下。

未评级 —— 由于在报告发出之时该股票不在本公司研究覆盖范围内，分析师基于当时对该股票的研究状况，未给予投资评级相关信息。

暂停评级 —— 根据监管制度及本公司相关规定，研究报告发布之时该投资对象可能与本公司存在潜在的利益冲突情形；亦或是研究报告发布当时该股票的价值和价格分析存在重大不确定性，缺乏足够的研究依据支持分析师给出明确投资评级；分析师在上述情况下暂停对该股票给予投资评级等信息，投资者需要注意在此报告发布之前曾给予该股票的投资评级、盈利预测及目标价格等信息不再有效。

### 行业投资评级的量化标准：

看好：相对强于市场基准指数收益率 5%以上；

中性：相对于市场基准指数收益率在-5% ~ +5%之间波动；

看淡：相对于市场基准指数收益率在-5%以下。

未评级：由于在报告发出之时该行业不在本公司研究覆盖范围内，分析师基于当时对该行业的研究状况，未给予投资评级等相关信息。

暂停评级：由于研究报告发布当时该行业的投资价值分析存在重大不确定性，缺乏足够的研究依据支持分析师给出明确行业投资评级；分析师在上述情况下暂停对该行业给予投资评级信息，投资者需要注意在此报告发布之前曾给予该行业的投资评级信息不再有效。

## 免责声明

本证券研究报告（以下简称“本报告”）由东方证券股份有限公司（以下简称“本公司”）制作及发布。

。本公司不会因接收人收到本报告而视其为本公司的当然客户。本报告的全体接收人应当采取必要措施防止本报告被转发给他人。

本报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，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证券研究报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的证券研究报告之外，绝大多数证券研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人作出邀请。

本报告中提及的投资价格和价值以及这些投资带来的收入可能会波动。过去的表现并不代表未来的表现，未来的回报也无法保证，投资者可能会损失本金。外汇汇率波动有可能对某些投资的价值或价格或来自这一投资的收入产生不良影响。那些涉及期货、期权及其它衍生工具的交易，因其包括重大的市场风险，因此并不适合所有投资者。

在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者自主作出投资决策并自行承担投资风险，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。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面协议授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容。不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

经本公司事先书面协议授权刊载或转发的，被授权机构承担相关刊载或者转发责任。不得对本报告进行任何有悖原意的引用、删节和修改。

提示客户及公众投资者慎重使用未经授权刊载或者转发的本公司证券研究报告，慎重使用公众媒体刊载的证券研究报告。

---

## 东方证券研究所

地址：上海市中山南路 318 号东方国际金融广场 26 楼

电话：021-63325888

传真：021-63326786

网址：[www.dfzq.com.cn](http://www.dfzq.com.cn)