

债券量化系列之一：

宏观数据与非线性模型



报告日期：2022 年 1 月 5 日

★主要内容

本文探讨三个利率量化择时中可能遇到的问题：一是**宏观指标的滚动窗口线性预测回归**，宏观指标对利率的影响呈现一定的周期性特征，通过加入经济周期状态变量可一定程度上提升有效预测占比。二是**降维模型的应用**，传统的因子降维模型往往建立在输入数据集都包含有效信息的基础上，更大量数据的加入并不一定会带来更有效的结果。基于单因子滚动窗口筛选的结果再进行降维，或可以提升模型稳健性。三是**机器学习模型在债券超额收益预测中的应用实证**，基于海外市场研究，机器学习模型在债券超额收益预测中取得远优于传统模型的效果，我们采用 XGBoost 模型在中债市场的实证研究亦得到相似的结论。

从债券超额收益的样本外预测误差角度对不同模型效果进行总结：首先，机器学习 XGBoost 回归树模型具有非常显著的模型优势；其次，对于线性预测框架而言，进行宏观数据的单因子滚动窗口筛选具有一定的提升效果；最后，对于不同期限超额收益预测效果对比，传统模型对中长期预测效果相比较弱，机器学习模型结果对于不同期限预测均较为稳健。

利率债久期择时策略表现方面，机器学习模型相对于线性模型具有显著优势，加入宏观数据的择时效果更优：采用机器学习模型 XGBoost 久期择时策略效果方面，基于国债财富指数实现年化收益 5.05%，最大回撤-2.58%，年化波动 2.56%，夏普比率 1.97，平均持仓周期 3.4 个月，回测区间为 2011.04 至 2021.12；基于国开债财富指数实现年化收益 6.6%，最大回撤-2.2%，年化波动 3.2%，夏普比率 2.08，平均持仓周期 3.2 个月，回测区间为 2013.02 至 2021.12。

★风险提示

量化模型有效性基于历史数据得出，不排除失效的可能。

★致谢

感谢东方证券金工首席分析师朱剑涛老师的指导与帮助。

王冬黎 高级分析师(金融工程)
从业资格号： F3032817
投资咨询号： Z0014348
Tel: 8621-63325888-3975
15802167045
Email: dongli.wang@orientfutures.com

目录

1. 主要内容	5
2. 经济周期与宏观指标滚动窗口回归	6
2.1. 指标选择与数据处理	6
2.2. 经济周期划分的参考指标与划分结果	7
2.3. 加入周期哑变量的回归模型构建	9
3. 基于债券超额收益预测进行久期择时	10
3.1. 零息债券超额收益计算与归因	10
3.2. 基于远期利率与宏观数据预测债券超额收益	12
3.2.1. 债券超额收益预测效果分析	13
3.2.2. 中债财富指数的复制和预测	15
3.2.3. 基于持有收益预测的久期轮动策略	16
4. 附件	20
4.1. 经济指标处理口径调整	20
4.2. 经济周期划分方法	22
4.2.1. 回顾分析: Bry-Boschan 算法	22
4.2.2. 前瞻判断: 趋势+绝对值突破	24
5. 风险提示	26

图表目录

图表 1: 单因子回归数据处理说明.....	6
图表 2: 经济周期划分——通胀、景气度与信贷指标走势.....	7
图表 3: 三因子拐点时间节点列表 (Bry-Boschan 方法)	8
图表 4: 不同经济周期下的利率表现总结整理 (Bry-Boschan 方法)	8
图表 5: 不同经济周期下宏观指标的有效预测占比 (ER)	8
图表 6: 不同期限利率预测单因子有效预测占比 (近两年平均)	9
图表 7: 国债零息债券 (3Y) 持有收益拆解	11
图表 8: 国债零息债券 (7Y) 持有收益拆解	11
图表 9: 不同期限国债零息债券远期利率、水平因子和其他因素波动贡献.....	11
图表 10: 债券超额收益预测不同输入/模型介绍.....	13
图表 11: 债券超额收益预测模型基本设定.....	13
图表 12: 不同模型下国债零息债券各久期超额收益预测样本外 R2.....	14
图表 13: 不同模型中债零息债券超额收益预测 (N=5)	14
图表 14: 不同模型中债零息债券超额收益预测 (N=10)	15
图表 15: 中债国债 7-10 年财富指数复制权重	15
图表 16: 中债国债 7-10 年财富指数复制净值.....	15
图表 17: 中债国债财富指数一年期持有收益预测值 (process6)	16
图表 18: 中债国债久期择时的配置期限与利率走势 (process6)	17
图表 19: 中债国债财富指数久期轮动策略净值.....	17
图表 20: 中债国债财富指数久期轮动策略净值分析	18
图表 21: 中债国开债财富指数久期轮动策略净值.....	18
图表 22: 中债国开债财富指数久期轮动策略净值分析	19
图表 23: 经济周期因子之通胀走势.....	20
图表 24: 经济周期因子之景气度走势.....	20
图表 25: 口径调整后的社会融资规模当月值.....	21
图表 26: 国债地方债净融资与社融政府债券分项.....	21
图表 27: 社会融资规模口径调整结果 (MA12)	21
图表 28: 经济周期因子之信贷脉冲走势	21
图表 29: 通胀因子走势与周期拐点确认结果 (基于 Bry-Boschan 方法)	23
图表 30: 景气度因子走势与周期拐点确认结果 (基于 Bry-Boschan 方法)	23
图表 31: 信贷因子走势与周期拐点确认结果 (基于 Bry-Boschan 方法)	24
图表 32: 通胀因子走势与周期拐点确认结果 (基于势平滑加辅助判断条件)	24

图表 33: 景气度因子走势与周期拐点确认结果 (基于势平滑加辅助判断条件)	25
图表 34: 信贷因子走势与周期拐点确认结果 (基于势平滑加辅助判断条件)	25

1. 主要内容

本文主要研究重点在于对三个利率量化择时中可能遇到的问题进行探讨：**一是宏观指标的单因子回归预测**。我们研究发现宏观指标对利率的影响呈现一定的周期性特征，进而降低单因子滚动窗口预测的显著性，我们通过纳入经济周期所处的状态作为哑变量辅助宏观指标的单因子回归的方法，可以一定程度上提升有效预测占比。

二是降维模型的应用。传统的因子模型在对于宏观数据降维的应用中存在问题，其中较为突出的是模型对输入极其敏感，因其假设就建立在输入数据集都包含有效信息的基础上，更大量数据的加入并不一定会带来更有效的结果。我们对传统降维模型的应用进行一定改进，在上述单因子滚动窗口筛选的框架下，对有效单因子进行 PCA 降维，进而确保输入信息的有效性。

三是机器学习模型在债券超额收益预测中的应用实证。债券超额收益的预测在海外固收量化类文献中已有较多的探讨，债券超额收益即为 N 年期债券的持有收益相对持有期即期利率的超额回报。预测的基准做法是基于远期利率曲线直接对超额收益进行预测，海外文献 (Bond Risk Premia, Cochrane and Piazzesi, 2005) 研究表明远期利率曲线对债券超额收益具有预测能力，我们也基于国内市场的实证研究得到较为一致的结论。此外，在远期利率曲线的基础上，不少文献研究宏观指标的加入是否能够显著提升预测效果，Macro Factors in Bond Risk Premia (Ludvigson and Ng, 2009) 采用动态因子的方法对宏观数据进行分析，论证了宏观指标可提升预测效果。宏观数据对利率影响往往存在非线性的特征，随着机器学习技术在金融领域应用的深入，Bond Risk Premia with Machine Learning (Bianchi, Buchner and Tamoni, 2017) 研究表明验证采用回归树、神经网络等机器学习方法基于远期利率曲线和宏观数据的集合预测债券超额收益的取得远优于传统模型的效果。

本文的文章结构的划分上，我们将在第二部分主要介绍基于宏观指标的单因子回归处理，包括宏观数据选择与处理、经济周期划分以及宏观指标单因子回归模型构建。文章第三部分我们将基于债券超额收益预测框架，探讨基于远期利率曲线与宏观数据进行债券超额收益预测的实证效果，并基于债券财富指数构建久期轮动策略，包括机器学习非线性模型与传统降维模型的效果对比，考虑到算力有限机器学习模型方面我们主要使用 XGBOOST 回归树进行预测。

结论方面，**基于债券超额收益的样本外预测误差角度对不同模型效果进行总结**：1) 机器学习 XGBoost 回归树模型具有非常显著的模型优势，对于不同解释变量数据集的处理效果均远优于 PCA 降维与线性预测模型；2) 对于线性预测框架而言，进行宏观数据的单因子滚动窗口筛选具有一定的提升作用效果。**基于对中债国债与国开债财富指数的久期择时效果模型比较方面**：1) 机器学习模型相对于线性模型具有显著优势；2) 加入宏观数据的久期择时效果相对于仅基于远期利率曲线效果更优。**采用机器学习模型 XGBoost 久期择时策略效果方面**，基于国债财富指数实现年化收益 5.05%，最大回撤 -2.58%，年化波动 2.56%，夏普比率 1.97，平均持仓周期 3.4 个月，回测区间为 2011.04 至 2021.12；基于国开债财富指数实现年化收益 6.6%，最大回撤 -2.2%，年化波动 3.2%，夏普比率 2.08，平均持仓周期 3.2 个月，回测区间为 2013.02 至 2021.12。

2. 经济周期与宏观指标滚动窗口回归

在线性回归框架下，基于宏观数据进行利率择时中遇到的主要问题在于指标解释能力不高，滚动窗口预测结果多为不显著。在不改变线性回归模式的情况下，我们可以通过引入经济周期所处状态作为哑变量对模型进行简单改进。其背后的逻辑在于，不同角度的宏观数据对于价格的影响程度很可能是与周期所处位置有关的，即经济周期轮动下市场主要矛盾的演绎。我们基于通胀、景气度与信贷三因子模式对经济周期进行划分，由于数据长度有限粗略划分为三个主要周期（每个周期约 50-60 个月左右的累计时长），不同周期下利率表现具有显著差异，周期研究或为大类资产配置的主要内容，对于利率择时而言，简单通过周期划分得到的信息比较粗糙，但其与宏观指标单因子相结合可有效提升回归解释能力。我们将基于前瞻性方式判断的周期状态作为宏观指标单因子回归的额外哑变量，对滚动窗口回归的有效预测占比起到较好的提升作用。尽管复杂的非线性模型如机器学习在解决非线性问题上效果更为优异，我们依然对简单模型进行优化，主要原因在于首先简单模型下更有助于量化与主观逻辑的结合，其次简单模型的参数拟合误差风险低。

2.1. 指标选择与数据处理

我们拟定部分宏观数据与收益率动量数据作为待筛选指标池，宏观数据包括 PMI 及其分项、PPI 及其分项、CPI 及其分项，以及社融与信贷数据原值、同比多增与存量同比；收益率动量数据为关键期限中债曲线收益率、利率互换与货币市场利率利差相关指标差分形式。考虑对原始数据的不同处理形式，指标池具有包含经济周期状态哑变量在内的共计 216 个指标。

图表 1：单因子回归数据处理说明

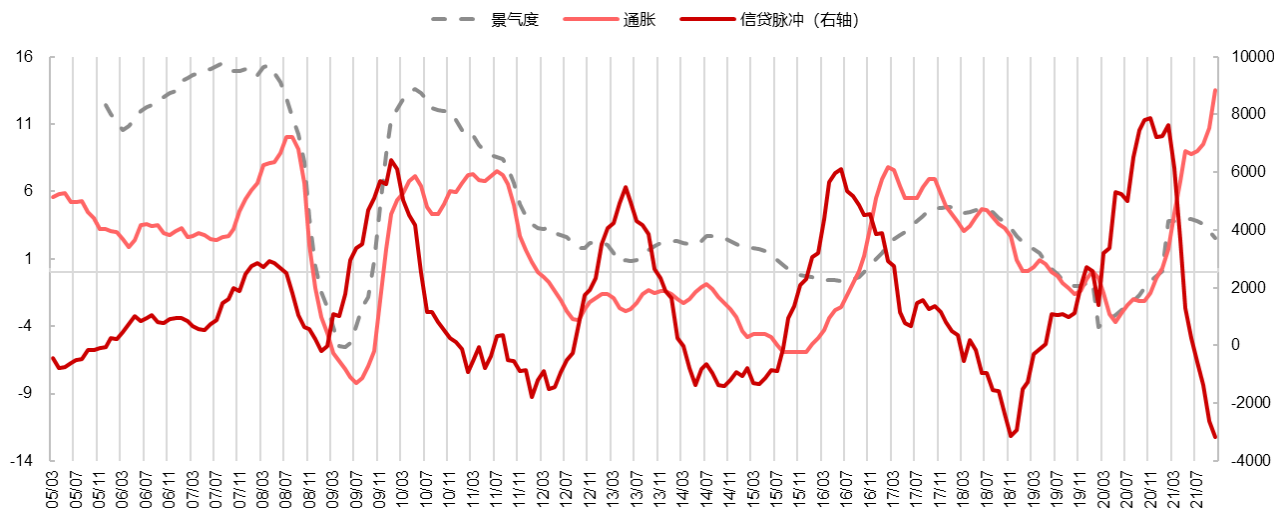
数据处理	详细说明
数据频率	月度频率、高频数据取最新一期的值处理成月度数据
可得性处理	宏观数据除 PMI 外均统一滞后一个月处理，动量等高频数据按其数据频率滞后一期
数据变化形式	宏观数据保留原值与差分，动量数据仅保留差分，注意数据差分计算均为月度频率的差分
数据标准化	基于中国经济周期规律特征，选取三年窗口，对宏观数据原值、差分，动量数据差分进行滚动标准化（ZSCORE）
回归假设检验	单因子线性回归前均对所有指标进行 ADF 检验，不通过则再次计算差分，若依然不通过则回归时跳过该指标。（机器学习模型无需此步骤）
回归被解释变量	中债收益率曲线某期限即期利率的月度变化

资料来源：东证衍生品研究院

2.2. 经济周期划分的参考指标与划分结果

我们基于通胀、景气度与信贷三因子确认周期位置，对应指标选择为 PPI、PMI 与社融增量，首先我们对划分经济周期的主要参照指标进行数据处理，为去除一定的季节性并尽量统一指标形式，采用移动平均的平滑处理。PPI 本身为同比序列无需额外调整。PMI 为环比数据，我们将其进行 12 个月移动平均处理得到同比数据。对于信贷指标，我们基于社融规模指标我们基于政府债券历史数据进行口径调整后，计算社融规模增量的 12 个月移动平均作为信贷脉冲。经济周期划分的方法方面，我们先基于 Bry-Boschan (1971) 的方法划分用于简单分析，该方法具有事后性质不能用于预测模型，但因其得到的周期划分更清晰有助于分析研究。其次采用趋势模型进行前瞻性的周期状态判断，进而用于策略中。经济周期指标处理、两种经济周期划分方法介绍与对三因子拐点的判断结果详见附件。

图表 2：经济周期划分——通胀、景气度与信贷指标走势



资料来源：Wind，东证衍生品研究院

基于 Bry-Boschan 周期划分算法与传统的通胀、景气度与信贷三因子框架进行周期划分，参考市场上对经济周期的研究我们先试图划分为四个主要周期：复苏（景气度下行、信贷上行）、繁荣（景气度上行）、滞胀（景气度下行、通胀上行）与衰退（景气、通胀与信贷均下行），但由于这样划分出来的滞胀与衰退期较短（分别为 26 个月和 23 个月），不满足线性回归对数据样本的要求，因而将二者进行合并，最终得到复苏（景气度下行、信贷上行）、繁荣（景气度上行）和衰退（景气度下行、信贷下行）三种周期状态，三种周期状态分别对应时长合计分别为 69 个月、62 个月和 49 个月，对应的月份个数相对较为平均便于分析研究。

图表 3：三因子拐点时间节点列表（Bry-Boschan 方法）

	通胀		景气度		信贷	
	谷	峰	谷	峰	谷	峰
1	2007-07-31	2008-08-31	2006-03-31	2007-08-31	2007-05-31	2008-04-30
2	2009-07-31	2011-07-31	2009-05-31	2010-05-31	2009-01-31	2010-01-31
3	2012-09-30	2014-07-31	2013-04-30	2014-08-31	2012-01-31	2013-05-31
4	2015-09-30	2017-02-28	2016-07-31	2018-01-31	2014-10-31	2016-06-30
5	2020-05-31		2020-02-29	2021-05-31	2018-11-30	2020-11-30

资料来源：东证衍生品研究院

图表 4：不同经济周期下的利率表现总结整理（Bry-Boschan 方法）

	state1（繁荣）	state2（复苏前期）	state3（衰退&滞胀）
月份数目	69	62	49
周期划分所需指标	PMI=1	PMI=-1 Credit=1	PMI=-1 Credit=-1
国债收益率变化均值（BPs）			
国债收益率：1Y	6.42	-1.96	-6.11
国债收益率：2Y	6.55	-1.70	-6.43
国债收益率：3Y	6.20	-1.68	-6.21
国债收益率：5Y	5.30	-1.91	-4.85
国债收益率：10Y	5.57	-2.94	-4.48
国债收益率期限利差变化均值（BPs）			
国债收益率：10Y-2Y	-0.98	-1.24	1.95
国债收益率：10Y-1Y	-0.85	-0.97	1.63
债券净价指数波动率均值			
中债净价指数：1to3Y	3%	3%	3%
中债净价指数：3to5Y	6%	5%	7%
中债净价指数：5to7Y	9%	8%	10%
中债净价指数：7to10Y	12%	10%	13%
中债净价指数：10Y	18%	18%	20%

资料来源：东证衍生品研究院

图表 5：不同经济周期下宏观指标的有效预测占比（ER）

	ER_I（通胀）	ER_M（信贷）	ER_P（景气度）	ER（所有指标）
State1（繁荣）	5%	5%	19%	8%
State2（复苏）	33%	11%	50%	24%
State3（衰退&滞胀）	5%	0%	19%	4%
全样本	5%	2%	13%	4%

资料来源：东证衍生品研究院

此外，我们对不同经济周期下的宏观指标预测能力进行分析，具体做法分基于周期三种不同状态将解释与被解释变量划分为三个集合分别进行全样本的回归分析。结果表明，不同经济周期状态下，宏观指标的预测能力确实有较为显著的差异，基于有效预测占比（ER）指标的统计结果可以得到如下主要结论；复苏周期下宏观指标对利率的解释能力最佳，我们对复苏期的定义为景气度向下信贷向上，即中国特色的稳增长时期，该时期景气度细项指标具有显著解释能力的数目占比高达 50%，其背后的逻辑或由于稳增长时期确立了资金与信贷相对宽松的环境，而影响利率走势的边际变化则取决于稳增长是否起效（景气度与价格指标的跟踪）以及稳增长力度的变化（信贷指标的跟踪），使得利率市场迎来一个对经济数据的敏感期。

2.3. 加入周期哑变量的回归模型构建

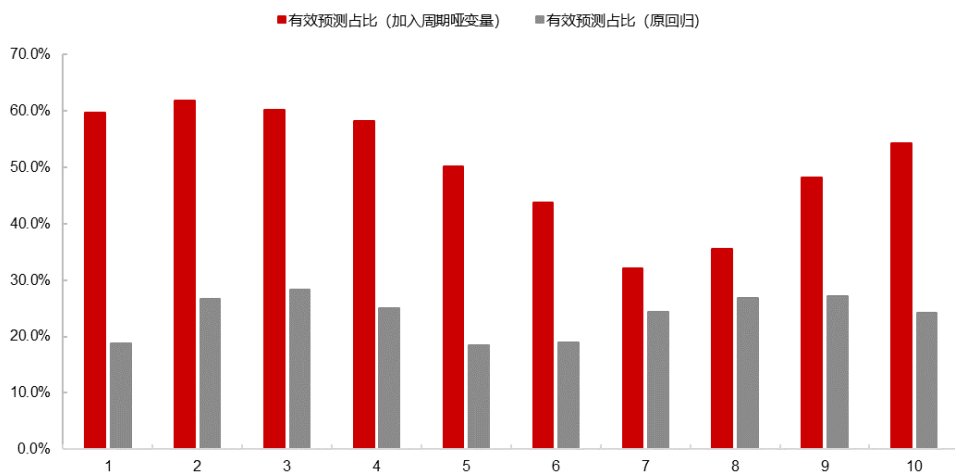
接下来我们考虑在单因子滚动回归框架下加入前瞻性判断的经济周期划分结果作为哑变量。由此，我们将较为粗糙的经济周期划分结果与更为精细的单因子预测回归相结合，借助周期的信息提升单因子的预测能力，借助单因子回归提升周期状态对利率波动影响的精细程度，进而在简单模型下实现一个较为可行的利率择时的单因子筛选方案。

具体回归方程方面，在选择三种周期状态的情况下，我们对原回归方程新引入两个哑变量：

$$(1) \quad y = \alpha_0 + \alpha_1 x + \beta_1 D_{state1} + \beta_2 D_{state2}$$

其中 x 为外生变量， D_{state1} 为是否处于 state1 的 0/1 哑变量， D_{state2} 为是否处于 state2 的 0/1 哑变量，常数项中包含是否处于 state3 的信息。

图表 6：不同期限利率预测单因子有效预测占比（近两年平均）



资料来源：东证衍生品研究院

3. 基于债券超额收益预测进行久期择时

3.1. 零息债券超额收益计算与归因

我们将债券超额回报描述为持有久期为 N 年的债券 H 期的回报率相对于 H 期即期利率的差值为该债券的超额回报，相对于简单地通过利率方向判断进行长短久期调整的策略，采用基于整条曲线的超额回报估计模型可以更加精准地进行组合构建。我们基于中债即期利率曲线构建“虚拟”零息债券，基于零息债券的框架进行债券超额收益的研究。

利率与零息债券对数价格之间满足如下关系：

$$(2) \quad \begin{aligned} y_t^{(n)} &= -\frac{1}{n} \ln p_t^{(n)} \\ f_t^{(n)} &= \ln p_t^{(n-1)} - \ln p_t^{(n)} \end{aligned}$$

则持有 n 年期零息债券一年的持有收益 $r_{t+1}^{(n)}$ 与超额收益 $rx_{t+1}^{(n)}$ 可以写为：

$$(3) \quad \begin{aligned} r_{t+1}^{(n)} &= \ln p_{t+1}^{(n-1)} - \ln p_t^{(n)} \\ rx_{t+1}^{(n)} &= r_{t+1}^{(n)} - y_t^{(1)} \end{aligned}$$

其中：

$\ln p_t^{(n)}$ ：为 t 时刻期限为 n 的零息债券的对数价格

$y_t^{(n)}$ ：为 t 时刻期限为 n 年期即期利率

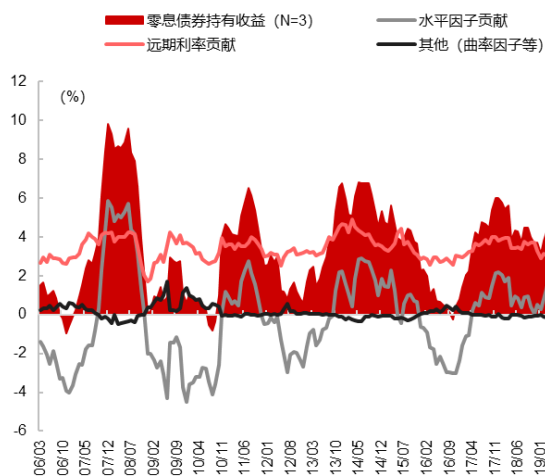
$f_t^{(n)}$ ：为 t 时刻 $n-1$ 到 n 的远期利率

直观理解零息债券持有一年的持有收益构成，其主要包含几个部分：1) 远期利率部分，即 t 时刻的远期利率，其决定了若曲线形态不发生变化的持有收益；2) 水平因子部分，即 t 时刻至 $t+1$ 时刻的即期利率变化，利率水平的波动是持有收益中较大的不确定因素；3) 曲率因子部分，即 t 时刻至 $t+1$ 时刻曲线的凸性变化，亦是未来曲线形态变化的不确定性影响，通常情况下影响较小，当市场波动较大或行情较为极端时影响可能放大。定量方面基于公式推导我们可以得到如下债券持有收益贡献的公式拆解：

$$(4) \quad \begin{aligned} r_{t+1}^{(n)} &\approx f_t^{(n)} - (1 + f_t^{(n)}) \times \frac{D_{N-1}}{1 + r_{N-1}} \times (y_{t+1}^{(n-1)} - y_t^{(n-1)}) \\ &\quad + \frac{1}{2} \times (1 + f_t^{(n)}) \times C_t^{N-1} \times (y_{t+1}^{(n-1)} - y_t^{(n-1)})^2 \end{aligned}$$

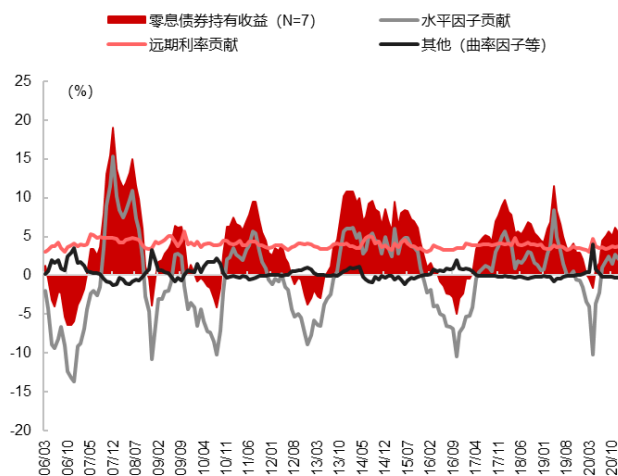
通过历史数据分析，对于远期利率、水平因子、曲率因子对不同期限零息债券的一年持有回报的实际贡献情况进行拆解与归因，下图展示了短期与中长期零息债券持有期为一年的实际持有收益的拆解结果。持有收益的主要影响因素中，远期利率的贡献相对最为稳定，利率水平因子的波动对持有收益结果影响更大，且随着债券期限增加其主导程度增加，即反应了久期效应，远期利率贡献在长债持有收益中重要性与波动贡献较低。

图表 7：国债零息债券（3Y）持有收益拆解



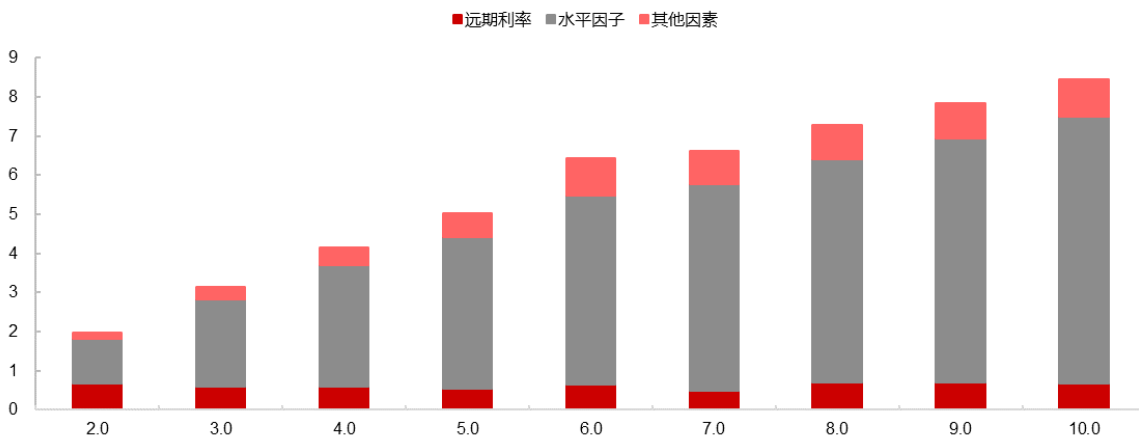
资料来源：东证衍生品研究院

图表 8：国债零息债券（7Y）持有收益拆解



资料来源：东证衍生品研究院

图表 9：不同期限国债零息债券远期利率、水平因子和其他因素波动贡献



资料来源：东证衍生品研究院

3.2. 基于远期利率与宏观数据预测债券超额收益

本文进行债券超额收益预测基于 Cochrane and Piazzesi (2005) 提出的远期利率曲线预测债券超额收益的框架，并尝试通过不同模型处理宏观数据作为除远期利率曲线之外的额外信息输入验证其对预测债券超额收益的增强效果。我们先对基于远期利率曲线的线性预测模型原理进行简单回顾，延续上文公式 (2) 至 (3) 的定义，将期限为 n 的债券超额收益线性预测方程写为：

$$(5) \quad rx_{t+1}^{(n)} = \beta_0^{(n)} + \beta_1^{(n)} y_t^{(1)} + \beta_2^{(n)} f_t^{(2)} + \cdots + \beta_{10}^{(n)} f_t^{(10)}$$

基于回归结果分析发现，所有期限债券的持有收益可基于同样的远期利率函数预测，期限较长的债券只是对函数有着更大的暴露，因而可将回归方程改写为：

$$(6) \quad rx_{t+1}^{(n)} = b_n \gamma^T F_t + \varepsilon_{t+1}^{(n)}$$

其中， F_t 为远期利率矩阵， γ^T 为影响不同期限债券超额收益的公共因子， b_n 为期限 n

对于公共因子的暴露， $n=2,3,\dots,10$ 。

基于海外文献以及我们此前对中债市场的实证研究，均证实远期利率曲线对未来债券超额收益具有显著预测能力。正如我们在文章第一部分中提到，海外文献在线性模型的基础上已经进行了很多拓展研究，其中引入关注的点主要在于宏观数据的加入对预测效果的提升，以及前沿的机器学习非线性模型在债券超额收益预测中的应用，而我们接下来也将基于国内市场进行进一步的实证探究。

具体模型说明与实证设定方面，我们将基于如下六种不同的建模方案进行债券超额收益进行预测。前三种方法仅基于远期利率曲线进行预测，后三种方法基于远期利率曲线加宏观数据进行预测。第一种方法，基于远期利率曲线线性回归预测超额收益；第二种方法，采用 PCA 降维法对远期利率曲线进行降维后再进行预测；第三种方法，基于机器学习 XGBoost 回归树模型使用远期利率曲线进行预测；第四种方法，基于远期利率曲线加宏观指标 PCA 降维方法使用线性预测回归模型进行预测；第五种方法，我们对宏观指标先进行滚动窗口单因子显著性测算，对显著的指标进行 PCA 降维后，再叠加远期利率曲线进行线性模型预测；第六种方法，我们基于机器学习 XGBoost 模型对远期利率曲线与宏观数据整个集合进行滚动窗口的训练与预测。通过关键期限 ($N=1,2,\dots,10$) 零息债券超额收益样本外预测准确程度 (R2) 评判上述方法的比较优劣。

图表 10：债券超额收益预测不同输入/模型介绍

模型代号	模型说明
process1	远期利率曲线线性模型预测
process2	远期利率曲线 PCA 降维预测
process3	远期利率曲线 XGBoost 模型预测
process4	远期利率曲线加宏观指标 PCA 降维特征基于线性模型预测
process5	远期利率曲线加经单因子筛选后的宏观指标 PCA 降维特征基于线性模型预测
process6	远期利率曲线加宏观数据基于 XGBoost 模型预测

资料来源：东证衍生品研究院

图表 11：债券超额收益预测模型基本设定

要点	详细说明
回测时间	2011.04.30 至 2021.12.24
解释变量	远期利率曲线与宏观与动量数据（2.1 节），远期利率曲线滞后一个月，其他数据均进行可得性处理。
被解释变量	中债国债零息债券月度频率持有期为一年的实际超额收益。
预测窗口设定	预测采取具有重叠窗口的方式，交易信号与模型滚动窗口均为月度，预测对象为持有一年的预期超额收益，类似“看长做短”。
机器学习模型	每一个滚动窗口（月度）重新训练模型，训练采取随机划分样本集与测试集的方式，基于训练结果进行取往后一期的预测值。基于部分历史数据样本优化超参数。
线性回归模型	债券超额收益预测的线性回归模型我们均使用 IVX 回归，宏观数据标准化处理后基本上是满足平稳性要求的，但远期利率曲线数据并不满足，使用传统 OLS 回归或带来较大估计误差。
PCA 降维模型	基于 PCA 对远期利率曲线降维与对宏观数据降维中，目标维度数据均选取三维。与机器学习模型设定类似，降维均基于滚动窗口操作，对于通过解释能力检验（基于回归 t 值检验）的降维特征采用线性回归的方式向后预测一期。

资料来源：东证衍生品研究院

3.2.1. 债券超额收益预测效果分析

基于不同期限零息债券超额收益预测值与实际值的样本外预测 R2 指标进行统计，六种预测方法的效果不一。首先对于仅使用远期利率曲线的前三种方法而言，process1 基于线性模型得到的 R2 结果相当于模型比较基准；process2 基于 PCA 降维对远期利率曲线进行特征提取并预测的方法降低了解释能力，预测结果不理想；process3 采用机器学习 XGBoost 回归树模型基于远期利率曲线对超额收益的预测能力进行非线性框架下的训练与预测，实现三种方法里相对非常突出的预测表现。其次，对于加入宏观数据的后三种模型，process4 基于 PCA 对大量数据进行直接降维其得到特征因子对超额收

益预测的改善不佳；process5 对于宏观数据的处理方面，先基于单因子预测回归筛选显著指标再使用 PCA 降维,相对直接大量数据 PCA 的方法预测样本外 R2 有显著提升，且相对于基准线性模型（process1）预测效果有一定增强；process6 基于机器学习 XGBoost 模型对远期利率曲线与宏观数据集进行训练,实现的样本外预测效果远优于其他宏观数据处理方法，且好于单纯基于远期利率曲线的 XGBoost 预测效果。

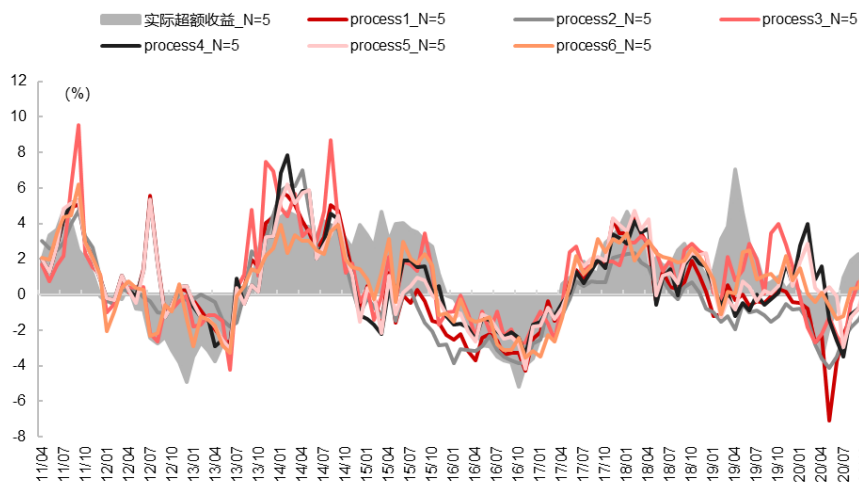
对于不同模型方法的评价进行总结，首先，机器学习 XGBoost 回归树模型具有非常显著的模型优势；其次，对于线性预测框架而言，进行宏观数据的单因子滚动窗口筛选具有一定的提升效果；最后，对于不同期限零息债券超额收益预测效果对比方面，传统模型对中长期零息债券预测效果相比较弱，而机器学习模型结果对于不同期限预测效果差异不大。

图表 12：不同模型下国债零息债券各久期超额收益预测样本外 R2

	process1	process2	process3	process4	process5	process6
BRP (n=2)	0.22	0.25	0.40	0.15	0.18	0.61
BRP (n=3)	0.30	0.24	0.45	0.24	0.30	0.60
BRP (n=4)	0.30	0.24	0.45	0.24	0.27	0.64
BRP (n=5)	0.23	0.16	0.40	0.19	0.25	0.61
BRP (n=6)	0.21	0.24	0.43	0.18	0.26	0.59
BRP (n=7)	0.17	0.14	0.42	0.18	0.25	0.62
BRP (n=8)	0.25	0.28	0.48	0.23	0.24	0.61
BRP (n=9)	0.28	0.29	0.50	0.29	0.30	0.61
BRP (n=10)	0.28	0.27	0.47	0.31	0.29	0.61

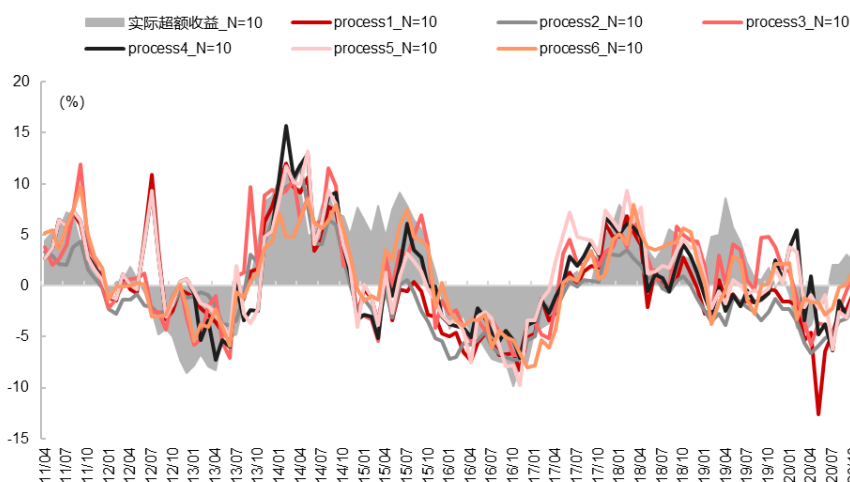
资料来源：东证衍生品研究院

图表 13：不同模型中债零息债券超额收益预测（N=5）



资料来源：东证衍生品研究院

图表 14：不同模型中债零息债券超额收益预测 (N=10)

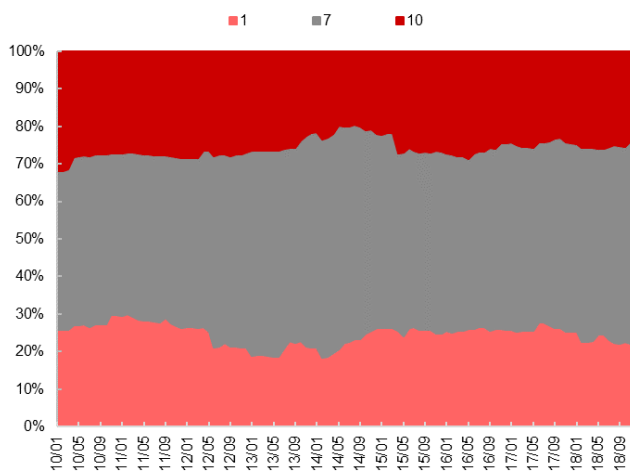


资料来源：东证衍生品研究院

3.2.2. 中债财富指数的复制和预测

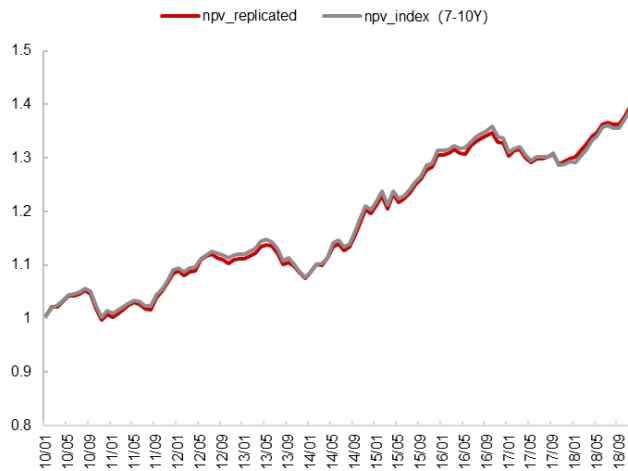
前文中我们均基于中债利率曲线反算得到的“虚拟”零息债券进行债券超额收益预测的研究，中债曲线亦是基于市场交易数据拟合得到的，因而从我们对关键期限零息债券超额收益的预测信息转为对于债券投资的久期择时问题并不复杂，我们通过对中债国债、国开债财富指数进行动态复制的方式，得到基于三个零息债券多空组合的特定指数的复制净值。例如，下图展示了我们基于期限为【1,7,10】的三个零息债券组合得到中债国债 7-10 年财富指数的复制指数，其净值走势与原指数非常接近。进而，我们基于复制指数的零息债券动态权重与零息债券持有收益的预测值得到特定中债财富指数的持有收益预测值，选择预期持有收益较高的指数作为债券久期择时策略的标的。

图表 15：中债国债 7-10 年财富指数复制权重



资料来源：东证衍生品研究院

图表 16：中债国债 7-10 年财富指数复制净值



资料来源：东证衍生品研究院

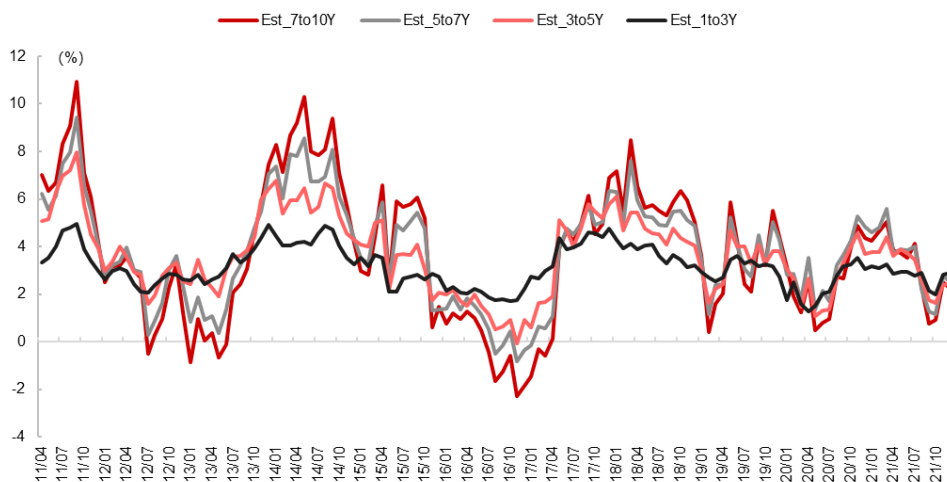
3.2.3. 基于持有收益预测的久期轮动策略

我们基于中债国债财富指数四个主要指数构建债券久期轮动策略（中债国债 1-3 年、中债国债 3-5 年、中债国债 5-7 年与中债国债 7-10 年）。通过指数动态复制的方法得到基于三个关键期限零息债券对各个指数的动态复制权重，进而从零息债券持有收益预测值得到指数预期持有收益，在四个期限中选择持有收益最高的期限作为债券久期择时的配置期限。通过模型久期的选择与利率走势的比较，可以发现模型在利率周期的顶部与底部的信号结果最稳正确与稳定，对于利率在中枢位置震荡时期可能出现交易信号在中等久期附近震荡的现象。

我们测算六种模型下中债国债指数与国开债指数久期择时净值效果，在久期择时策略中决定策略效果更重要的是模型对不同久期预期持有收益的排序是否准确，与前文比较模型优劣采用的超额收益预测值的误差在评判逻辑上有一定程度的差异，结果有细节出入但整体上也较为一致。从净值结果表现来看，效果较好具有显著超额收益的模型是 process3、process4、process5、process6，前两个模型超额收益不明显，后四个模型中，年化收益方面表现较为接近，加入宏观数据的模型实现较为明显的最大回撤下降，process6 整体表现最为突出。

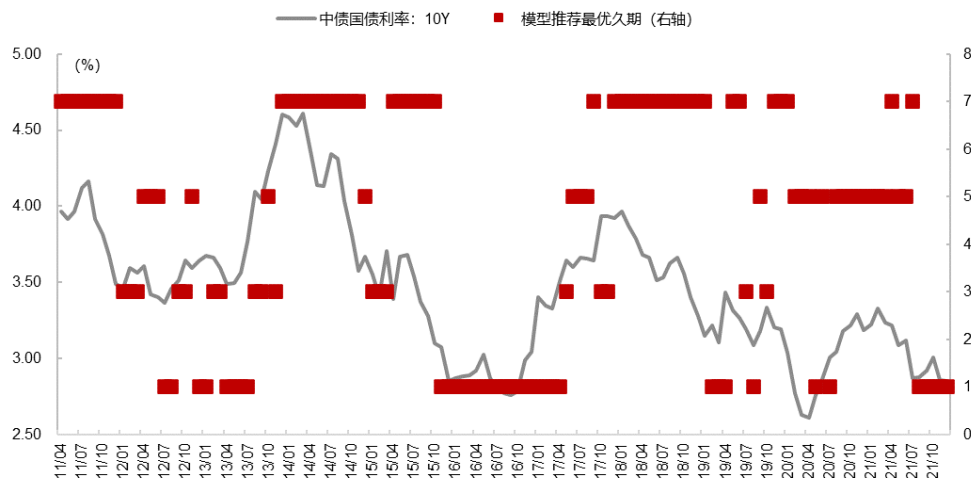
策略表现总结方面，机器学习模型相对于线性模型具有显著优势，加入宏观数据的择时效果更优，机器学习模型实现经济数据对利率非线性影响关系计量与预测方面更优的效果。基于 XGBoost 模型对中债国债四个关键期限财富指数的久期择时策略实现年化收益 5.05%，最大回撤-2.58%，年化波动 2.56%，夏普比率 1.97，平均持仓周期 3.4 个月，回测区间为 2011.04 至 2021.12；对相应中债国开债指数的久期择时策略实现年化收益 6.6%，最大回撤-2.2%，年化波动 3.2%，夏普比率 2.08，平均持仓周期 3.2 个月，回测区间为 2013.02 至 2021.12。

图表 17：中债国债财富指数一年期持有收益预测值（process6）



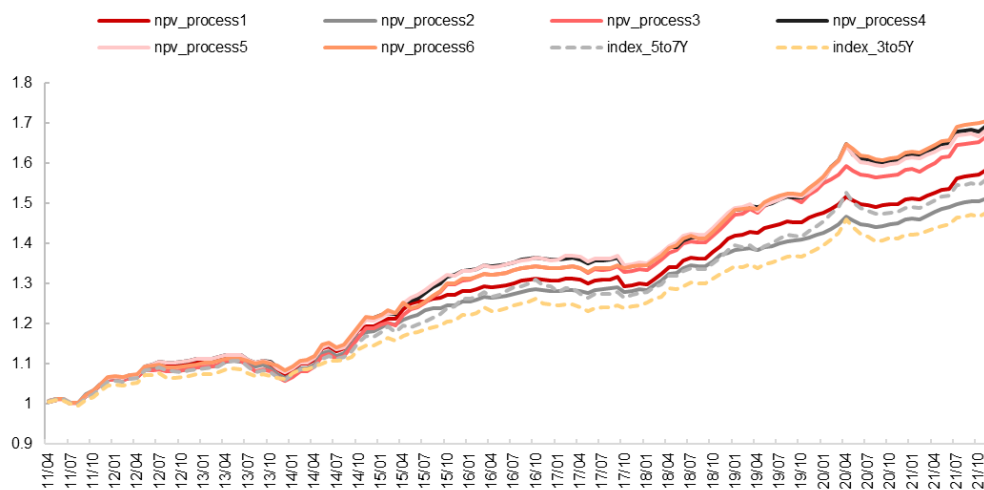
资料来源：东证衍生品研究院

图表 18: 中债国债久期择时的配置期限与利率走势 (process6)



资料来源: 东证衍生品研究院; 注: 图表中右轴配置久期是对债券指数大致匹配的代号, 并不必然对应真实久期建议, 其对应方式为选择 1-3 年指数配置对应 1, 选择 3-5 年指数配置对应 3, 选择 5-7 年指数配置对应 5, 选择 7-10 年指数配置对应 7。

图表 19: 中债国债财富指数久期轮动策略净值



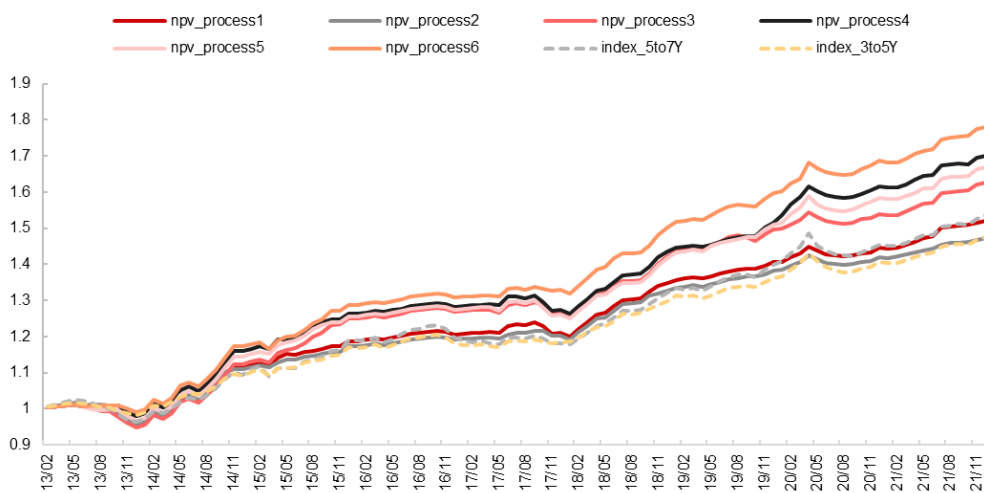
资料来源: 东证衍生品研究院

图表 20: 中债国债财富指数久期轮动策略净值分析

	净值结果 process1	净值结果 process2	净值结果 process3	净值结果 process4	净值结果 process5	净值结果 process6	基准净值 (3-5Y)	基准净值 (5-7Y)
累计收益率	58.00%	50.44%	65.99%	68.96%	67.93%	69.77%	47.02%	54.99%
年化收益率	4.35%	3.87%	4.83%	5.00%	4.94%	5.05%	3.68%	4.19%
年化波动率	2.34%	2.13%	2.53%	2.62%	2.68%	2.56%	2.15%	2.69%
最大回撤率	-3.78%	-3.86%	-4.57%	-3.78%	-3.94%	-2.58%	-3.65%	-4.39%
胜率(M)	76%	77%	77%	76%	74%	73%	72%	69%
盈亏比	1.44	1.26	1.29	1.50	1.47	1.78	1.31	1.38
夏普比率	1.86	1.82	1.90	1.91	1.84	1.97	1.71	1.56
Calmar 比	1.15	1.00	1.06	1.32	1.25	1.96	1.01	0.96
平均持仓(M)	2.9	5.2	2.6	3.2	3.4	3.4	--	--

资料来源: 东证衍生品研究院

图表 21: 中债国债财富指数久期轮动策略净值



资料来源: 东证衍生品研究院

图表 22：中债国开债财富指数久期轮动策略净值分析

	净值结果 process1	净值结果 process2	净值结果 process3	净值结果 process4	净值结果 process5	净值结果 process6	基准净值 (3-5Y)	基准净值 (5-7Y)
累计收益率	51%	46%	62%	70%	66%	77%	47%	53%
年化收益率	4.8%	4.4%	5.6%	6.1%	5.9%	6.6%	4.4%	4.9%
年化波动率	3.0%	2.8%	3.4%	3.3%	3.3%	3.2%	2.6%	3.5%
最大回撤率	-5.5%	-5.5%	-6.3%	-4.0%	-4.5%	-2.2%	-3.7%	-5.8%
胜率(M)	75%	75%	72%	75%	75%	75%	71%	66%
盈亏比	1.24	1.27	1.42	1.49	1.37	1.90	1.38	1.41
夏普比率	1.58	1.55	1.65	1.86	1.75	2.08	1.68	1.40
Calmar 比	0.87	0.79	0.88	1.54	1.31	2.99	1.18	0.83
平均持仓(M)	4.1	7.1	2.6	4.1	3.6	3.2	--	--

资料来源：东证衍生品研究院

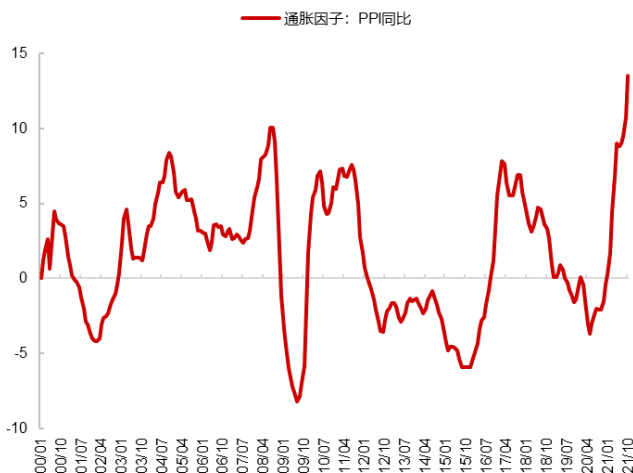
4. 附件

4.1. 经济指标处理口径调整

我们基于通胀、景气度与信贷三因子确认周期位置，在此补充介绍三因子的计算方法。

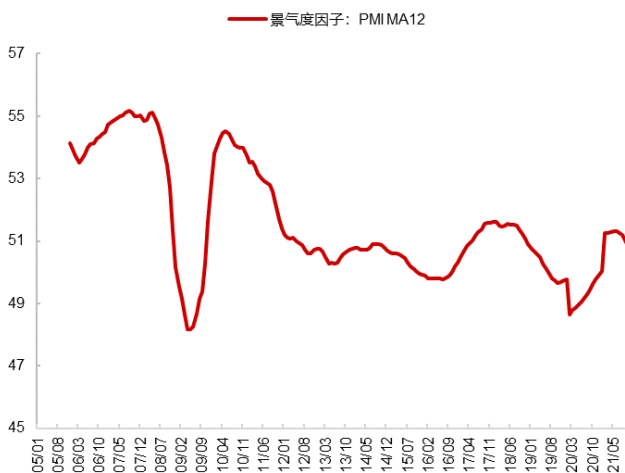
我们对划分经济周期的主要参照指标进行数据处理，指标选择为 PPI、PMI 与信贷增速，为去除一定的季节性并尽量统一指标形式，采用移动平均的平滑处理。PPI 本身为同比序列无需额外调整。PMI 为环比数据，我们将其进行 12 个月移动平均处理得到同比数据。

图表 23：经济周期因子之通胀走势



资料来源：Wind，东证衍生品研究院

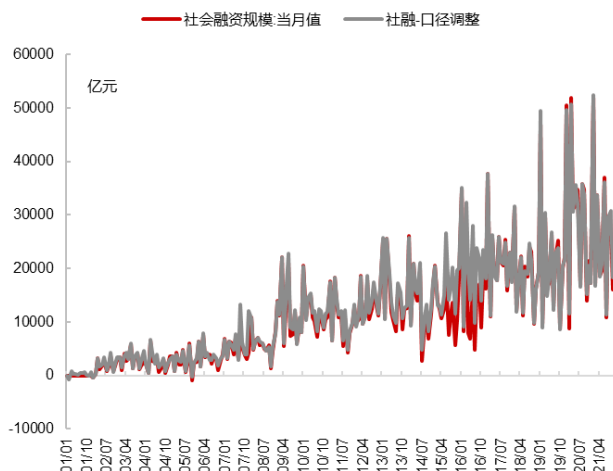
图表 24：经济周期因子之景气度走势



资料来源：Wind，东证衍生品研究院

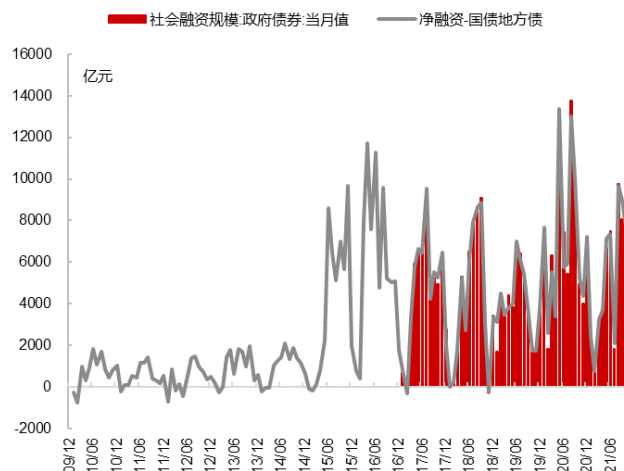
对于信贷指标，我们基于社融规模指标基于国债地方债净融资月度数据进行口径调整之后的结果，因社融数据未包含 2017 年之前的政府债券，我们基于国债地方债发行数据进行月度汇总估算 2017 年之前的政府债券净融资额，进而得到调整后的社融增量，进一步基于社融规模同比多增的 12 个月移动平均计算信贷脉冲。

图表 25: 口径调整后的社会融资规模当月值



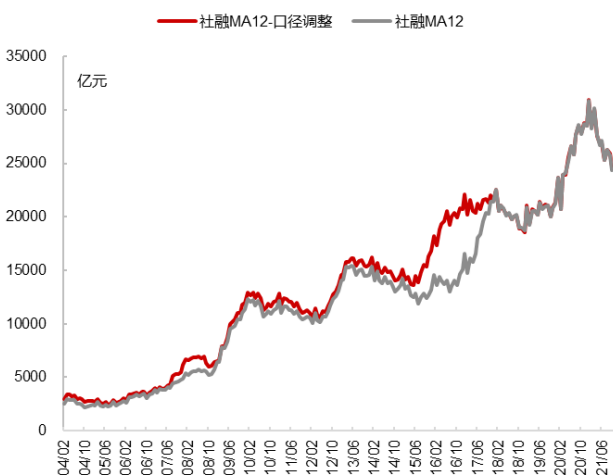
资料来源: Wind, 东证衍生品研究院

图表 26: 国债地方债净融资与社融政府债券分项



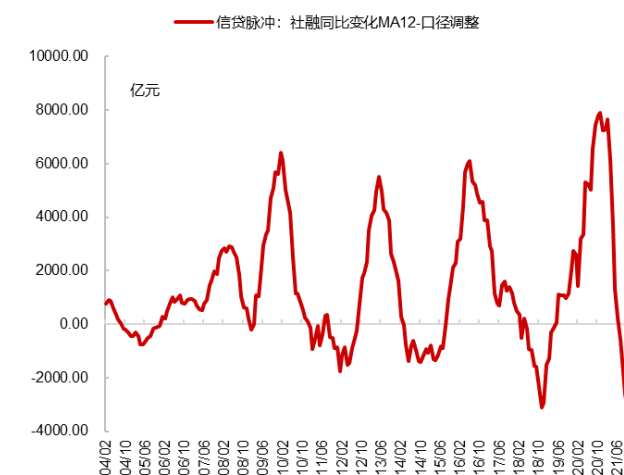
资料来源: Wind, 东证衍生品研究院

图表 27: 社会融资规模口径调整结果 (MA12)



资料来源: Wind, 东证衍生品研究院

图表 28: 经济周期因子之信贷脉冲走势



资料来源: Wind, 东证衍生品研究院

4.2. 经济周期划分方法

4.2.1. 回顾分析：Bry-Boschan 算法

Bry-Boschan (1971) 算法处理的对象是平滑后的数据序列，其本质其实是一种极值筛选法，它将 NBER 经济周期决策委员会对拐点的判断方式，归纳为一系列简单决策原则，用于层层筛选过滤，最后剩下的极值被认定为拐点，其判断原值包括：

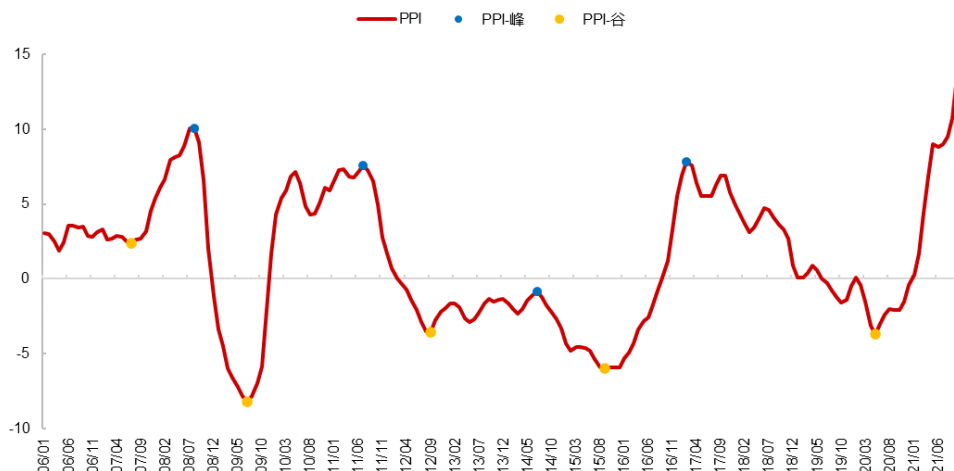
- i. 波峰、波谷必须交替出现；
- ii. 相邻波峰、波谷之间须至少间隔 6 个月；
- iii. 一个周期(峰与峰、谷与谷之间)需至少持续 15 个月；
- iv. 序列头尾 6 个月内识别出拐点无效，若头尾 24 个月内识别出局部峰（谷）且该值不是 24 月内最大（小）值，则该峰（谷）同样无效。

算法实现主要包含以下几个步骤：

- 第一步：寻找所有极值点(前后 N 期范围内最大/小值)；
- 第二步：检查相邻极值点之间该极值是否是最大/小值，若不是，则删除该极值点；
- 第三步：检查极值点是否交替出现，若有重复峰/谷，只保留时间最早或较大/小点；
- 第四步：检查周期长度是否满足最短时限要求，若不是，删除较小/大的极大/小值；
- 第五步：重复二、三步；
- 第六步：检查半周期是否满足最短长度要求，若不是，只保留偏离均值较大的极值；
- 第七步：重复二、三步，剩下的极值点认定为拐点。

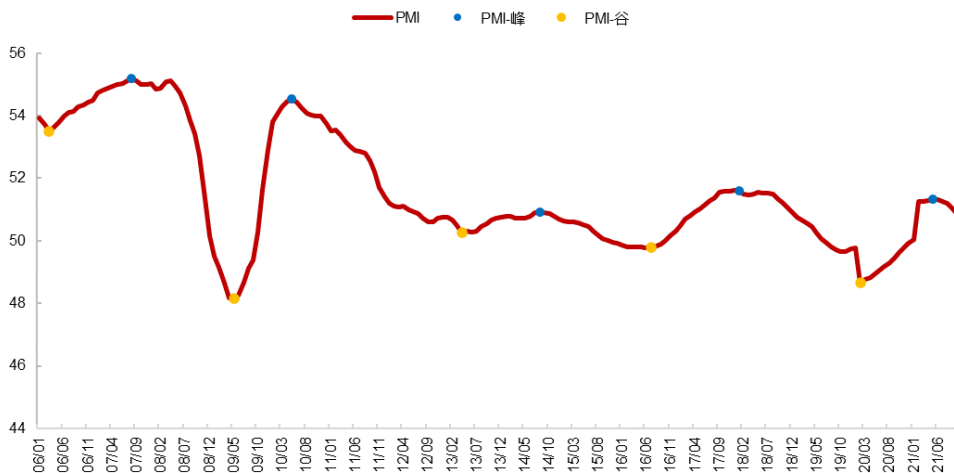
我们基于 BB 算法对通胀、景气度和信贷三因子进行周期划分的结果如下图展示。

图表 29：通胀因子走势与周期拐点确认结果（基于 Bry-Boschan 方法）



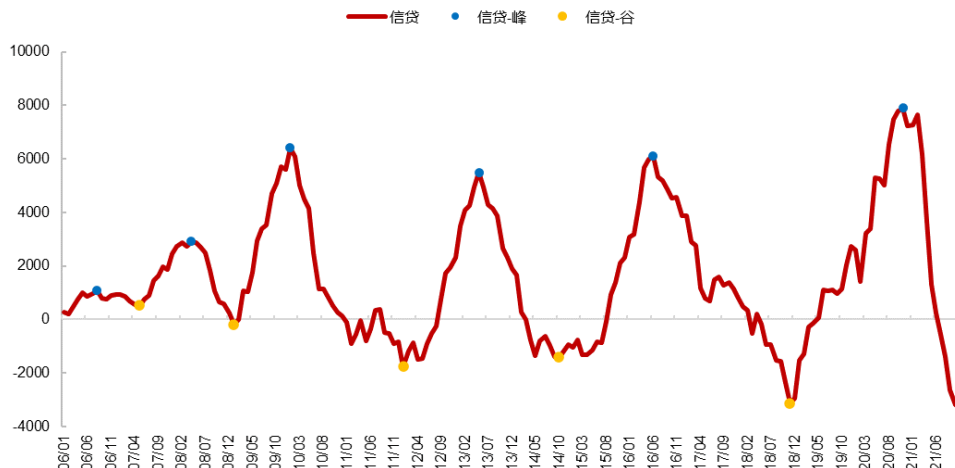
资料来源：Wind，东证衍生品研究院

图表 30：景气度因子走势与周期拐点确认结果（基于 Bry-Boschan 方法）



资料来源：Wind，东证衍生品研究院

图表 31: 信贷因子走势与周期拐点确认结果 (基于 Bry-Boschan 方法)

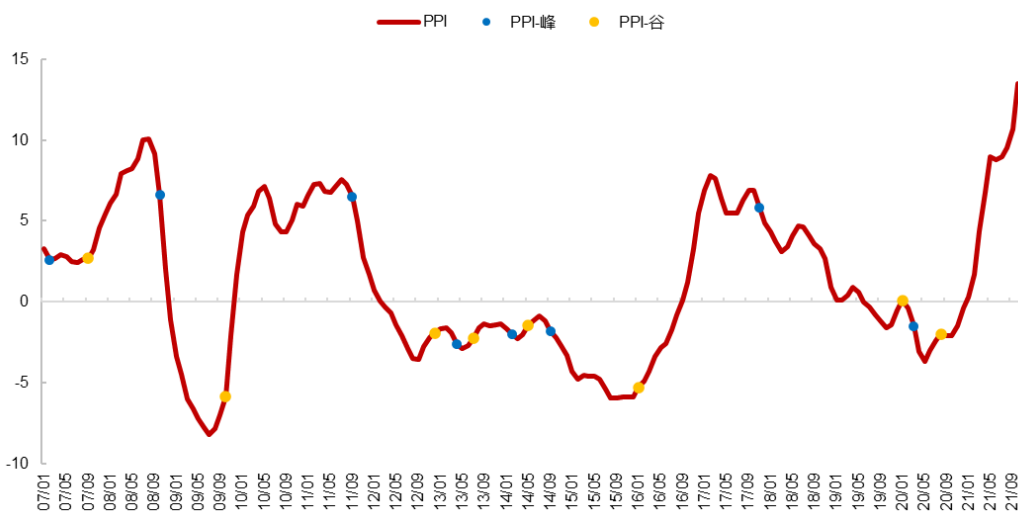


资料来源: Wind, 东证衍生品研究院

4.2.2. 前瞻判断: 趋势+绝对值突破

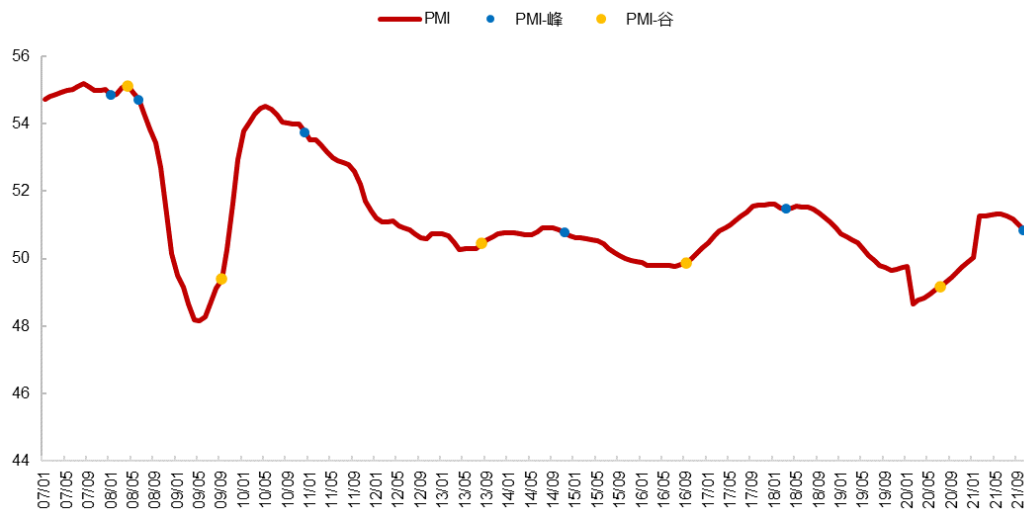
基于趋势平滑 (LLT) 信号加绝对水平突破的方式进行前瞻式周期状态判断用于策略中。趋势平滑主要基于高阶滤波器实现更为灵敏的趋势方向判断, 加入数据绝对水平阈值的判断是为了从经济数据绝对值上进行进一步的验证, 我们的做法是最新值突破前三期最大/最小值的方式确认趋势变化, 进而尽量避免无效转折信号。此种基于前瞻性判断的方式也是我们在策略回测中应用的周期状态因子结果, 相对于 BB 算法无法避免在数据波动较大时期存在一定的反复情况, 但整体上也能够较好地描述周期趋势。

图表 32: 通胀因子走势与周期拐点确认结果 (基于势平滑加辅助判断条件)



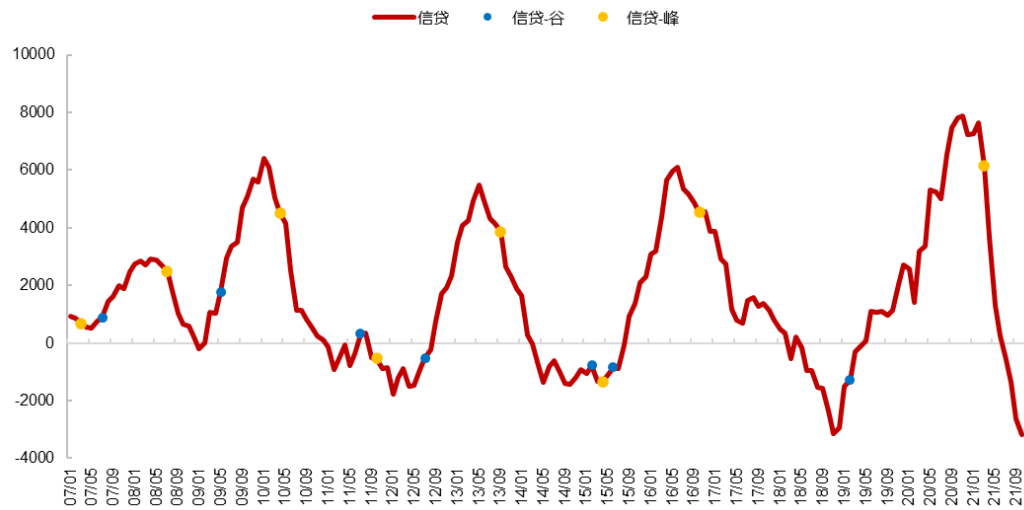
资料来源: Wind, 东证衍生品研究院

图表 33: 景气度因子走势与周期拐点确认结果 (基于势平滑加辅助判断条件)



资料来源: Wind, 东证衍生品研究院

图表 34: 信贷因子走势与周期拐点确认结果 (基于势平滑加辅助判断条件)



资料来源: Wind, 东证衍生品研究院

5. 风险提示

量化模型失效风险，指标的有效性基于历史数据得出，不排除失效的可能。

期货走势评级体系（以收盘价的变动幅度为判断标准）

走势评级	短期（1-3 个月）	中期（3-6 个月）	长期（6-12 个月）
强烈看涨	上涨 15%以上	上涨 15%以上	上涨 15%以上
看涨	上涨 5-15%	上涨 5-15%	上涨 5-15%
震荡	振幅-5%-+5%	振幅-5%-+5%	振幅-5%-+5%
看跌	下跌 5-15%	下跌 5-15%	下跌 5-15%
强烈看跌	下跌 15%以上	下跌 15%以上	下跌 15%以上

上海东证期货有限公司

上海东证期货有限公司成立于 2008 年，是一家经中国证券监督管理委员会批准的经营期货业务的综合性公司。东证期货是东方证券股份有限公司全资子公司，注册资本金 23 亿元人民币，员工近 600 人。公司主要从事商品期货经纪、金融期货经纪、期货投资咨询、资产管理、基金销售等业务，拥有上海期货交易所、大连商品交易所、郑州商品交易所和上海国际能源交易中心会员资格，是中国金融期货交易所全面结算会员。公司拥有东证润和资本管理有限公司，上海东祺投资管理有限公司和东证期货国际（新加坡）私人有限公司三家全资子公司。

东证期货以上海为总部所在地，在大连、长沙、北京、上海、郑州、太原、常州、广州、青岛、宁波、深圳、杭州、西安、厦门、成都、东营、天津、哈尔滨、南宁、重庆、苏州、南通、泉州、汕头、沈阳、无锡、济南等地共设有 33 家营业部，并在北京、上海、广州、深圳多个经济发达地区拥有 134 个证券 IB 分支网点，未来东证期货将形成立足上海、辐射全国的经营网络。

自 2008 年成立以来，东证期货秉承稳健经营、创新发展的宗旨，坚持市场化、国际化、集团化的发展道路，打造以衍生品风险管理为核心，具有研究和技术两大核心竞争力，为客户提供综合财富管理平台的一流衍生品服务商。

分析师承诺

王冬黎

本人具有中国期货业协会授予的期货执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰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接收到任何形式的报酬。

免责声明

本报告由上海东证期货有限公司（以下简称“本公司”）制作及发布。

本研究报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本研究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报告之外，绝大多数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投资标的的邀请或向人作出邀请。

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，投资者需自行承担风险。

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。

如需引用、刊发或转载本报告，需注明出处为东证衍生品研究院，且不得对本报告进行任何有悖原意的引用、删节和修改。

东证衍生品研究院

地址：上海市中山南路 318 号东方国际金融广场 2 号楼 22 楼

联系人：梁爽

电话：8621-63325888-1592

传真：8621-33315862

网址：www.orientfutures.com

Email：research@orientfutures.com