

【建投专题】*Nelson-Siegel* 理论在我国国债期货市场跨品种套利中的应用

何熙 CFC 宏观金工 2020-09-02

作者姓名：何熙
期货投资咨询从业证书号：Z0014559
电话：023-81157284
研究助理：王锴
期货从业证书号：F3075456

摘要

以 5 年和 10 年期国债期货合约为目的的跨品种套利是市场上较为常见的操作方式，传统理论认为不同债券期限利差在偏离到一定程度时会出现均值回归。但是实际当中经常出现利差突破轨道后继续上涨(下跌)并没有回归的情况。本文在这里讨论并测试基本面因子嵌套模型结合趋势回归模型的使用效果，为了避免多因子多重共线性问题，因此本文在研究利差影响因素时，在 *Nelson-Siegel* 模型的基础上综合使用主成分分析法和因子旋转-最大方差法，选取 12 个日频宏观因子，通过主成分分析提取出 3 个主成分，3 个主成分分别对应市场预期、风险溢价和凸性偏离影响，可以解释样本中 78%、6%和 1%的方差，然后进行因子旋转和 *Logistic* 回归，同时利用趋势回归模型做信号过滤对原模型进行改进。依据改进后的模型可以得到跨品种套利的开平仓信号。作为对比本文分别对只进行 *Logistic* 回归、主成分分析+*logistic* 回归、主成分分析+因子旋转+*logistic* 回归和主成分分析+因子旋转+*logistic* 回归+趋势回

归四种方法做了回测。实证结果显示，基于“主成分分析+因子旋转+*logistic* 回归”的嵌套基本面因子模型经过趋势回归信号过滤之后，年化收益获得较大提升，且最大回撤显著缩小。由此得到的国债期货 10-5Y 跨品种套利策略具有一定的实践指导意义。

关键词：国债期跨品种套利；*Logistic* 回归；趋势回归信号过滤

01

引言

以 5 年和 10 年期国债期货合约为目的的跨品种套利是市场上较为常见的操作方式。传统理论认为，5 年期和 10 年期国债收益率具有一致性，存在着较为稳定的利差中枢。根据 *Fama* 提出流动性偏好假说，投资者偏好投资短期债券，长期债券由于市场波动风险和变现的风险更大，持有长期债券就需要一定的收益率补偿，因此正常情况下收益率曲线会向上倾斜，收益率曲线的斜率也会存在一定的规律性。当利差明显高于（低于）利差中枢时，收益率曲线过于陡峭（平缓），此时 10 年期国债价值相对 5 年期国债被低估（高估）。预期过于陡峭（平缓）收益率曲线会出现自我修复，例如收益率曲线过于陡峭时，投资者倾向于买长期债券或者卖出短期债券，反之收益率曲线过于平缓时，投资者倾向于卖出长期债券，买入短期债券，从而带动利差中枢修复，收益率曲线斜率回归常态。在某些极端情况下，甚至有可能出现收益率曲线倒挂的情景，而此时收益率曲线回归正常形态的动能也就更强，确定性也更高。以上提

到的这些，也成为国债期货跨品种套利的理论基础，近几年随着 5 年期和 10 年期国债期货的相继上市，债券市场做空机制更加完善，可以选择的套利方式也更加多元化。

以上提到的只是理想情况，在实际操作中，经常会面临利差突破轨道后继续上涨(下跌)并没有回归的情况。其原因在于，一方面利差中枢需要一段时间跨度的历史数据来计算，时间跨度的选取具有经验性和不确定性，虽然债券收益率的变动确实存在一定的规律性，但影响债券市场的因素较为复杂，随着时间的推移，历史经验不会永远有效，进一步扩大了时间跨度选择的难度。另一方面，均值回归理论基于市场有效性和投资者风险偏好假设，但是现实的情况是市场强有效性是一种理论上的状况，实际的债券市场中经常出现对信息反映过度的情况，尤其是市场波动剧烈时，追涨杀跌的动能导致在很多时候非理性才是常态，同时利差影响因素包括宏观基本面信息，传统理论并不总是适用。本文在这里讨论并测试基本面因子嵌套模型结合趋势回归模型的使用效果，需要指出的是在传统的基本面因子研究中常用的线性回归模型很难规避多重共线性的问题，因此本文在 *Nelson-Siegel* 模型的基础上综合使用主成分分析法和因子旋转-最大方差法，然后进行 *Logistic* 回归，同时利用趋势回归模型做信号过滤对原模型进行改进。

02

本次研究对象与研究方法

本文以中金所 10 年期国债和 5 年期国债期货主力合约为主要研究对象，回测区间为 2017 年 1 月 4 日至 2020 年 8 月 4 日，在主成分分析时，根据经济功能、估值、流动性、资金和外部因素五个方面，选取 12 个日频的因子，最终对提取出的三个降维后的主成分做 Logistic 回归，3 个主成分分别对应市场预期、风险溢价和凸性偏离。利用趋势回归模型做信号过滤的时候，主要观测的是 2017 年 1 月 4 日至 2020 年 8 月 4 日期间 5 年期和 10 年期国债现券收益率的利差，我们采用 2.15 倍的移动 20 日标准差作为通道宽度，即观测利差超过通道宽度时作为开仓信号。

观测区间内债券市场经历了熊市转牛市又重新变成熊市的过程，以 10 年期国债收益率为例：2017 年 1 月 4 日-2017 年 11 月 23 日受到资管新规、宏观去杠杆等因素的影响，收益率由 3.15% 上行至 3.99%，上行幅度 84 个 BP；之后债券市场在经历了资管新规延期执行、宏观流动性放松、新冠疫情等影响开启了 1 年半的债券牛市，截止 2020 年 4 月 29 收益率下行至 2.49%，下行幅度达到 150 个 BP；随后随着疫情冲击的减弱，宏观经济企稳，货币政策不再边际宽松，5 月份以来债券市场收益率迅速反弹，截至 2020 年 8 月 4 日反弹至 2.94%，短短两个多月时间，收益率上行 45 个 BP。总体而言，虽然我们选取的回测区间并不长，但是期间经历了两次的牛熊转换，还是能够对策略模型做比较好的验证。

（一）Nelson-Siegel 模型

上世纪 80 年代 Nelson 和 Siegel 提出了一个用参数表示的瞬时远期利率的函数，该模型将利率期限结构分解为水平、斜率和曲率三个部分，分别可以解释实际中的远期利率市场预期、债券的风险溢价以及凸性偏离。相比传统的多项式和指数样条法，该模型对于短期收益率高于长期收益率的反向形态，以及国家宏观经济政策和微观市场环境的原因使长期收益率高于短期收益率的正向趋势具有可解释意义，符合市场理论预期，上世纪末欧洲银行率先使用后认可了该模型的可解释性。此外该模型只使用三个 β 参数，具有更高的灵活度来拟合不同类型的收益率曲线：

$$R(t) = \beta_0 + \beta_1 \frac{1-e^{-t/\tau}}{t/\tau} + \beta_2 \left(\frac{1-e^{-t/\tau}}{t/\tau} - e^{-t/\tau} \right) \quad (1)$$

其中 β_0 的系数为 1，是一个常数，可以理解为一个水平因子。 β_1 从 1 变为 0，代表曲线的斜率， β_2 先增后减，代表曲度。如何将市场基本面数据库中庞杂的特征变量转化为上述三因子模型并运用于实际策略是当前主要问题。

本文从以下三个方面对现有方法进行改进：第一，单独采用主成分分析对利差曲线拟合效果较差，仅能描述大致变化趋势，不能满足实际套利需要的预测精度。我们对因变量利差进行区间处理，使用 Logistic 回归模型进行拟合，加强了拟合程度和样本外泛化能力；第二，为了使整个模型更具有可解释性，运用最大化方差法旋转因子，使 Nelson-Siegel 模型内三个预测变量的整体贡献方差达到最大；第三，将基本面因子模型信号叠加趋势回归信号进行过滤，信号共振时开仓进场，信号背离时平仓离场。

（二）主成分分析

我们从基本面逻辑出发，结合 *Salomon Brothers(1983)* 在 *Overview of Forward Rate Analysis* 中对于利率曲线提及的经济因素，根据经济功能、估值、流动性、资金和外部因素五个方面，通过筛选利差波动的相关性，提取 12 个日频的因子，经过归一化和去极值后加入到模型中。

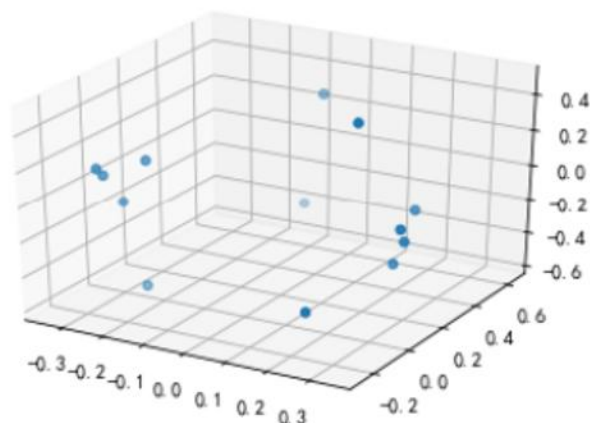
在此也对相关性比较高的几个因子做简要的介绍。与价格相关的日频因子，主要包括菜篮子产品批发价格 200 指数、HRB400 $\Phi 16-25\text{mm}$ 螺纹钢市场价、全国猪肉平均价、炼焦煤中信行业指数，这类因子对债券收益率和利差的影响有这样的路径，农产品和工业品价格的上涨意味着通胀的抬升，如果名义利率不变，实际利率是在下降的，如此一来投资者就会要求更高的名义利率，工业品价格的上涨，通常伴随着需求的上涨，经济基本面的改善，实体企业能够接受更高的融资利率，通胀的抬升和经济的走强还可能引发货币政策的收紧。以上提到的这几点均有可能造成债券收益率的上行和长短期利差的走阔。

与流动性相关的因子，主要包括 DR007、DR001、R007、R001、GC001 利率，流动性的宽松通常是由货币政策的宽松引起的，货币政策宽松又对债券收益率的下行有引导作用，金融市场利率的走低，意味着债券市场套息空间的加大，在套利投资者的参与下，债券的需求增加，也就意味着债券收益率的下行，实践中我们很容易发现，债券收益率的下行通常也会伴随着债券长短期利差的收窄。

与其他债券的利差因子，具体包括十年期国开债与国债的到期收益率利差、十年期铁道债与国债的到期收益率利差和十年期 **AAA** 企业债与国债的到期收益率利差。这些因子也不难理解，当这些利差出现明显变化时，也就意味着不同券种、不同期限债券的配置性价比出现变化，债券投资者就会调整投资行为，也会引国债利差的变化。

为了避免多重共线性，实现数据降维的同时最大程度减少原始数据信息的丢失，我们首先引入主成分分析法，基本思路是：从 p 个相关的解释变量中提出 k 个不相关的主成分，每一个主成分都是原始变量的线性拟合，第一个主成分最大程度地解释了原始变量数据的方差，具有最大的特征值。第二主成分与第一主成分之间不存在线性关系，它最大程度解释了剩余方差，以此类推。因此，参照（1）中的三个解释变量，我们从原始数据中提取出前三个主成分，分别可以解释样本中 **78%**、**6%** 和 **1%** 的方差，其解释占比的大小也接近经验性上对于市场预期、风险溢价和凸性偏离影响大小的判断。根据此前 *Salomon Brothers* 的相关研究，三个主成分的解释程度约为 **70%**、**20%** 和 **10%**，虽然受到选取因子、金融市场差异等影响存在一定的区别，但是大体上还是有一定的借鉴意义。

图 1：主成分载荷曲面

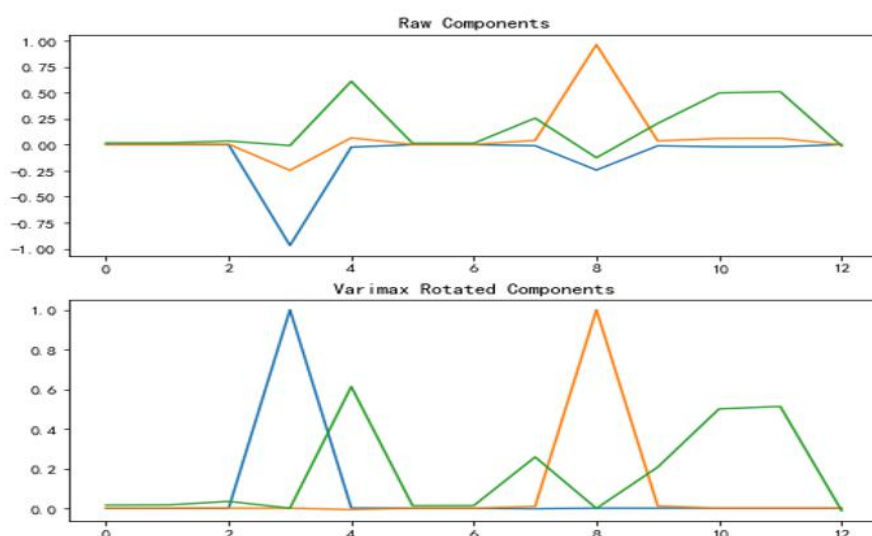


数据来源：Wind，中信建投期货

（三）因子旋转-最大方差法

通过主成分分析提取的三个主要成分均由原始数据的因子线性组合而成，这样的向量称之为载荷矩阵。多数情况下，估计出的因子载荷在各个因子上并不突出，在一定程度上影响了模型解释度。由于因子载荷矩阵并不唯一，可以通过旋转的方式突出因子的特征，使各个因子载荷之间的距离尽量拉大，即最大方差化。理论上，因子旋转后每个主成分都包含了一些较大和较小的因子载荷，通过突出较大载荷的因子，加强解释某一主成分所代表的原始信息。

图 2: 主成分因子载荷



数据来源: Wind, 中信建投期货

(四) Logistic 回归

跨品种套利聚焦利差的相对变化位置而不是连续的利差预测曲线，离散型变量相对于连续型变量在模型预测方面提供了更灵活的空间，输出结果与套利的方向操作有较大的兼容性。因此我们对因变量 $10-5Y$ 利差做二分类处理，扩大记为“1”，缩小记为“0”。Logistic 回归不同于线性回归，它不要求模型变量间具有线性的相关关系，不要求服从协方差矩阵相等和残差项服从正态分布等，使得模型较为简洁高效。通常来讲 Logistic 回归基于极大似然估计方法逐步选择重要的解释变量，无法避免多重共线性和对原始数据依附性的问题。由于在该案例中，原始数据已经由上述步骤所得的三个独立的主成分代替，规避了类似问题。

$$\ln\left(\frac{p_1}{p_2}\right) = \mu_1 + \omega_1\beta_0 + \gamma_1\beta_1 + \lambda_1\beta_2$$

$$p = \text{probability}(y = 1|x) \quad 1 - p = \text{probability}(y = 0|x)$$

(五) 趋势回归模型信号过滤

对于上述嵌套的基本面因子模型产生的信号，我们使用传统理论中的趋势回归信号进行过滤。可以理解两个模型信号产生共振时，进行开仓操作；信号背离时，平仓离场。通过对历史区间内训练样本进行参数遍历，我们采用 **2.15** 倍的移动 **20** 日标准差作为通道宽度，以较为严格的限制边界来降低该模型下信号产生的频率，提升过滤的稳定性。

表 1：模型信号

基本面因子信号 趋势回归信号	-1	1
-1	卖出利差	平仓或不进行操作
0	卖出利差	买入利差
1	平仓或不进行操作	买入利差

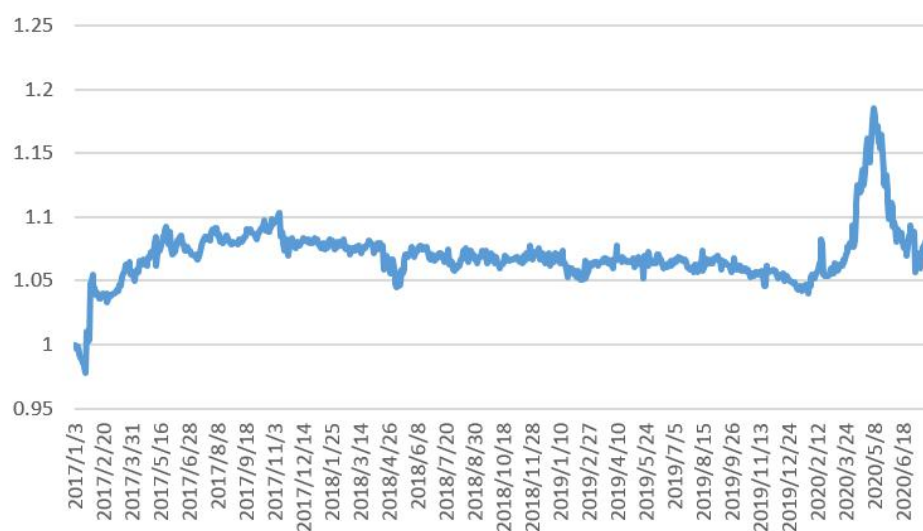
03

实证分析和研究结论

（一）实证分析

我们采用样本外数据作为测试集，考虑安全边际下 **6** 倍杠杆，冲击成本 **0.04%**，手续费每手 **3** 元。考虑到成交量流动性因素，主力合约在到期前 **20** 个交易日展期到下季合约。此外，根据对历史 **5** 年期和 **10** 年期国债现券的统计，我们采用久期配比法按照 **1:1.8** 的比例对 **T** 和 **TF** 合约进行开平仓操作。现假设每个交易日收盘后获得信号，下个交易日开盘价进行操作。

图 3: Logistic 回归净值曲线



数据来源: Wind, 中信建投期货

图 4: 主成分分析+logistic 回归净值曲线



数据来源: Wind, 中信建投期货

图 5：主成分分析+因子旋转+logistic 回归净值曲线



数据来源：Wind，中信建投期货

表 2：回测统计汇总

统计指标	logistic 回归	主成分分析+logistic 回归	主成分分析+因子旋转+logistic 回归	‘主成分分析+因子旋转+logistic 回归’经趋势回归过滤
样本内拟合效果统计				
AIC(赤池信息准则)	4.58e+04	3.23e+04	3.13e+04	
BIC(贝叶斯信息准则)	4.96e+04	3.37e+04	3.24e+04	
修正 R^2	0.61	0.73	0.76	
样本外收益表现				
择时胜率	51.21%	53.39%	59.37%	63.10%
年化收益率	2.16%	5.21%	5.95%	13.57%
累计收益率	7.64%	18.51%	20.83%	47.36%
夏普比率	0.32	0.58	0.78	1.32
最大回撤	10.82%	10.91%	7.15%	4.95%
波动率	4.98%	4.97%	6.85%	5.18%

（二）研究结论

实证研究显示，综合样本内拟合程度和样本外泛化能力两个方面，基于“主成分分析+因子旋转+logistic 回归”的嵌套基本面因子模型经过趋

势回归信号过滤之后，年化收益获得较大提升，且最大回撤显著缩小。

具体来看，主成分分析可以有效避免多重共线性的同时拟合理论模型，因子旋转方法突出了部分因子在主成分上的贡献率，在策略运用中控制了最大回撤。在经济数据普遍具有高维和内在相关性的背景下，上述三个方法的结合使用在尽可能减少原始数据信息丢失的前提下，构建了具有基本面属性和可操作性的利率嵌套模型，同时再结合利差中枢的趋势回归属性，显著优化了模型收益表现。

参考文献

【1】Antti Ilmanen. *Overview of Forward Rate Analysis Understanding the Yield Curve: Part 1*. Salomon Brothers.

【2】Robert J. Shiller, John Y. Campbell, Kermit L. Schoenholtz. *Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates*[J]. *Brookings Papers on Economics Activity*, 1983(1):173-223.

【3】Francis X. Diebold, Canlin Li. *Forecasting the term structure of government bond yields*[J]. *Journal of Econometrics*, 2005 (2).

【4】李宏瑾.利率期限结构的远期利率预测作用——经期限溢价修正的预期假说检验[J].金融研究,2012(08):97-110.

【5】周子康,王宁,杨衡.中国国债利率期限结构模型研究与实证分析[J].金融研究,2008(03):131-150.

【6】赵晶,张洋,尹浩明.中国国债利率期限结构的动态特征:基于 *NS* 和 *SV* 模型的实证判别[J].管理世界,2014(11):164-165.

【7】郭济敏,张嘉为.基于 *Nelson-Siegel* 模型预测中债国债收益率曲线形态[J].债券,2016(07):66-72.