

# 基于宏中观数据的市场利率动态模型研究

摘要：中国金融市场发展迅速，规模持续扩大。利率市场化的进程也在不断加快，市场利率水平受到金融市场供求关系，和国际环境等因素影响，其变动频率和幅度加大。准确刻画市场利率的动态变化，是有效进行利率风险管理的前提。本文宏观角度出发，分析国内经济基本面指标、境外经济指标、实体经济货币供需指标、价格指数、银行间市场流动性水平五个方面的宏观经济数据对市场利率的影响，用多元线性回归，Ridge 回归，LASSO 来构建这些因素与市场利率之间的动态模型。通过 2002.01-2019.08 的相关月度数据，验证模型的有效性。考虑到宏观指标只考虑整体，很难反映出实际利率传导路径，且具有时间滞后性。我们通过分析实效性更强的中观行业数据与利率之间的关系，观察需求在不同行业间的传导变化，来发现利率传导路径。丰富了基于行业基本面数据的市场利率理解，也对行业之间，特别是上下游行业的动态轮动关系进行了量化建模。通过该模型可以对未来行业景气度进行预测估计。通过相关行业数据来验证模型的有效性。

## 一.引言

国债是中央政府发行的政府债券，以国家信用作为担保，是业内公认的安全系数最高的投资工具之一，对宏观经济的调节意义重大。美国次贷危机、欧洲主权债务危机发生后，国债收益率的波动对整个国家的金融货币市场都有举足轻重的影响，而我国国债市场在金融市

场中占据相当大的比重。目前在国内能看到许多关于股票、票据、期货期权方面金融工具的探讨，但是对国债市场的研究尚不充分。

在市场经济国家，市场利率是制约国债利率的主要因素。市场利率一般是指证券市场上各种证券的平均利率水平。一般的原则是，国债利率要保持与市场利率大体相当的水平，且国债收益率与市场利率有一定程度上的正相关性。对比其他收益率曲线，国债收益率曲线具有无风险利率、市场化、期限完整、与其他市场化利率相关性强等优势[1]。国债收益率对内看是金融资源配置的指引，对外看则是一国的主权信用定价。因此，国债利率作为无风险利率，能够较好实现市场利率定价参考作用。

随着利率市场化改革的持续推进，发挥国债收益率曲线的市场基准作用更加迫切。为实现其利率定价引导作用，需建立在国债规模较大、流动性较高、市场影响力更明显的基础上，因此本文选择一年期中国国债到期收益率进行研究。同一期限下，国债到期收益率可能受金融市场利率水平、政府的信用状况和社会资金的供求状况等因素影响。目前，我国国内对这方面的探讨大都局限在经济学理论上的分析，甚少有全面地、大幅度的对各项宏微观指标进行量化并进一步建立模型的研究出现，而学术上需要一个严谨有效的国债收益率研究体系，业界也同样在寻求一个准确、实用的利率预测体系来优化其资产配置和融资决策。

我国自 2004 年推出货币市场基准利率的参考性指标以来，利率市场化的进程在不断加快。随着与国际贸易和金融的密切度加大，我

国的利率水平也受到很多国内外因素的影响，从而使利率风险加大。本文的目的是希望建立一个理论和实证应用的桥梁，为有效进行利率风险管理，准确刻画市场利率的动态变化，以国债收益率为研究对象，探讨基准利率体系的构建问题，对宏观指标、微观行业指标与市场利率之间的关系进行建模，实现基于宏观微观数据的市场利率动态模型。

## 二、文献综述

随着中国利率市场化的进程在不断加快，市场利率水平受到金融市场供求关系，和国际环境等因素影响，其变动频率和幅度加大。准确刻画市场利率的动态变化，是有效进行利率风险管理的前提。国内外学者更多集中研究利率风险。由于同业拆借市场利率是实现利率市场化的核利率之一，国内外许多学者已经对利率的影响因素以及利率预测进行了大量研究，目前对利率进行预测主要分为两部分，一部分学者通过对利率本身数据建立模型进行实证分析。根据历史文献，我国大部分学者目前主要采用自回归移动平均（ARIMA）模型对同业拆借利率进行预测[2]。ARIMA 模型的基本思想为，运用差分变换将非平稳时间序列变为平稳时间序列，然后用变量滞后值和随机误差项的现值对变量进行回归预测[3]。运用 ARIMA 模型短期预测能力较好，而对于长期预测，则误差波动较大，预测能力较差[4]。

后来，有些经济学家在 ARMA 模型基础上改进，使用 EGRACH 模型对我国中债国债收益率进行预测，EGRACH 模型是解决异方差、非对称问题的有效方法，模型参数少，易于操作和控制，效率高。运用

EGRACH 预测的平均预测误差仅为 1.4058%[5]。除外，有些研究发现使用两因素 Vasicek 模型对拟合 SHIBOR 有较好的预测效果，尤其是对 SHIBOR 在 3 个月或更长时间内的预测[6]。

除了上面直接用模型进行利率预测，还有很多经济学家从利率的影响因素入手对利率进行定量预测，比如赵燕梅(2018),选取政府支出、汇率、货币供应量 M2、固定资产投资、通货膨胀率和股票价格指数为解释变量，然后通过建立协整方程和误差修正模型（VEC），得出政府支出对利率正向冲击影响较大的结论[7]。刘晶晶(2009)则选择 M2、固定资产投资、CPI、LIBOR 和美元兑换人民币汇率等影响因素运用线性模型回归进行实证分析，最后给出 M2、LIBOR 和 CPI 三因素对利率的影响[8]。李海涛，王欣，方兆本（2008）运用 Garch 模型从我国央行短期货币市场供求、货币政策以及假期效应三个维度，分析 Shibor 隔夜利率波动的主要影响因素，最后发现发现新股发行对其影响显著远大于存款准备金制度，准备金制度对波动性的影响有增强趋势[9]。盛慧（2014）采用多元回归模型研究隔夜 SHIBOR,其波动的因素选取公开市场操作、消费价格指数和汇率、上证综合指数、短期 Libor、中长期 Libor 对中长期 Shibor,最后给出公开市场操作、物价水平和汇率对隔天 Shibor 的影响较大的结论[10]。HANS DEWACHTER, MARCO LYRIO, KONSTANTIJN MAES（2006）提出并估计了一个包含可观测宏观经济变量和潜在宏观经济变量的连续时间期限结构模型，能够精确地描述美国宏观经济变量和收益率曲线的联合动态[11]。将变量选择方法 Lasso 运用于时间序列数据的探讨，国内外

也有不少研究。Jushan Bai 和 SerenaNg(2008)探索用目标因子来预测经济时间序列,他们首先运用主成分分析方法筛选出主要的成分,紧接着利用 Lasso 在已选出的成分中进行变量选择,最后只选取了极少但信息量很丰富的因子[12]。孙昕(2017)基于 Lasso 回归模型针对变量的选择进行第一步稀疏化处理,挑选出最有影响力的宏观变量[13]。

通过历史文献的研究可以发现对利率定量预测一直是学者研究的重点,研究过程中不同学者根据各种方法选取利率的影响因素,最后通过模型实证分析得出结论。这为我们研究市场利率的动态模型提供很好思路。基于历史文献,本文从宏观角度出发,根据国内经济基本面指标、境外经济指标、实体经济货币供需指标、价格指数、银行间市场流动性水平五个方面的宏观经济数据对市场利率的影响,构建描述这些因素与市场利率之间的动态模型。

### 三. 基于宏中观视角的市场利率分析框架

本文首先从宏观经济指标出发,分析其与市场利率直接的动态关系。通过研究发现,影响市场利率波动的宏观因素有国内经济基本面指标、境外经济指标、实体经济货币供需指标、价格指数、银行间市场流动性水平。分析相关宏观指标与市场利率的动态关系,并对其关系进行建模;同时,考虑到宏观数据无法及时体现行业层面供需关系的变化,这里从中观数据出发,分析实效性更强的中观行业数据与利

率之间的关系，观察需求在不同行业间的传导变化，来发现利率传导路径。丰富了基于行业基本面数据的市场利率理解，也对行业之间，特别是上下游行业的动态轮动关系进行了量化建模。基于宏观中观视角的市场利率模型框架如图 1 所示。

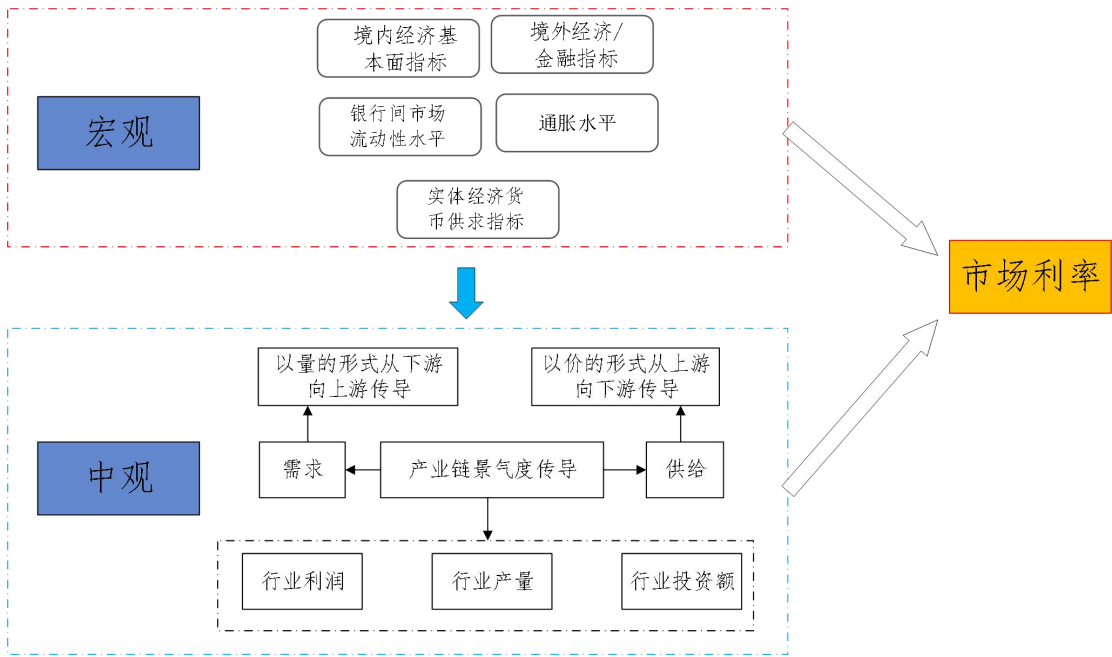


图 1. 基于宏观中观视角的市场利率分析框架

### 1. 基于宏观经济指标的利率预测模型

首先，对于宏观数据，本文将相关影响因素根据五大类展开为 20 项经济指标。通过历史数据进行线性建模来描述他们与市场利率的关系，思路如图 2 所示。

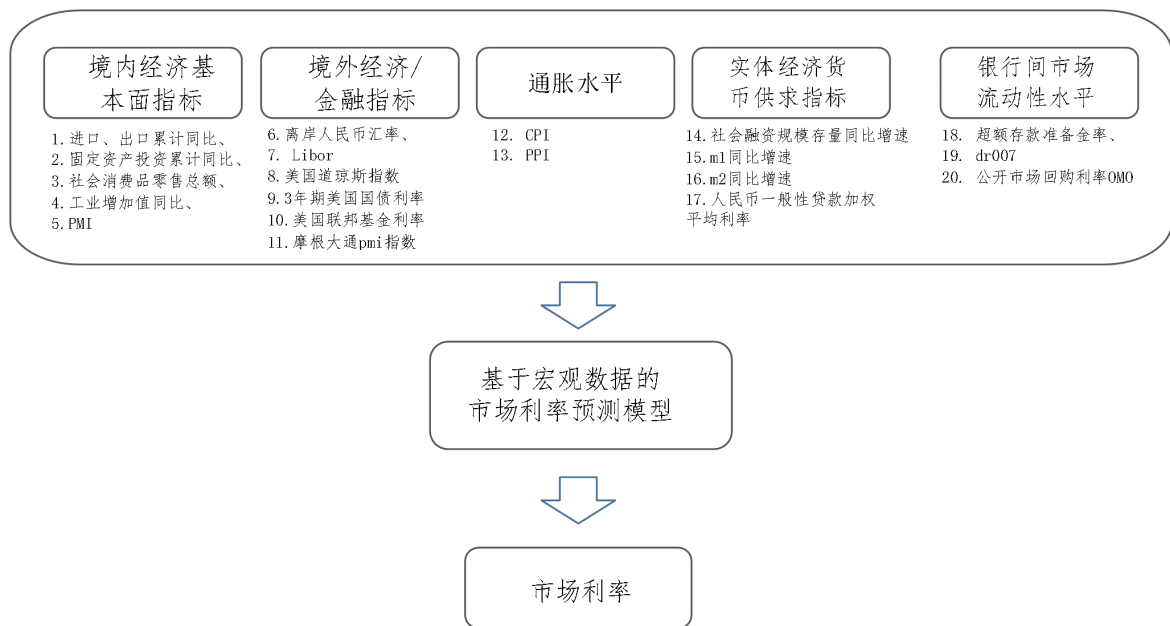


图 2. 基于宏观数据的市场利率预测模型框架

## 2. 基于行业轮动原理的利率传导模型

考虑到市场利率本质上资金的供求关系。如果资金供给大于资金需求，那么利率将下降；如果资金需求大于资金供给，那么利率将上升。资金由谁供给？主要是两个层面：一是企业和居民的闲散资金，二是中央银行，而金融机构的存在是为了提高资金使用效率，把资金引导到需要的地方。如何将资金引流到最需要的地方？无论在时间还是空间上，利率都是最有效的指标，因为它标志着资金的回报水平。因此，利率的变化从根本上应该跟着实体经济的资产回报率走。

从这个逻辑出发，分析中观行业数据，构建其与利率的关系。特别是不同经济周期下，行业有着不同的经济活跃度，进而通过资金的供需反应该行业的景气程度。从空间上，分析行业经济数据的分布对利率的影响。从时间上，观察行业间的轮动规律。一个完善的行业轮动框架，通常涵盖对宏观变量的解读、对行业景气度的分析，以及微观的企业特征等多个维度。目前，行业轮动也是学界和市场界探索得

比较多的一个领域。所谓行业轮动，是指从宏观的角度出发，研究行业轮动中有效的宏观驱动因素及逻辑，并探索其中可量化的因素，并进行建模，逻辑框架见行业逻辑图 3。

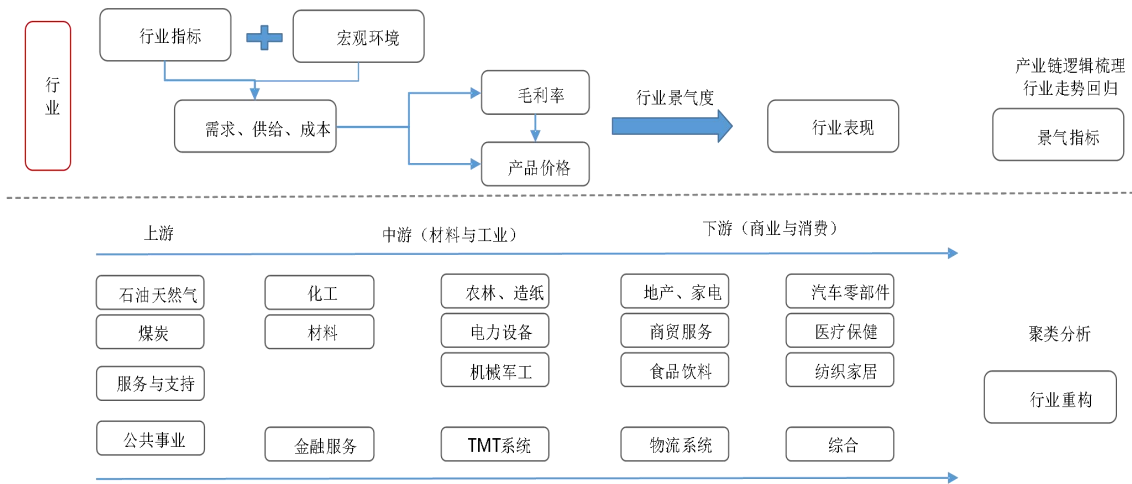


图 3. 行业分析逻辑图

宏观变量对所有行业来说都是唯一的总量指标，不存在横向差异，因此通常采用时间序列方法进行分析， 以下是比较常见的四个不同分析角度：

### 1) 样本分类法

这种方法在于首先分析时间样本的统计规律，根据统计规律在不同的时间点进行不同行业的配置。 常见的“美林时钟”就属于该类方法，在不同经济周期运行状态下配置合适的行业。该方法虽逻辑与结论简单直观，且易于操作， 但由于对样本的分类存在显著的滞后性，实际应用中偏差与争议较大。

### 2) 时间序列回归

此类方法采用回归的方法，利用宏观数据或者市场情绪等数据对未来的行业指数走势进行拟合，从而利用拟合的趋势对行业指数走势



进行预测。这一方法也较为简单，但缺点在于缺乏足够清晰的逻辑以及容易陷入过拟合，若考虑宏观指标的实际披露存在时间滞后，计算常见宏观变量与行业收益率相关系数，发现二者并不存在显著的相关性，单纯采用宏观因子完整序列进行一个行业轮动分析，难以获得显著结果。

### 3) 模式匹配法

在行业历史数据中寻找模式，基于行业自身相对变化来预测行业的下一阶段的变化，比如技术分析。 这种方法的假设是历史会重复发生，缺点是缺乏因果关系，并且行业数量有限， 在统计意义上并不能构成大样本。

### 4) 宏观事件驱动法

与时间序列回归法不同的是，事件驱动法只针对少数情况，当宏观指标处于极端状态并触发相应“宏观事件”时，才研究其对特定行业的影响，该方法所得到的结论通常更加直观，且由于剔除了大量可能无效的数据信息，分析结果有效性也更加显著。因此，在行业分析上，拟尝试从宏观事件驱动的角度，利用统计学方法，寻找宏观事件和行业表现之间的关系，为行业配置提供建议。 在具体的因子选取中，既考虑宏观指标，也纳入各行业特有的指标进行测算。总体而言，该方法较为合理，后文中宏观因子的选取以此方法为主。

由于各行业的表现不仅受宏观因素的影响，还受不同的产业变量的影响，尝试对产业逻辑进行梳理，从需求、供给、成本等影响行业景气的驱动因素出发， 除通用的宏观因子外， 构建每个行业都有特

有或产业上下游的相关指标，附录表 1 是各行业的一部分重要指标展示。

每类宏观指标，均衍生定义以下 8 种事件模式（如附录表 2 所示），对于每个行业通过将不同的宏观因子或行业指标和不同的事件模式相结合，构成特定的事件，并统计分析在不同的事件下各行业指数的表现，从而筛选出对于特定行业指数具有显著影响力的事件。因子事件中所包含的参数 K1、K2、K3 将通过样本内优化的方法确定。

对于特定的某个因子事件，我们观察历史上该因子事件发生后下一期各个行业的超额收益。如果历史上每次某个因子事件发生之后，某行业在下一个时间段的超额收益比较稳定，那么我们就认为该因子事件对于该行业未来一段时间内的收益存在较为显著的影响。将申万一级 28 个行业和通用的宏观事件及各自的事件库两两交叉，测算出每个行业最有效的预测事件。

在为各行业筛选历史上对其未来收益率影响较为显著的因子事件时，可采用两个标准进行筛选：历史上因子事件的发生次数、以及事件对下一期的行业轮动预测的有效性。为保证因子事件具有参考意义，在筛选时需要考虑因子事件历史上发生的次数，当对于某个因子的极值事件发生次数不低于 8 次时，我们才考虑其对于未来资产收益率的预测能力；同时，我们定义因子事件 IR 为：每次因子事件发生后，下一期平均超额收益/超额收益的标准差，如果因子事件 IR 的绝对值越大，则该因子事件对于某一个资产未来收益的影响将更加稳定，设置 IR 绝对值阈值为 0.7。

## 四. 线性模型介绍

### 1. 模型背景

在对数据建模的过程中，伴随着维数的增加，所需的样本数也急剧增加。在样本数不足的情况下，容易发生过拟合。解决此类问题的其中一种方法是增加样本量，但考虑到有些数据样本量有限，我们需要使用减少样本特征的方式来进行降维。因此，LASSO 和岭回归成为不错的选择。

### 2. LASSO 回归与岭回归

LASSO 回归[14]和岭回归[15]都采用增加正则化项作为惩罚项来防止过拟合的出现。正则化项一般是模型复杂度的单调递增函数，而经验风险负责最小化误差，使模型偏差尽可能小。经验风险越小，模型越复杂，正则化项的值越大。要使正则化项也很小，那么模型复杂程度受到限制，因此能够有效地防止过拟合。

$$\min_{f \in F} \left[ \frac{1}{m} \sum_{i=1}^m L(y_i, f(x_i)) + \lambda J(f) \right] \quad (1)$$

正则化项一般具有如下形式的优化目标：

其中， $\lambda \geq 0$  是用来平衡正则化项和经验风险的系数。正则化项可以是模型参数向量的范数，经常使用的是  $l_1$  范数和  $l_2$  范数。当使用  $l_1$  范数时，我们得到 LASSO 回归；使用  $l_2$  范数时，得到岭回归。

### 3. 线性回归

上述原理介绍强调了正则化项的优越性，但其实 LASSO 回归和岭回归的优点同样体现在其经验风险的多样性上。本文将使用线性回归中的均方差来度量 LASSO 回归和岭回归的经验风险。

倘若设自变量为  $x$ ，因变量为  $y$ ，则有一般的线性回归模型为

$$y = x\beta = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (2)$$

线性回归的优化目标通常为最小化均方差，即

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2 \quad (3)$$

我们可采用最小二乘的方法求得最优的参数向量为

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (4)$$

其中  $X$  为已知的信息矩阵。

LASSO 回归的模型即为

$$y = X\beta + \alpha \|\beta\|_1 = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \alpha(|\beta_0| + |\beta_1| + \dots + |\beta_n|) \quad (5)$$

Ridge 回归的模型即为

$$y = X\beta + \alpha \|\beta\|_2^2 = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \alpha(\beta_0^2 + \beta_1^2 + \dots + \beta_n^2) \quad (6)$$

## 五. 变量选取与数据的处理

### 1. 变量的选取

市场利率是由资金市场上的供求关系决定的，影响其波动的主要因素有：国内经济基本面指标、境外经济指标、实体经济货币供需指标、价格指标、银行间市场流动性水平等。

本文对市场利率作建模分析,选取了可能影响市场利率的多个变量以进行下一步的研究筛选:

### 1) 境外经济/金融指标

用于衡量境外经济、金融对境内的传导。包括:

- (1). 离岸人民币汇率。这里我们选择使用美元兑人民币的平均汇率。
- (2). LIBOR。即伦敦银行同业拆放利率,是目前国际普遍参考及采用的市场基准利率。这里我们选用的期限为七天。
- (3). 美国道琼斯指数。此处指道琼斯工业平均指数,是世界上影响最大并且最有权威的股票价格指数。可以反映美国乃至整个世界的经济运行情况。
- (4). 1 年期美国国债利率。即由美国政府发行的短期国库券。由于其流动性很高且无信用风险的性质,被视为是十分具有参考价值的利率。
- (5). 美国联邦基金利率。即美国银行同业拆借利率,是美国金融市场最重要的短期利率之一。
- (6). 摩根大通 PMI 指数。即摩根大通编制的全球综合采购经理指数,可以反映全球经济总体运行情况以及其增速。

### 2) 境内经济基本面指标

用来衡量经济发展情况(指标取累计同比或转换成当月同比)包括:

- (1). 进口、出口累计同比增速。反映国家一定时间内进出口水平。
- (2). 固定资产投资累计同比增速。此处我们具体采用的是房地产开发投资累计同比的增速。
- (3). 社会消费品零售总额同比增速。

(4). 工业增加值同比增速。

(5). PMI。即采购经理人指数。

其中，进出口、固定资产投资和消费三个指标是衡量社会总需求的，工业增加值和 PMI 是衡量生产的。

### **3) 价格类指标**

(1). CPI,即消费者价格指数。是反映应总体价格水平以及通货膨胀或紧缩程度的重要指标。这里我们使用的是其同比的增长率。

(2). PPI,即生产者价格指数。对总体物价水平也具有指导作用,同样我们采用其同比增长率。

### **4) 实体经济货币需求及供给的指标**

(1). 社会融资规模存量同比增速。即实体经济进行融资获得的资金的同比增长率。是反映实体经济市场和金融货币市场之间联系的重要指标,与利率具有较高的相关性。

(2). M1 同比增速。即狭义货币供应量的同比增速。可以直接反映宏观经济中的现实购买力,也是国家货币政策中的关键因素。

(3). M2 同比增速。即广义国币供应量的同比增速。相比 M1 可以反映经济中总体的潜在购买力。

(4). 人民币一般性贷款加权平均利率。即经过加权计算后的能反映混合贷款成本的利率。

### **5) 银行间市场流动性水平指标**

(1). 超额存款准备金率。即保证存款支付和资金清算的货币资金占存款总额的比率,反映了银行的支付能力以及流动性。

(2). DR007。即银行间存款类机构以利率债为质押的 7 天期回购利率。

是较为有效的反映银行间流动性水平的指标。

(3). 公开市场回购利率 OMO。即公开市场的 7 天国债回购利率。

## 2. 数据的处理

我们从万得数据库中下载得到 2004 年至今的以上数据并进行了如下的处理：

1) 对于少数同比增长数据的缺失值,由于其数量较少,可直接对其赋值为后一期的数值进行处理。

2) 所有的数据均为月度数据,对更新频率大于一个月的进行简单插值处理。

3) 为消除不同量纲对模型带来的影响,以及考虑到部分同比增长率数据存在负值,对数据进行归一化处理：

$$X_i^* = \frac{2(X_i - X_{\min})}{X_{\max} - X_{\min}} - 1 \quad (7)$$

其中  $X_i$  为原数据,为  $X_i^*$  归一化后的数据,  $X_{\max}, X_{\min}$  为原数据的最大,最小值。归一化后所有数据的范围均在区间  $[-1, 1]$  上。

至此我们便可以进入到预测模型的训练以及结果分析部分。

# 六.利率预测模型实证研究

## 1. 模型描述

上文提到, LASSO 回归和岭回归是在一般的线性回归基础上,在方程式最后分别加上参数的  $l_1$  范数和  $l_2$  范数作为惩罚项,起到收缩

参数大小的作用，来减轻模型的过拟合。

因此，我们分别采用 LASSO 回归、岭回归和无惩罚项的一般线性回归对上述变量进行拟合，预测 1 年期国债利率。实验证明，对比一般线性回归和岭回归，LASSO 回归得到的结果具有稀疏性这一优良性质，该模型会对输入变量进行一定的特征选择，使得最终的结果更好。

由于 LASSO 回归采用了  $l_1$  范数，因此为求解参数带来了一定的麻烦。带有绝对值的函数是不可导的，但绝对值会驱使结果有稀疏性，即使得结果中出现更多的 0 分量，这是一个良好的性质。在本论文使用的数据中，由于变量众多，其中一些金融数据不可避免的存在一些共线性，并且金融数据大多具有一定的自相关性，因此采用简单的多元线性回归与岭回归不见得优于 LASSO 回归。

数据的时间跨度为从 2004 年 2 月至今，时间频率的最小周期是月，数据的搜集工作与预处理已在前文叙述。我们将所有数据按照数据特点与经济周期划分为三个时间段，分别为 2004 年 2 月至 2009 年 10 月、2009 年 11 月至 2016 年 12 月、2016 年 12 月至 2019 年 8 月。我们对以上三段模型进行分别建模，模型均采用 LASSO 回归。最终，这三个时间段中实际采用的变量数据为：

04-09 模型包括变量：

美元兑人民币：平均汇率；美国道琼斯指数；工业增加值同比；CPI 同比增长；PPI 同比增长；m1 同比增长；m2 同比增长；DR007；滞后一阶国债利率

09-16：



摩根大通 PMI 指数；DR007；滞后一阶国债利率

16-19:

房地产开发投资累计同比；社会消费品零售总额同比增长；社会  
融资规模存量同比增速；DR007；滞后一阶国债利率

2. 模型评估

通常情况下，对线性回归模型效果的评估主要采取的评估指标为：  
拟合优度  $R^2$ ，均方根误差 RMSE。三个 LASSO 回归模型的表现如表 1 所  
示

表 1. LASSO 拟合度和误差

时间	$R^2$	RMSE
04-09	0.9475536	0.1280997
09-16	0.9072091	0.1526858
16-19	0.9227256	0.1575957

$R^2$  能够切实的反映出模型的拟合效果，可见表中的  $R^2$  均大于 0.9，  
说明模型效果较好。同时，RMSE 也较小，符合预期。

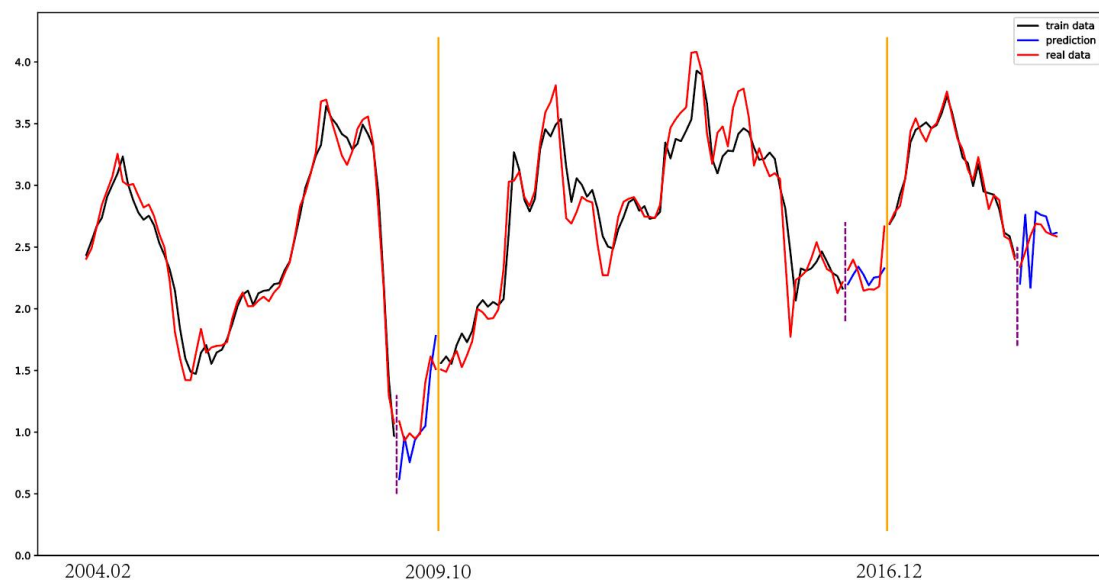


图 4. LASSO 拟合效果图

图 4 为 LASSO 模型拟合图，红色线代表真实数据，黑色线为模型在训练集中的表现，蓝色线为模型测试结果。我们每次使用前  $t$  个自变量去预测第  $t+1$  个因变量，以这种方法来预测每一时间阶段的最后 8 个样本点，可从图中看出，预测效果较好。

我们也将线性回归及岭回归对模型的模拟结果与 LASSO 模型进行对比，将线性回归与岭回归的拟合表现如表 2 和表 3 所示。

表 2. 线性回归的拟合度和误差

时间	$R^2$	RMSE
04-09	0.9789464	0.0811621
09-16	0.9354166	0.1273816
16-19	0.9710369	0.0964825

表 3. 岭回归的拟合度和误差

时间	$R^2$	RMSE
----	-------	------

04-09	0.9781961	0.0825958
09-16	0.9285349	0.1339964
16-19	0.9506233	0.1259758

综合对比三个模型的  $R^2$  和 RMSE，不难发现三个模型的指标值相差不大，如果仅从指标值的角度出发，似乎线性回归是最好的模型。但 LASSO 模型使用变量较少，其模型具备鲁棒性，并且 LASSO 模型使用了更少的变量，就提取出了较多的主要信息，说明其模型性质较好。然而，在实际情况中，金融数据间往往存在一定相互关系，仅使用少量变量来对国债利率做出拟合预测，从解释变量的角度来看并不充分。因此，我们也可直接采用线性回归模型来作为最终的模型。这里我们也给出其余两个模型对应的拟合图，如图 5，图 6 所示。

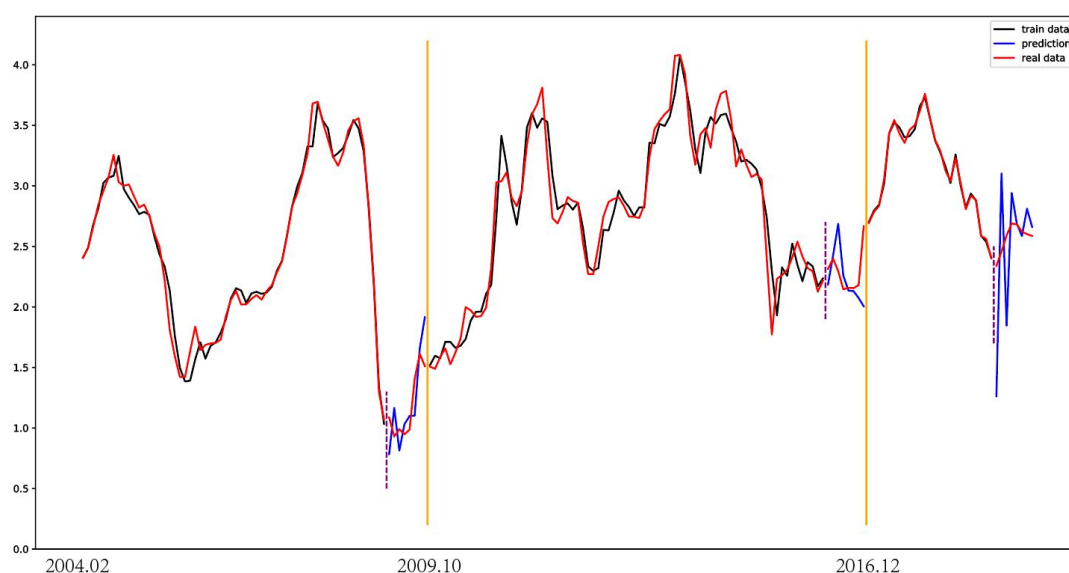


图 5. 多元线性回归拟合效果图

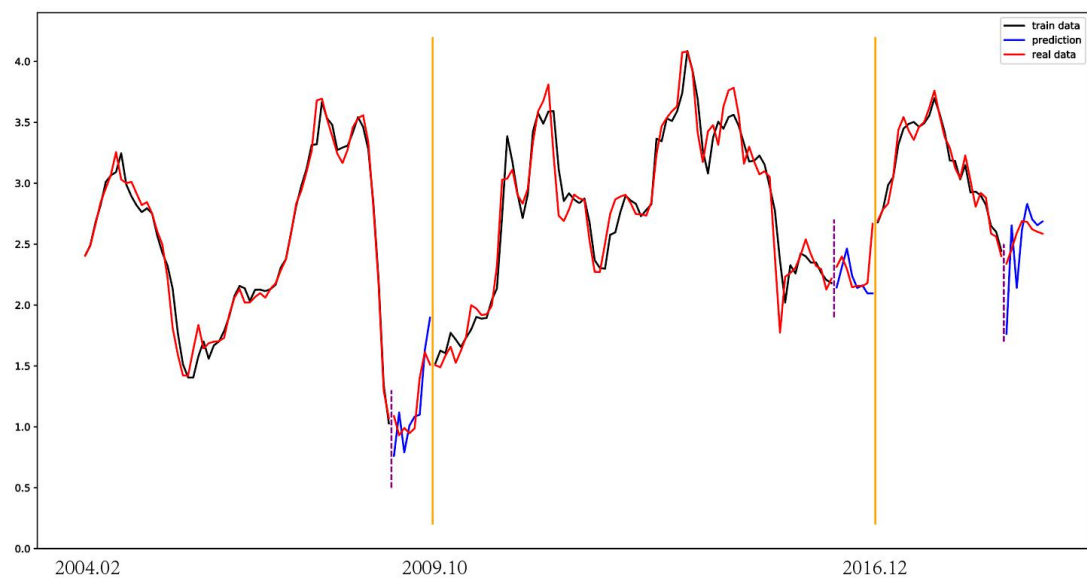


图 6. Ridge 回归拟合效果图

除上述对模型的评估及对比外，我们也尝试将预测变量的一阶滞后变量删去，使用原始变量来做拟合。同样地，我们列示出相关的指标值如表 4：

线性回归		
时间	$R^2$	RMSE
04-09	0.9392537	0.1378638
09-16	0.8652608	0.1839895
16-19	0.9565707	0.1181457
LASSO 回归		
时间	$R^2$	RMSE
04-09	0.8926737	0.1832497
09-16	0.6833611	0.2820514

16-19	0.7759628	0.2683406
岭回归		
时间	$R^2$	RMSE
04-09	0.9274383	0.1506759
09-16	0.8436695	0.1981837
16-19	0.9366908	0.1426460

可以看出，将国债利率的一阶滞后值去掉后，所有模型均出现了一定程度的拟合效果下降，其中 LASSO 模型已经难以从少量的变量中提取中更多的信息。下面展示出 Ridge 回归的拟合图：

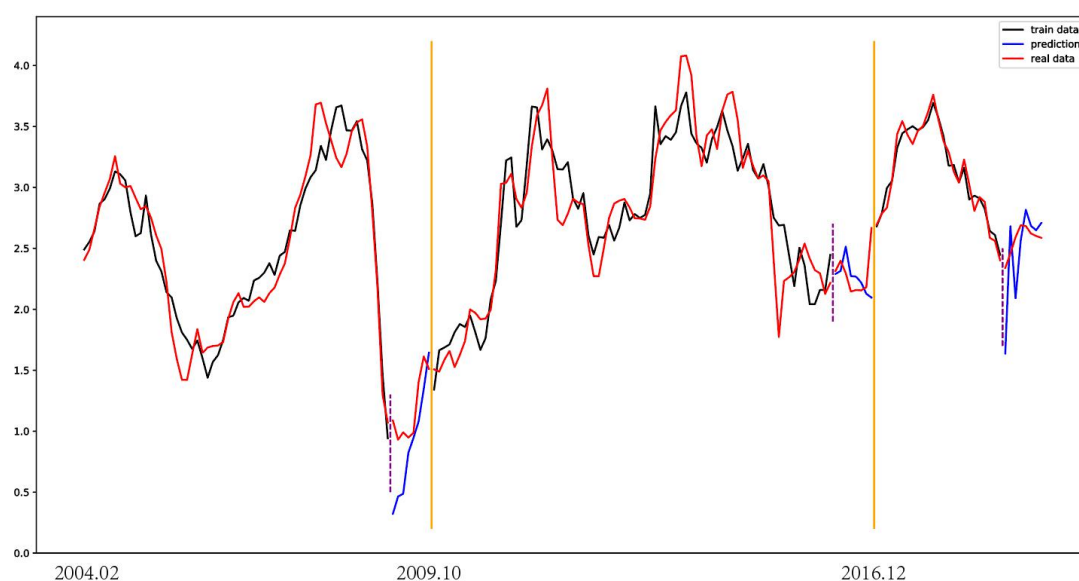


图 7. Ridge 回归拟合效果图（去除  $y_t$ ）

可以看到，拟合效果确实出现了下降的情况，这说明国债一阶滞后的变量中包含了一定的信息，即反映出预测变量具有自相关性。同样，这里也给出其余两个模型的拟合图作为对比，如图 8，图 9。

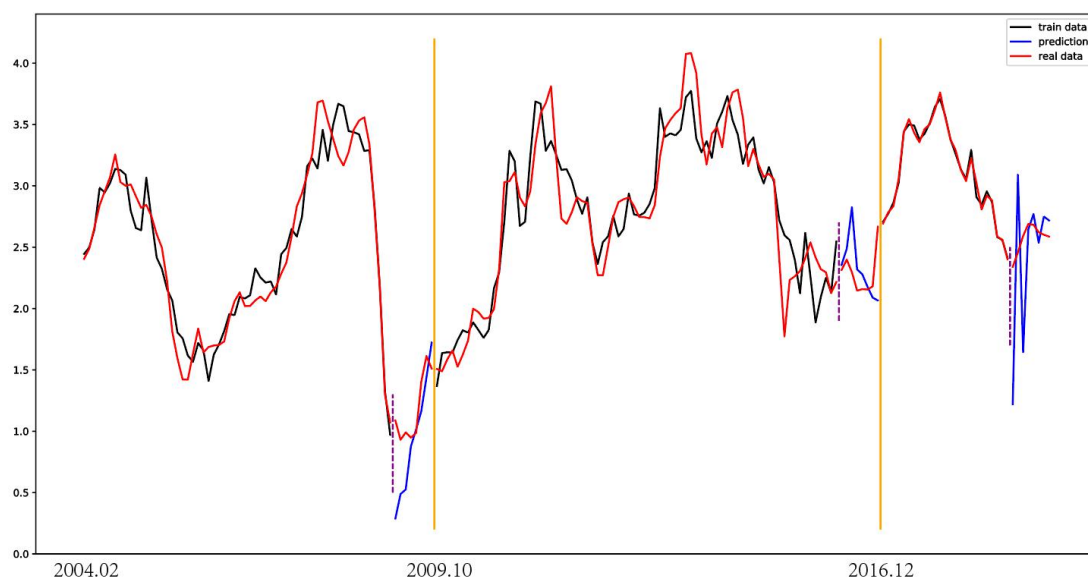


图 8. 多元线性回归拟合效果图（去除  $y_t$ ）

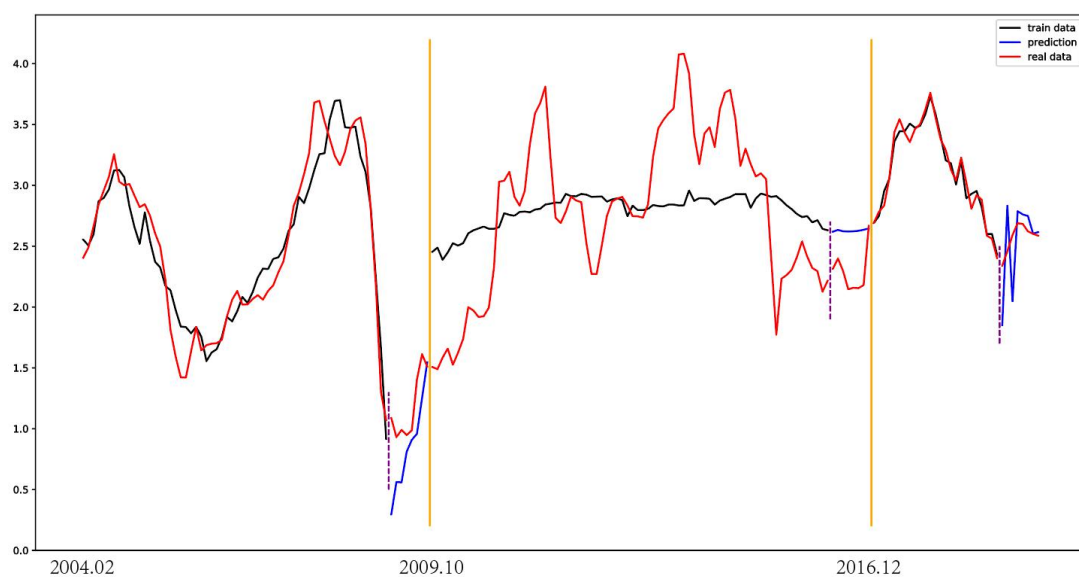


图 9. LASSO 回归拟合效果图（去除  $y_t$ ）

可以看到中段的 LASSO 回归效果很差，这也间接说明滞后一阶的国债利率确实会对预测造成较大的影响。

### 3. 残差分析

我们再来看各模型残差的箱型图，如图 10 所示。

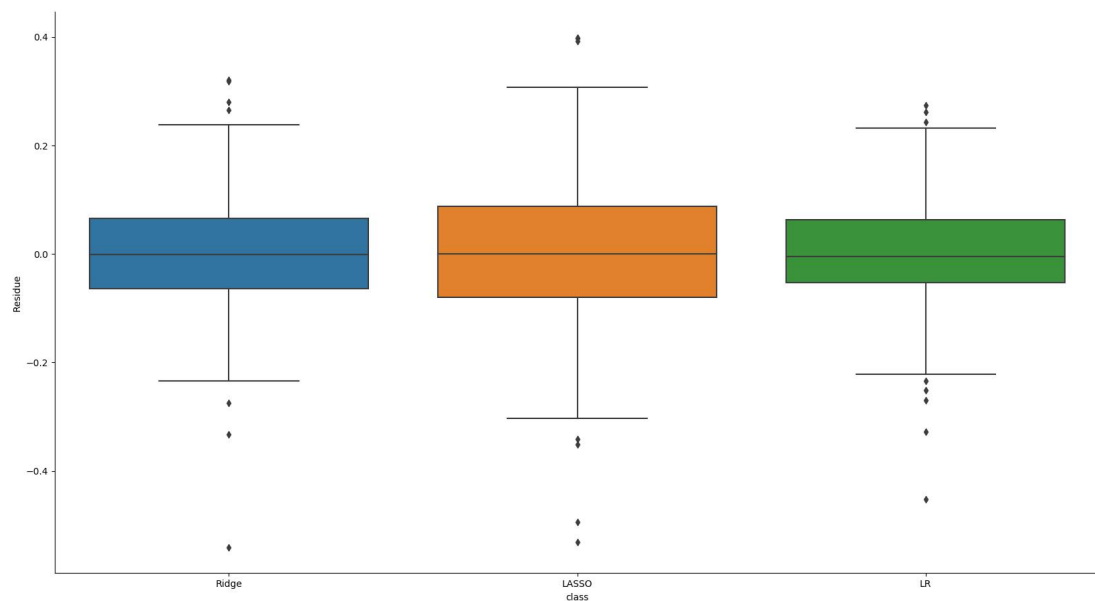


图 10. 残差箱型图

可从残差箱型图中看出，残差围绕 0 上下波动，这是正常的；但对于每个模型而言，均有一些异常值突破箱体，说明我们的模型依旧有可以提高的地方。

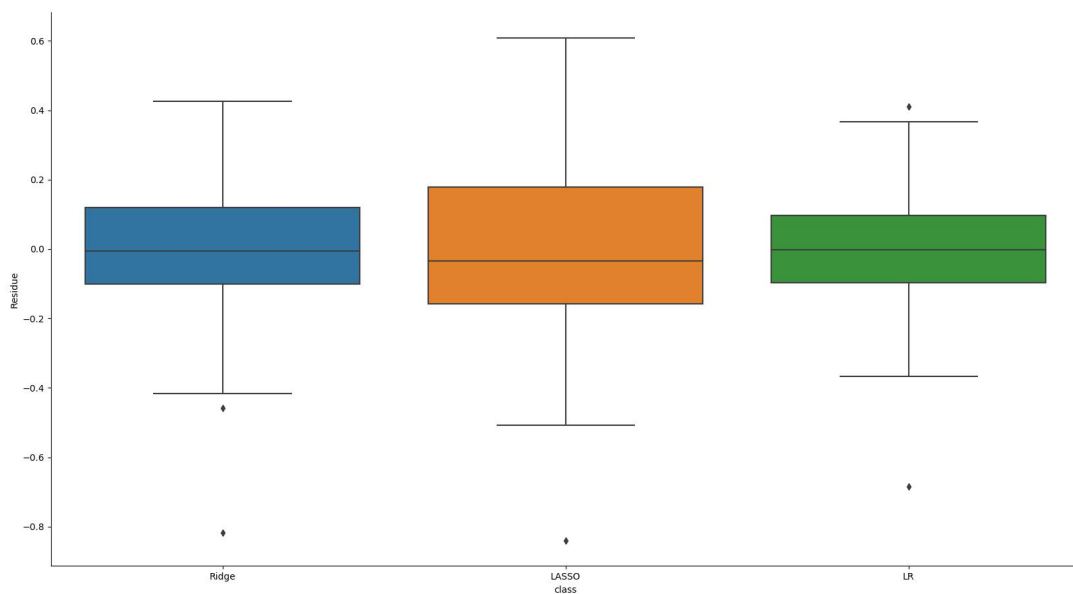


图 11. 残差箱型图（去除  $y_t$ ）

图 11 为将国债的一阶滞后变量去除后得到的残差箱型图，可以看到，突破箱体的异常值显著减少。这说明，虽然从  $R^2$  的角度

来看，将国债的一阶滞后变量去除后拟合优度有所下降，但其残差却呈现出更好的性质。同时，LASSO 回归模型的箱体显著增大，说明此时 LASSO 回归的表现差强人意。

## 参考文献

- [1] 关于利率市场化背景下国债收益率曲线作为定价基准的专题调查 ——以马鞍山辖内金融机构为例\_徐锦瑞、尹源、丁梦、陈秀云
- [2] 郑尧天，杜子平，ARIMA 模型在同业拆借利率预测中的应用，2007
- [3] 孙继国，伍海华.我国银行间同业拆借利率的时间序列预测模型[J].统计与决策，2004(5):33-34.
- [4] 宋华，姚晓军基于 ARMA 模型的银行间同业拆借利率预测，淮南师范学院学报，（2018）02-0011-05
- [5] 舒服华，陈传杰.基于 EGRACH 模型的我国中债国债收益率预测,金融教育研究,2017(9):30-5
- [6] Chaoqun Ma, Jian Liu, and Qiujuan Lan, 2014, Studying Term Structure of SHIBOR with the Two-Factor Vasicek Model, Hindawi Publishing Corporation
- [7] 赵燕梅.我国银行间同业拆借利率的影响因素分析——基于 VEC 模型的实证研究，（2018）05—0030—13
- [8] 刘晶晶，基于利率市场化 SHIBOR 利率影响因素的研究，2009
- [9] 李海涛，王欣，方兆本，基于偏 t 分布的 Shibor 隔夜拆借利



率影响因素分析, 2008

[10] 盛慧, 上海银行间同业拆借利率影响因素的实证分析, 2014

[11] HANS DEWACHTER, MARCO LYRIO, KONSTANTIJN MAES, A joint model for the term structure of interest rates and the macroeconomy, 2006

[12] Jushan Bai, Serena Ng (2008) Forecasting economic time series using targeted predictors, 2008

[13] 孙昕, 基于 LASSO 变量选择方法的投资组合及实证分析, 2017

[14] Robert Tibshirani, Regression Shrinkage and Selection Via the Lasso, 1996

[15] Arthur E. Hoerl & Robert W. Kennard, Ridge Regression: Applications to Nonorthogonal Problems, 1970

[16] BLACK, F., DERMAN, E., AND TOY, W. 1990. "A One-Factor Model of Interest Rates and its Application to Treasury Bond options," Financial Analysts Journal, 33-39

[17] COURTADON, G. 1982. "The Pricing of Options on Default-Free Bonds", Journal of Financial and Quantitative Analysis 17:75-100