

doi:10.19920/j.cnki.jmsc.2021.01.007

高风险低收益？基于机器学习的动态 CAPM 模型解释^①

姜富伟, 马甜, 张宏伟

(中央财经大学金融学院, 北京 100081)

摘要: 我国股票市场存在高风险股票反而伴随较低收益的低风险定价异象, 这有悖于传统资产定价理论. 本文使用宏观经济和微观企业特征构建了六百多个变量的宏微观混合大数据集, 并结合多种经典机器学习算法开发了基于大数据和机器学习的智能动态 CAPM 模型, 检验了时变系统性风险对我国股市收益解释能力. 实证结果表明: 本文的智能动态 CAPM 定价模型能够显著解释我国股市低风险定价异象; 随机森林等非线性机器学习算法表现最佳; 影响股票时变系统风险的主要因素是市场类因子, 基本面因子居次. 本文对于我国股市系统性风险测度、动态资产定价模型构建和金融与大数据和人工智能融合创新有重要理论与实践指导意义.

关键词: 系统性风险; 动态 CAPM; 机器学习; 金融大数据

中图分类号: F832.5 **文献标识码:** A **文章编号:** 1007-9807(2021)01-0109-18

0 引言

股票市场系统性风险测度以及风险与收益之间的权衡关系是现代资产定价研究的核心问题. 作为最为著名的均衡资产定价理论, 资本资产定价模型 (capital asset pricing model, CAPM) 用市场贝塔 (β) 来测度股票的系统性风险暴露, 并预测高贝塔风险的股票应该获得较高收益. 然而, 随着资产定价研究的深入, 学者们近年来发现实际情况并非如此. 比如, Fama 和 French^[1, 2] 发现在控制规模和账面市值比因子后, 风险-收益定价线变得平坦; Frazzini 和 Pedersen^[3] 进一步发现高贝塔股票呈现较低的超额收益, 这类现象现在一般统称为资产定价的“低风险定价异象” (low risk anomalies).

按照贝塔风险测度是否随时间动态变化, 现有研究主要分为静态 CAPM 和动态 CAPM 两类. 静态 (unconditional) CAPM 模型假设贝塔固定不变, 采用股票收益数据和线性回归建模贝塔. 而动

态 (conditional) CAPM 模型假设贝塔是动态时变的, 并使用条件信息集改进贝塔测度. 比如, Oliver 等^[4], Cederburg 和 O’Doherty^[5], 丁志国等^[6], 邓可斌等^[7] 指出包含宏观经济信息的时变贝塔可以更精准的刻画不同时期的股票系统性风险, 有更强的股票收益解释力. 但现有动态 CAPM 文献多数只考虑少数几个宏观变量, 并使用简单线性回归建模时变贝塔, 而忽视了大数据和机器学习技术的快速发展.

本文创新构建了基于大数据和机器学习的智能动态 CAPM 模型, 并检验时变系统性风险对低风险定价异象的解释能力.

主要有三大学术贡献. 首先, 本文指出, 我国股市时变系统性风险不仅受到宏观经济的影响, 同时还受到微观企业经营状况的影响. 比如, 宏观经济下行会导致股市系统性风险上升, 而其中具有低市值、高波动和高杠杆率等微观企业特征的股票的系统性风险可能会升高更多. 因此, 呼应金融大数据的飞速发展, 构建了包括 74 个微观企业

^① 收稿日期: 2019-07-29; 修订日期: 2020-08-01.

基金项目: 国家自然科学基金资助项目 (72072193; 71872195); 国家社科基金资助重大项目 (19ZDA098).

作者简介: 姜富伟 (1983—), 男, 山东烟台人, 教授, 博士生导师. Email: fwjiang@foxmail.com

特征和8个宏观经济指标在内的共600个宏微观混合大数据集,并利用宏微观大数据进行股票系统性风险测度.在数据维度和颗粒精度上较现有文献均有了大幅提升,从而可以更准确、更全面的衡量我国不同股票在不同时期的时变风险特征.

其次,应用人工智能和机器学习算法开展股票系统性风险建模并构建了智能动态CAPM模型.传统动态CAPM模型通常基于简单线性回归进行时变贝塔建模,面对高维大数据时容易陷入过度拟合和维度陷阱,同时在复杂数据分析上忽视了因子结构(latent factor)、稀疏性(sparsity)和非线性等重要信息,造成模型精度下降和解释能力不足.因此,本文使用包括主成分回归、偏最小二乘回归、弹性网络和随机森林等多种主流机器学习算法替换线性回归模型,对600个宏微观大数据进行数据降维、变量选择和非线性系统性风险建模.本文使用前沿的机器学习算法改进股票时变贝塔建模,能充分挖掘大数据蕴藏的预测信息,从而可以更灵活、更智能的测度我国股票市场系统性风险.

最后,高风险低收益的低风险定价异象给传统资产定价模型极大的挑战,Frazzini和Pedersen^[3]等主流文献多使用行为金融和有限套利理论来解释这一异象.基于本文提出的智能动态CAPM模型,提出时变贝塔也有助于解释这一异象,使模型的超额收益或阿尔法(α)趋于零.为避免传统机器学习存在的“黑箱”问题,本文强调复杂数据和算法的可解释性,发现随机森林等非线性算法表现好,市场类因子最重要,基本面因子居次,并给出经济学解释.从时变风险角度解释了低风险定价异象,提高了智能动态CAPM资产定价理论在我国股票市场的适用性.

本文的社会经济价值体现在金融科技、金融大数据与人工智能等交叉融合创新.大数据和人工智能技术正在深刻改变世界尤其是金融系统,迫切需要学术界加强基础研究和交叉创新,推动金融科技落地.对于服务金融大数据和人工智能产业发展,服务金融业数字化和科技化改造有着重要的应用价值.

本文的社会经济价值还体现在维护国家金融安全和防范化解系统性金融风险.近年来,我国股票市场快速稳步发展,整体行业产值占国民生产

总值比重不断加大,对我国经济和金融安全有着越来越重要影响,以“股灾”为代表的股市系统性风险事件会对金融系统和实体经济产生重要负面冲击^[8-10].本文对健全我国股市系统性风险测度、归因和预警,促进经济社会平稳有序发展有着重要的应用价值.

1 理论模型和相关文献

1.1 动态CAPM理论模型

资本资产定价模型(CAPM)使用贝塔来衡量股票的系统性风险暴露,传统的静态CAPM理论假设贝塔是不变的,而动态CAPM理论考虑股票风险受到企业自身微观特征和外在宏观经济环境影响,是动态变化的,即贝塔(β)是时变的.因此,基于时变 β 构建的动态CAPM模型相较静态模型理应可以更加准确测度股市系统性风险和股票收益的动态变化.

本文采用Ferson和Sigel^[11]构建的经典动态CAPM理论框架

$$\alpha_{i,t} \equiv E(R_{i,t} | Z_{i,t-1}) - \beta_{i,t} E(R_{m,t} | Z_{i,t-1}) = 0 \quad (1)$$

为上期已获知的条件信息, $E(R_{i,t} | Z_{i,t-1})$ 为第*i*只股票在第*t*期在信息集 $Z_{i,t-1}$ 下的条件预期收益, $E(R_{m,t} | Z_{i,t-1})$ 为市场组合在第*t*期的预期收益.

$\beta_{i,t}$ 是股票系统性风险测度,是本文研究的核心变量.传统静态CAPM模型一般假设 $\beta_{i,t}$ 固定不变的, $\beta_{i,t} = \beta_i$;而本文采用的动态CAPM则假设 $\beta_{i,t}$ 是动态时变的.在线性动态CAPM和线性时变贝塔文献中, $\beta_{i,t}$ 一般表示为 $Z_{i,t-1}$ 的线性函数, $\beta_{i,t} = \beta_{i,0} + \beta_{i,1} Z_{i,t-1}$, $\beta_{i,0}$ 是常数项, $\beta_{i,1}$ 是 $Z_{i,t-1}$ 对 $\beta_{i,t}$ 影响系数,而 $Z_{i,t-1}$ 通常是少数几个宏观经济变量.

关于动态CAPM模型的相关研究有:Hansen和Richard^[12]较早研究了线性动态CAPM理论和时变贝塔模型;Jagannathan和Wang^[13]、Lewellen和Nagel^[14]发现动态CAPM模型可以更好的解释美国市场横截面股票收益;Ferson和Sigel^[11]研究了条件信息对于构建组合策略有效性的影响;Cederburg和O'Doherty^[5]通过滞后期的 β 值、市场分红率和信用利差等宏观变量构建动态CAPM,发现可以有效的减少风险定价异象;Cose-

mans 等^[15]在估计股票动态 β 时引入了宏观变量（信用利差等）和微观变量（账面市值比等）以及两者的交叉乘积项。针对中国市场，丁志国等^[6]在早年研究中发现中国市场中系统风险 β 具有明显的时变特征，邓可斌等^[7]构建了宏微观混合 β 的动态 CAPM 模型来研究股市的系统性风险，发现加入宏观变量后模型可以更好的估计市场风险，尹力博等^[16]使用时变 CAPM 模型研究了中国股市各类市场异象在不同阶段的显著性，蒋志强等^[17]同样使用动态 CAPM 模型对我国股市可预测性进行了探索。

1.2 基于大数据和机器学习的动态 CAPM 模型

目前多数研究仅使用小数据集和线性回归算法估计时变贝塔 β ，忽视了大数据的信息含量，也忽视了因子降维、变量选择和非线性建模等前沿机器学习算法能带来的更准确的预测能力。考虑现有文献的不足，本文构建了基于宏微观大数据和机器学习算法的中国股市动态 CAPM 模型

$$R_{i,t} = a_i^c + (\beta_{i,0} + \beta_{i,1}f_t(Z_{i,t-1}))R_{m,t} + \mu_{i,t} \quad (2)$$

如果模型成立，则 $a_i^c = 0$ ，即没有错误定价。股市系统性风险使用时变贝塔测度

$$\beta_{i,t} = \beta_{i,0} + \beta_{i,1}f_t(Z_{i,t-1}) \quad (3)$$

其中 $\beta_{i,0}$ 为常数项， $\beta_{i,1}$ 是宏微观大数据 $Z_{i,t-1}$ 对股市系统性风险 $\beta_{i,t}$ 的影响系数。

考虑宏观经济指标 X_{t-1} 和每个企业的微观特征指标 $C_{i,t-1}$ 对 $\beta_{i,t}$ 的共同作用，所以设定预测变量 $Z_{i,t-1}$ 为宏微观变量的交叉乘积项

$$Z_{i,t-1} = X_{t-1} \otimes C_{i,t-1} \quad (4)$$

其中 $C_{i,t-1}$ 为 $P_C \times 1$ 企业微观特征矩阵， X_{t-1} 是 $P_X \times 1$ 宏观变量矩阵，最终的宏微观混合数据集 $Z_{i,t-1}$ 共有 $P_Z = P_X \times (P_C + 1)$ 个预测变量。本文中，参考了 Welch 和 Goyal^[18]，Jiang 等^[19]，姜富伟等^[20]，陈国进等^[21]研究，共考虑 $P_X = 8$ 个宏观经济指标和 $P_C = 74$ 个微观企业特征，最后得到共 $P_Z = 600$ 个宏微观混合预测变量大数据集。然后，使用机器学习这一前沿大数据建模工具对宏观微观大数据进行贝塔风险建模。

将动态 CAPM 模型(2)展开，可以得到等价的多因子资产定价模型

$$R_{i,t} = a_i^c + \beta_{i,0}R_{m,t} + \beta_{i,1}f_t(Z_{i,t-1})R_{m,t} + \mu_{i,t} \quad (5)$$

其中 $f_t(Z_{i,t-1})$ 是基于机器学习算法对宏大数据 $Z_{i,t-1}$ 进行信息提取得到线性或非线性函数。 $f_t(Z_{i,t-1})R_{m,t}$ 是条件信息 $f_t(Z_{i,t-1})$ 与市场组合收益 $R_{m,t}$ 的乘积，它的经济学含义是一个动态管理的投资组合。式(5)可以理解成使用基于 $R_{m,t}$ 和 $f_t(Z_{i,t-1})R_{m,t}$ 的高维多因子资产定价模型对个股超额收益 $R_{i,t}$ 进行精准预测的问题。

机器学习建模过程要通过划分训练集和验证集进行参数选择。按照机器学习文献的惯例，本文使用较远年限数据作为训练集拟合模型，然后使用较近年限数据作为验证集进行参数迭代并选择最优值，最终将最优参数代入当期各股变量计算，即可得到各机器学习算法模型对应各股的月度时变贝塔值。

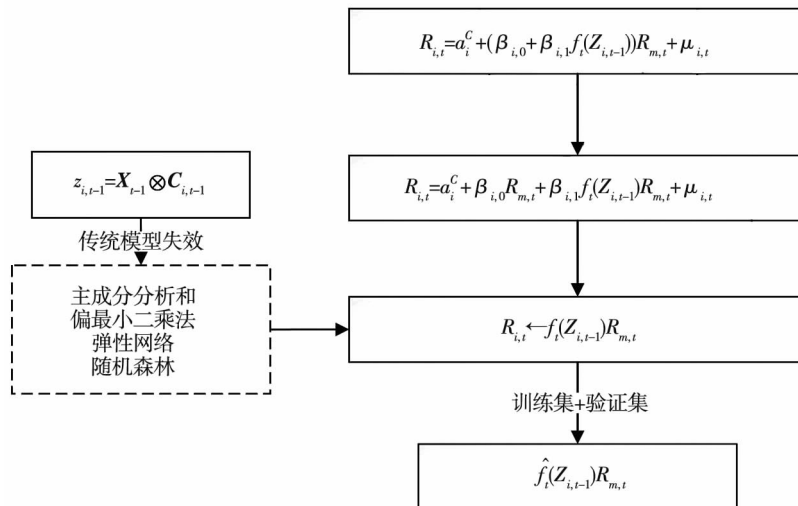


图 1 基于机器学习算法的动态 CAPM 模型建模流程

Fig. 1 Conditional CAPM modeling process based on machine learning algorithm

下文将简要介绍各类机器学习算法. 首先介绍主成分分析和偏最小二乘法模型建模; 接着介绍带惩罚函数的 LASSO、岭函数和弹性网络模型; 最后引入非线性模型即随机森林模型.

1.2.1 主成分回归和偏最小二乘法回归

主成分分析是考察多个变量间相关性的统计方法. 多变量之间的相关性一般可以认为变量反映的信息有一定的重叠. 而主成分分析能使用尽可能少的变量最大化提取变量里的公共信息. 当考虑将原变量的维度降低为 K 维时, 新的变量表示可为

$$Z_{i,t-1} R_{m,t} \times \Omega_K \tag{6}$$

其中 Ω_K 为 $P \times K$ 维矩阵, P 为原变量维度即 600 维, K 为新设定的降维维度 $K = 1, 2, 3 \dots$, 矩阵中每一列 ω_j 为构造新维度的线性权重即特征向量, 目的是找出变量间方差最大的坐标来进行降维, 且不同特征变量间保持正交消除重复信息, 目标函数表示为

$$\begin{aligned} w_{j,t} = \arg \max \text{Var}(Z_{i,t-1} R_{m,t} w_t), \\ \text{s. t. } w_t' w_t = 1, \text{Cov}((Z_{i,t-1} R_{m,t}) w_t, \\ (Z_{i,t-1} R_{m,t}) w_{t,l}) = 0 \\ l = 1, 2, \dots, j - 1 \end{aligned} \tag{7}$$

通过构建训练集和验证集进行调参: 使用较远年限 $t \sim 60$ 到 $t \sim 25$ 共 36 个月度数据作为训练集, 较近年限 $t \sim 24$ 到 t 共 25 个月度数据作为验证集. 使用训练集样本根据式(7)得到特征向量矩阵 Ω_K , 经过降维处理原 600 维数据变为 K 维的新数据, 接着迭代计算得到回归模型(5)中参数 $\hat{\beta}_{i,0,k}, \hat{\beta}_{i,1,k}$ 的最优值, 将 $\hat{\beta}_{i,0,k}, \hat{\beta}_{i,1,k}$ 以及 Ω_K 代入验证集样本并参考 Gu 等^[22] 计算其预测优度

$$R^2 = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^{N_i} (r_{i,t} - \hat{r}_{i,t,k})^2}{\sum_{t=1}^T \sum_{i=1}^{N_i} r_{i,t}^2} \tag{8}$$

其中 $\hat{r}_{i,t,k} = \hat{\beta}_{i,0,k} R_{m,t} + \hat{\beta}_{i,1,k} Z_{i,t-1} R_{m,t} \Omega_K, r_{i,t}$ 为股票当期真实收益, R^2 越高表明预测精度越高. 在正常情况下, 随着 K 值的增加验证集预测优度在上升到一定程度时会出现下降. 此时认为模型可能出现过拟合的情况而中断迭代, 最优维度 $K1$ 是使得验证集预测优度最高的参数, 此时的参数值为 $\hat{\beta}_{i,0,K1}, \hat{\beta}_{i,1,K1}$, 而当期各股的时变 $\beta_{i,t} = \beta_{i,0,K1} +$

$$\beta_{i,1,K1} Z_{i,t-1} \Omega_K.$$

偏最小二乘法 (partial least square, PLS) 与主成分分析类似, 也是一种因子数据降维方法. 在选择特征向量矩阵 Ω_K 时 PLS 的目标公式为

$$\begin{aligned} w_{j,t} = \arg \max \text{Cov}^2(R_{i,t}, Z_{i,t-1} R_{m,t} w_t), \\ \text{s. t. } w_t' w_t = 1, \text{Cov}((Z_{i,t-1} R_{m,t}) w_t, \\ (Z_{i,t-1} R_{m,t}) w_{t,l}) = 0 \\ l = 1, 2, \dots, j - 1 \end{aligned} \tag{9}$$

即寻找自变量 $Z_{i,t-1} R_{m,t}$ 与因变量 $R_{i,t}$ 协方差最大的坐标来进行数据降维. 相比 PCA, PLS 更关注特征变量与所预测的 $R_{i,t}$ 之间的相关关系. 考虑到数据过度挖掘的问题, 本文将维度 K 控制在 70 以内然后进行选择.

1.2.2 LASSO、岭函数和弹性网络模型

LASSO、岭函数和弹性网络模型这三类模型类似, 均是在传统线性回归模型中引入惩罚函数, 进行变量选择和降低过拟合, 区别在于惩罚函数的形式不同

$$g(Z_{i,t-1} R_{m,t}; \theta) = (Z_{i,t-1} R_{m,t})' \theta \tag{10}$$

式(10)为线性回归模型, θ 为拟合参数即 $[\hat{\beta}_{i,0}, \hat{\beta}_{i,1}]$, 求解最优参数时的目标损失函数为

$$L(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (R_{i,t} - g(Z_{i,t-1} R_{m,t}; \theta)) \tag{11}$$

当损失函数最低时得到的 θ 即为最优参数值. 在原损失函数中加入惩罚函数 $\varphi(\theta;)$

$$\phi(\theta;) = \lambda(1 - p) \sum_{j=1}^p |\theta_j| + \frac{1}{2} \lambda p \sum_{j=1}^p \theta_j^2 \tag{12}$$

其中当 $P = 0$ 时, 模型称为 LASSO; 当 $P = 1$ 时, 模型称为岭函数; 当 P 介于 0 到 1 时, 模型称为弹性网络. λ 为惩罚因子, 用来改变惩罚函数占总损失函数的比值. 通过加入惩罚函数, 当在计算迭代损失函数时, θ_j 值会降低甚至被剔除, 从而解决变量间共线性问题, LASSO 类模型可以直接剔除冗余变量即部分参数值变为 0, 因此也被称作“变量选择”方法, 而岭函数类方法被称为“变量收缩”法, 可以使变量参数趋近于 0 值, 但不会剔除变量, 弹性网络法介于两者之间. 本文在实证中使用弹性网络算法时取平均权重设定 $P = 0.5$.

在建模时, 同样将数据分为训练集和验证集来做参数选择. 首先使用训练集样本优化损失函数 $L(\theta;) = L(\theta) + \phi(\theta;)$, 得到不同 λ 下的最优参

数值 $\theta_{j,\lambda}$ ，接着将 $\theta_{j,\lambda} = [\hat{\beta}_{i,0,\lambda}, \hat{\beta}_{i,1,\lambda}]$ 代入验证集样本并计算其预测优度，最终的最优参数 $\theta_{j,\lambda 1}$ 是使得验证集预测优度最高的参数，此时的参数值为 $\hat{\beta}_{i,0,\lambda 1}, \hat{\beta}_{i,1,\lambda 1}$ ，而当期各股的时变 $\beta_{i,t} = \beta_{i,0,\lambda 1} + \beta_{i,1,\lambda 1} Z_{i,t-1}$ 。

1.2.3 随机森林

使用随机森林和决策树来建模宏观微观大数据的非线性交互作用。由于变量间的交互效应数随着变量数增加而以几何速度增加，传统的通过将交叉项引入线性回归方程的做法在大数据面前缺乏可操作性。因而，本文引入决策树模型。决策树模型与传统线性回归模型有着完全不同的构造，例如在二叉树模型中的每一个分岔点，模型通过计算得到最优的分割因子将数据分为两部分得到下一层树，最终通过不断的分岔将具有相同属性的数据归类并进行预测。图 2 给出了一种简单的决策树结构，首先按照企业规模对数据进行一次划分，高于 0.5 的归为类别 3，低于 0.5 的再次按照账面市值比进行划分，低于 0.3 的归入类别 1，高于 0.3 的归入类别 2，而最终的预测值为归类后同类数据的均值。

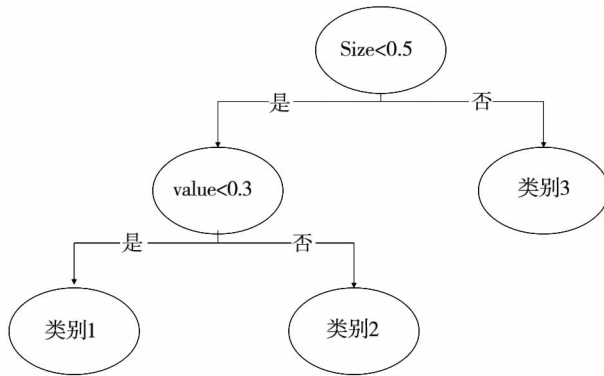


图 2 简单的决策树示例
Fig. 2 Decision tree example

随机森林模型是在简单决策树模型基础上的优化，避免陷入过度拟合。随机森林引入集成学习 (ensemble) 的概念，设定上百棵树进行计算，最终的结果综合所有树结果。最优情况下，当每棵树为互相独立时预测能力最佳，在实际中模型通过降低相关性来达到这一目的。例如对于全部数据而言企业规模是最重要的一个预测变量，那么模型在构造每棵树时会降低企业规模这一变量的使用权重，在初始几层中使用别的变量进行分岔，最终降低了树之间的关联。一个随机森林模型中需要

设定的参数包括最大特征数以及最大深度等。其中最大特征数为单个分类树模型中使用的特征变量数量，而最大深度是指单个树的最大分层数。较多的特征数和较深的深度会提高分类的精度，但同时也会导致模型过拟合，因此同之前的建模过程一致，同样使用训练集和验证集的方式，通过训练集找到在不同特征数和深度下模型的最优解，并通过验证集计算预测优度 R^2 来确定当期最优模型参数。

与本文类似的基于各类机器学习的金融经济研究文献主要偏重于收益预测和风险管理，比如 Rapach 等^[23]应用了 LASSO 方法对全球的股票市场进行了分析预测。此外，国内学者也同样进行了研究，苏治等^[24]对近年来在金融实证里应用的深度学习模型进行了文献梳理，并对未来中国金融市场中机器学习的应用给出了评述；陈卫华等^[25]基于深度学习和股票论坛数据对股市波动率进行了预测，发现深度学习模型显著提升了预测精度。上述文献都没有针对股市系统性风险进行分析，也没有使用多种机器学习算法进行对比。

2 实证研究和结果分析

2.1 样本数据

本文数据来源于国泰安和万得数据库，样本区间为 2003 年 7 月 ~ 2017 年 6 月 A 股市场所有股票收益数据、财务数据和宏观数据，考虑机器学习算法需要一定时长的训练期，实证结果区间为 2008 年 7 月 ~ 2017 年 6 月。股票收益频度为日度数据和月度数据，财务数据和宏观数据统一为月度指标。无风险收益采用月度化的一年期国债收益率，市场组合收益率使用 A 股市场所有股票收益流通市值加权平均得到。此外考虑到 A 股市场存在 ST 等特殊类股票，构建股票组合时剔除了 20% 市值最小股票，同时剔除了新发行股票前 3 个月的交易数据和金融类股票数据，组合收益为各股收益流通市值加权平均值。

2.2 中国股票市场低风险定价异象和静态 CAPM 模型

在每年 7 月初根据前 12 个月股票日度收益

按照式(13)和式(14)计算得到的各股 β 值大小排序构建 10 个股票组合^②

$$r_{i,j} = a_i + \beta_{i,0}r_{m,j} + \beta_{i,1}r_{m,j-1} + \beta_{i,2} \left[\frac{r_{m,j-2} + r_{m,j-3} + r_{m,j-4}}{3} \right] + \varepsilon_{i,j} \quad (13)$$

其中 $r_{i,j}$ 为股票 i 当期超额收益, $r_{m,j}$ 为市场组合当期超额收益, $r_{m,j-1}$ 为市场组合上期超额收益, $r_{i,j-2} + r_{i,j-3} + r_{i,j-4}$ 为市场组合滞后 2 期到滞后 4

期的超额收益之和,采用滞后项主要考虑市场信息扩散速率不一致问题,最终 β 值为由式(13)计算得到的各项参数估计值之和

$$\hat{\beta}_{i,t} = \hat{\beta}_{i,0} + \hat{\beta}_{i,1} + \hat{\beta}_{i,2} \quad (14)$$

组合构建完成后持有 12 个月直到下一年 6 月底,之后重新按照上述过程构建,最终得到 10 个组合的 2008 年 7 月 ~ 2017 年 6 月的共 1 080 个月度收益数据.

表 1 各组合收益均值和方差

Table 1 The mean and variance of each portfolio's return

| 组合 | $\bar{\beta}$ | 收益均值/% | 收益方差 | 组合 | $\bar{\beta}$ | 收益均值/% | 收益方差 |
|----|---------------|--------|-------|----|---------------|--------|-------|
| 1 | 0.52 | 0.994 | 0.070 | 6 | 1.23 | 0.746 | 0.098 |
| 2 | 0.83 | 0.915 | 0.083 | 7 | 1.31 | 0.754 | 0.098 |
| 3 | 0.96 | 1.043 | 0.091 | 8 | 1.40 | 0.575 | 0.095 |
| 4 | 1.06 | 0.833 | 0.093 | 9 | 1.52 | 0.641 | 0.105 |
| 5 | 1.15 | 1.152 | 0.095 | 10 | 1.80 | 0.498 | 0.111 |

表 1 报告了各组合流通市值加权静态贝塔 β 均值和收益均值与方差. 高 β 组合(组合 6 ~ 组合 10) 收益均低于低 β 组合(组合 1 ~ 组合 5),表明基于静态 CAPM 模型构建的风险指标无法合理解释我国股市收益,高风险股票无法获得相应高收益,且随着组合 β 增大,收益反而呈递减趋势,与 CAPM 模型预测相违背,因此我国股票市场存在低风险定价异象.

为了进一步探究 β 对组合收益影响的显著性,在样本期每个月(2008:07 ~ 2017:06)使用式(15)对上述得到的组合收益 $R_{i,t}$ 和估计的 $\hat{\beta}_{i,t}$ 进

行横截面回归

$$R_{i,t} = a_t + b_t \hat{\beta}_{i,t} + \mu_{i,t} \quad (15)$$

回归得到截距 α_t 和斜率 b_t .

图 3 报告了式(15)估计的时变截距 α_t 和斜率 b_t ,其中 α_t 均值为 0.012, t 检验 1.80,在 10% 水平内显著; b_t 均值为 -0.003 5, t 检验 -0.61. b 值为负但不显著,说明国内股票市场 CAPM 模型存在斜率过于平坦的现象, α_t 显著为正表明存在组合套利,即通过买入低 β 组合卖出高 β 组合可以长期获得超额利润.

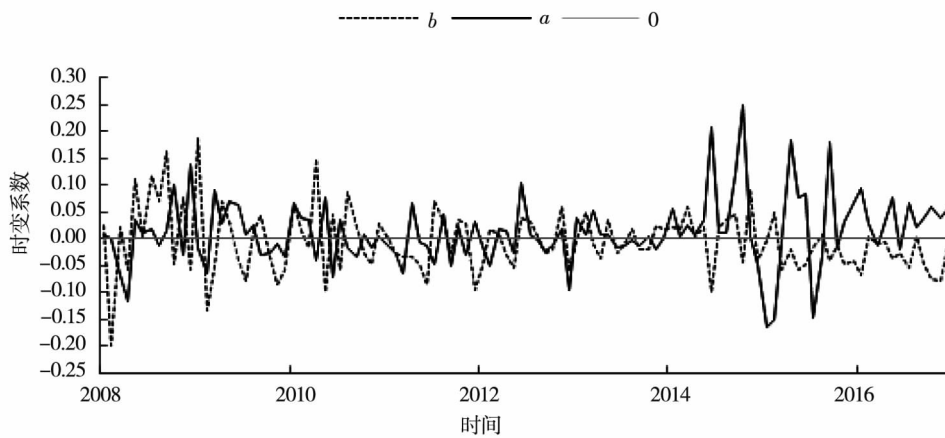


图 3 CAPM 模型时变截距和斜率

Fig. 3 Time varying intercept and slope of CAPM model

② 为保证构建的有效性,要求组合包含的每只股票每年度最少有 150 天的交易数据.

选取组合 10(H) 和组合 1(L) 时序收益, 利用式(16) 进一步分析了静态 CAPM 模型的时间序列定价表现

$$R_{i,t} = \alpha_i + \beta_i R_m + \mu_{i,t} \quad (16)$$

其中 $R_{i,t}$ 和 R_m 分别为组合 i 和市场组合的超额收益.

表 2 报告了静态 CAPM 模型下的拟合结果. H 组合样本区间内 β 估计值为 1.26, L 组合为 0.79, 高 β 组合(H 组合) 相对低 β 组合(L 组合) 有着明显的负超额收益 α , 而通过做多 H 组并卖空 L 组, 多空 HL 组合每月的超额收益为 -0.86% (年化为 -10.32%), 在水平 5% 内显著, 说明静态 CAPM 模型无法解释贝塔 β “异象”.

表 2 HL 组合静态 CAPM 模型回归拟合结果

Table 2 Regression results of HL portfolios' unconditional CAPM model

| 组合 | α (%) | β | R^2 |
|----|--------------|------------|-------|
| H | -0.49(-1.30) | 1.26(28.0) | 0.88 |
| L | 0.37(1.46) | 0.79(25.6) | 0.86 |
| HL | -0.86(-1.59) | | |

表 3 宏观经济变量的描述性统计

Table 3 Descriptive statistics of macroeconomic variables

| | 均值 | 最大值 | 最小值 | 标准差 | 偏度 | 峰度 |
|-------------|-------|-------|-------|------|-------|------|
| <i>LVOL</i> | 4.10 | 4.75 | 3.44 | 0.23 | -0.42 | 3.55 |
| <i>BM</i> | -1.02 | -0.33 | -1.76 | 0.31 | 0.06 | 2.60 |
| <i>DY</i> | -0.27 | 0.54 | -1.02 | 0.35 | 0.16 | 2.67 |
| <i>EP</i> | -4.00 | -2.96 | -5.00 | 0.40 | -0.08 | 2.34 |
| <i>CEI</i> | 1.08 | 1.27 | 0.96 | 0.48 | 0.25 | 3.98 |
| <i>INF</i> | 0.21 | 2.60 | -1.30 | 0.60 | 0.43 | 3.69 |
| <i>HJ</i> | 1.01 | 1.06 | 0.97 | 0.17 | 0.33 | 2.87 |
| <i>TO</i> | 3.08 | 8.44 | 0.44 | 1.74 | 0.90 | 3.59 |

企业特征指标依据 Jiang 等^[19] 使用企业财务数据和股票收益数据构建, 按照特征分为 6 个大类指标: 估值与成长类、投资类、盈利类、惯性类、交易摩擦类和无形资产类, 各特征统计描述见表 4.

同样选取上述 H 和 L 组合时序收益数据构建基于机器学习的动态 CAPM 模型, 起始训练期窗口从 2003 年 7 月 ~ 2006 年 6 月共 36 个月, 验证期为 2006 年 7 月 ~ 2008 年 7 月共 25 个月, 首期月度时变 β 估计月为 2008 年 7 月, 之后随着样

2.3 基于机器学习的动态 CAPM 模型

在构建动态 CAPM 模型时, 条件信息 $Z_{i,t-1}$ 包括 8 个宏观市场指标和 74 个企业特征指标, 其统计描述见表 3.

- 波动率(对数) *LVOL*: 各股月度波动率流通市值加权平均的对数值.

- 分红率(对数) *DY*: 股票红利除以滞后一期价格的对数值, 其中分红为最近 12 个月的滚动的值.

- 盈利比(对数) *EP*: 股票盈利除以当期的价格的对数值, 其中盈利为最近 12 个月的滚动的值.

- 账面市值比(对数) *BM*: 股票账面权益同当期市值的比值.

- 通货膨胀率 *INF*: 国家统计局每月给出的居民采购指数(CPI) 环比值.

- 经济景气指数 *HJ*: 根据国家统计局给出的指数除以 100 得到.

- 消费者预期指数 *CEI*: 根据国家统计局给出的指数除以 100 得到.

- 换手率 *TO*: A 股市场当月所有股票换手率算数平均值.

本外数据推移, 验证期窗口和训练期窗口滚动移动, 最终得到样本外区间为 2008 年 7 月 ~ 2017 年 6 月共 108 个月的数据.

表 5 报告了 H 组合和 L 组合在不同算法下动态 CAPM 模型下回归参数及其 t 值, 以及卖空 H 组合并买多 L 组合产生的超额收益 α 及其 t 值(表中 HL 行), 同时为了便于比较给出静态 CAPM 模型的拟合数据. 其中表中“ $f_t(Z_{i,t-1}) R_{m,t}$ ”项为各算法在式(5) 计算得到的 $\beta_{i,1} f_t(Z_{i,t-1}) R_{m,t}$ 项.

表 4 企业特征指标描述性统计

Table 4 Descriptive statistics of enterprise characteristic

| 特征因子 | 均值 | 方差 | 特征因子 | 均值 | 方差 |
|---------------------------|--------|-------|------------------------------|--------|-------|
| 资产市值比(<i>AM</i>) | -0.042 | 0.511 | 现金流负债比(<i>CFD</i>) | 0.073 | 0.930 |
| 账面市值比(<i>BM</i>) | 0.024 | 0.833 | 流动比率(<i>CR</i>) | 0.012 | 0.984 |
| 现金流股价比(<i>CFP</i>) | 0.034 | 0.975 | 流动比率增长(<i>CRG</i>) | -0.006 | 0.851 |
| 债务股本比(<i>DER</i>) | -0.055 | 0.416 | 速动比率(<i>QR</i>) | 0.008 | 0.984 |
| 长期债务股本比(<i>DLME</i>) | 0.011 | 0.997 | 速动比率增长(<i>QRG</i>) | -0.007 | 0.868 |
| 红利价格比(<i>DP</i>) | 0.006 | 0.998 | 销量现金比(<i>SC</i>) | -0.032 | 0.691 |
| 市盈率(<i>EP</i>) | 0.065 | 0.601 | 销量存货比(<i>SI</i>) | 0.000 | 1.005 |
| 债务增长(<i>LG</i>) | 0.006 | 0.884 | 交易量趋势(<i>VOLT</i>) | 0.026 | 0.990 |
| 营运现金流价格比(<i>OCFP</i>) | -0.005 | 0.786 | 交易量惯性(<i>VOLM</i>) | 0.010 | 0.966 |
| 股息股价比(<i>PY</i>) | -0.008 | 0.913 | 反转(<i>REVL</i>) | 0.035 | 1.013 |
| 可持续增长率(<i>SG</i>) | 0.019 | 0.759 | 价格(<i>PRC</i>) | -0.042 | 0.950 |
| 销量增长与存货增长差(<i>SMI</i>) | -0.004 | 0.970 | 规模(<i>SIZE</i>) | -0.005 | 0.802 |
| 销量价格比(<i>SP</i>) | 0.012 | 1.002 | 蒂姆森系数(<i>B_DIM</i>) | 0.012 | 0.966 |
| 纳税增长(<i>TG</i>) | -0.001 | 0.952 | 下行系数(<i>B_DN</i>) | 0.045 | 0.826 |
| 应计收入(<i>ACC</i>) | 0.025 | 0.821 | 法马弗伦奇系数(<i>B_FF</i>) | -0.001 | 0.660 |
| 百分比应计收入(<i>PACC</i>) | -0.003 | 0.991 | 弗兰兹尼与皮特森系数(<i>B_FP</i>) | -0.024 | 0.560 |
| 资本开销增长率(<i>CAPEG</i>) | 0.018 | 1.037 | 洪与斯拉尔系数(<i>B_HS</i>) | 0.031 | 0.958 |
| 股东权益变化(<i>dBe</i>) | 0.041 | 0.836 | 人民币交易量(<i>RVOL</i>) | 0.053 | 0.936 |
| 固定资产与存货变化率(<i>dPIA</i>) | 0.061 | 0.821 | 6个月惯性变化(<i>CHMOM</i>) | 0.000 | 0.875 |
| 投资资产比(<i>IA</i>) | 0.031 | 0.904 | 换手率(<i>TURN</i>) | 0.005 | 1.096 |
| 存货变化率(<i>IVC</i>) | 0.039 | 0.989 | 最大日收益(<i>MAXRET</i>) | -0.044 | 0.608 |
| 存货增长率(<i>IVG</i>) | 0.003 | 0.990 | 收益率波动率(<i>RETVOL</i>) | -0.053 | 0.625 |
| 净运营资本(<i>NOA</i>) | 0.063 | 0.596 | 人民币交易量波动率(<i>STD_RVOL</i>) | -0.025 | 0.918 |
| 资产换手率(<i>ATO</i>) | 0.050 | 0.999 | 换手率波动率(<i>STD_TURN</i>) | 0.006 | 1.099 |
| 现金流资产比(<i>CFOA</i>) | 0.052 | 0.718 | 非流动性(<i>ILLIQ</i>) | -0.061 | 0.619 |
| 现金生产率(<i>CP</i>) | -0.020 | 0.318 | 零交易天数(<i>ZEROTRADE</i>) | -0.023 | 0.811 |
| 现金资产比(<i>CTA</i>) | 0.039 | 0.967 | 系统性风险系数(<i>BETA</i>) | -0.019 | 0.919 |
| 资本换手率(<i>CTO</i>) | 0.028 | 0.970 | 系统性风险系数方差(<i>BETASQ</i>) | -0.024 | 0.871 |
| 息税前收益(<i>EBIT</i>) | -0.001 | 0.895 | 延迟股价(<i>PRCDEL</i>) | -0.018 | 0.917 |
| 企业收益率(<i>EY</i>) | 0.102 | 0.787 | 特质波动率(<i>IVOL</i>) | -0.058 | 0.888 |
| 边际毛利(<i>GM</i>) | 0.045 | 0.892 | 毛利率(<i>GP</i>) | 0.004 | 1.012 |
| 净利润(<i>NPOP</i>) | 0.002 | 0.914 | 资产收益率(<i>ROA</i>) | 0.001 | 1.005 |
| 净运营资产收益(<i>RNA</i>) | -0.001 | 0.948 | 股权收益率(<i>ROE</i>) | -0.001 | 1.002 |
| 投资型资产收益率(<i>ROIC</i>) | -0.003 | 0.812 | 1个月惯性(<i>MOM1</i>) | 0.019 | 0.152 |
| 应税所得与账面资产比(<i>TBI</i>) | 0.012 | 1.013 | 6个月惯性(<i>MOM6</i>) | 0.095 | 0.344 |
| Z评分(<i>Z</i>) | 0.026 | 1.049 | 12个月惯性(<i>MOM12</i>) | 0.213 | 0.541 |
| 公司年龄(<i>AGE</i>) | -0.030 | 0.988 | 36个月惯性(<i>MOM36</i>) | 0.658 | 0.934 |

表 5 静态和基于机器学习动态 CAPM 模型比较

Table 5 Comparison of unconditional and conditional CAPM models based on machine learning

| 模型 | | α (%) | p | $R_{m,t}$ | $f_t(Z_{i,t-1})R_{m,t}$ | R^2 |
|--------|----|--------------|-------|------------|-------------------------|-------|
| 静态模型 | H | -0.49(-1.30) | | 1.26(28.0) | | 0.88 |
| | L | 0.37(1.46) | | 0.79(25.6) | | 0.86 |
| | HL | -0.86(-1.59) | n/a | | | |
| PCA | H | -0.55(-1.37) | | 1.20(8.65) | 0.05(0.44) | 0.88 |
| | L | -0.22(-0.88) | | 0.60(14.1) | 0.40(5.89) | 0.89 |
| | HL | -0.33(-0.68) | 0.001 | | | |
| PLS | H | 0.00(0.00) | | 0.58(6.32) | 0.57(8.02) | 0.92 |
| | L | 0.21(0.85) | | 0.67(13.9) | 0.22(3.30) | 0.87 |
| | HL | -0.21(-0.47) | 0.004 | | | |
| EN + H | H | -0.26(-3.03) | | 0.01(0.54) | 0.99(44.5) | 0.99 |
| | L | -0.19(-2.21) | | 0.05(1.69) | 0.95(30.2) | 0.98 |
| | HL | -0.07(-0.75) | 0.06 | | | |
| RF | H | 0.18(0.80) | | 0.23(3.00) | 0.97(14.0) | 0.95 |
| | L | 0.15(0.57) | | 0.69(12.9) | 0.20(2.35) | 0.87 |
| | HL | 0.03(0.06) | 0.002 | | | |

首先,表 5 显示基于主成分分析(PCA)和偏最小二乘法(PLS)的动态 CAPM 模型的 HL 组合月度超额收益为不显著,分别为 -0.33% 和 -0.21% (t 值 -0.88 和 -0.47)。其中 PLS 模型 H 组合超额收益由 -0.39% 变为 0.00% , L 组合由 0.37% 变为 0.21% ,模型对 H 组合 β 改善要高于 L 组合,HL 组合超额收益减少了近 80% 。还使用单边显著性检验比较各动态模型多空组合超额收益与静态模型多空组合超额收益的大小,并在表 5 中给出了单边检验的 p 值。PCA 和 PLS 模型 p 值为小于 0.001 和 0.004 ,在 1% 显著水平内显著小于静态模型。

其次带有惩罚项的弹性网络算法(EN)构建的动态模型 HL 组合月度超额收益为 -0.07% (t

值 -0.75), p 值 0.06 ,模型拟合优度 R^2 为 0.99 和 0.98 ,相比静态模型提升明显。

最后报告了随机森林(RF)模型分析结果。模型的 HL 组合月度超额收益变为不显著的 0.03% (t 值 0.06),相较静态模型减少近 96% (p 值为 0.002),模型表现优于上述线性模型,组合拟合优度 R^2 为 0.95 和 0.87 ,高于静态模型。综上,引入动态 CAPM 模型可以解释股票市场低风险定价异象,随机森林等非线性模型表现尤其好。

图 4 报告 4 类动态 CAPM 模型中 H 和 L 组合的 β 时序图,相比 PCA 和 PLS,弹性网络和随机森林模型 H 组合 β 波动剧烈,而 L 组合中随机森林模型 β 波动较小,整体更符合实际,其最终表现也要优于其他三类线性模型。

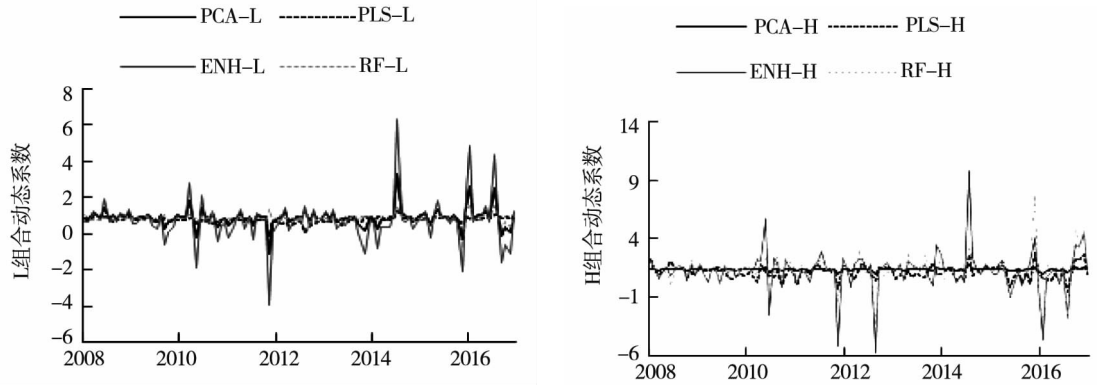


图4 动态CAPM模型中H和L组合的 β_i 时序图

Fig. 4 HL portfolios' β_i sequence diagram in conditional CAPM model

图5给出了各模型复杂程度. 主成分分析(PCA)和偏最小二乘法(PLS)给出的是在样本周期内模型降维后的维度数时序图, 维度均值分别为48和39, PCA使用信息维度高于PLS但模型表现力较差, 进一步验证了PLS优于PCA对“噪声”数据的控制力. 弹性网络模型给出的是使用的变量数的时序图, 可以发现模型使用的变量指

标在40~250之间波动, 在2012年~2015年变量数相较于其他时期明显增大, 2012年正逢我国股市大力改革时期, 市场有效性提高, 各类信息的使用率也在上升. 本文还统计了随机森林模型的模型深度时序图, 可以发现最优深度在50至32之间波动, 这说明模型深度加大会导致过拟合, 而模型过浅会影响预测效果.

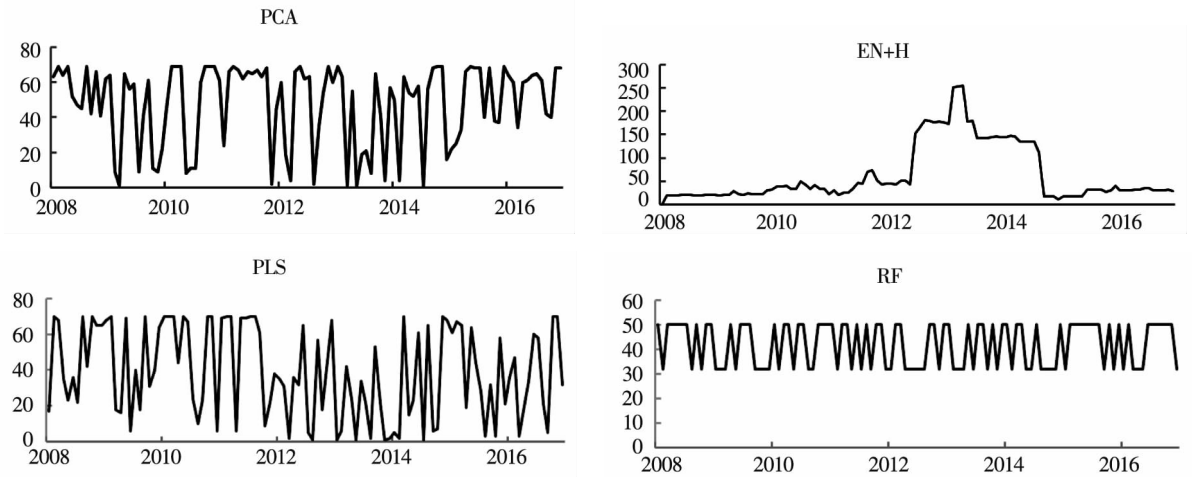


图5 各类机器学习模型复杂度比较

Fig. 5 Complexity comparison of various machine learning models

图6报告了样本周期内随机森林模型使用的各因子平均重要度^③, 横向为8类宏观因子, 纵向为74类微观因子, 交叉块颜色越深代表合成因子的重要度越高, 因子后的数字代表相比同类因子的重要度. 重要度的计算主要参考Gu等^[22], 对于任一个特征变量, 计算其剔除后模型拟合优度即

式(8)相对于未剔除此变量时的差异, 拟合优度下降越大即变量重要度越高.

本文发现, 最重要的1个宏观因子为市场换手率 TO ; 最重要的5个微观因子为一个月动量 $MOM1$, 日度换手率 $TURN$, 销售库存增长差 SMI , 日度换手率标准差 STD_TURN 和销售库存比 SI .

③ 本文计算了上述4类机器学习算法的变量重要度, 限于篇幅, 这里只给出拟合效果最好的随机森林算法的变量重要度结果.

5 个重要微观因子具体分析如下：

(1) 动量因子

动量效应作为经典的市场异象学者们对其进行了长期的研究, 主要从传统风险收益模型和行为金融两个方向, 但包括 Fama-French 因子模型在内至今没有形成统一的解释. 动量效应仍存在于众多股票市场中, 对于中国市场通常认为存在短期的动量效应和中长期的反转效应, 这与我国股票市场短期投资者如散户、游资等占比较大有关, 因而相比其他动量因子, 一个月动量 (MOM1) 具有更好的解释能力, 股票短期内持续的高收益状态内含了其高的风险状态, 因而对动态 β 具有显著的解释预测作用.

(2) 换手率

换手率 (TURN) 通常用来反映市场流动性, 换手率越高表示股票交投活跃, 流动性好, 变现能力强, 投资者更倾向于购买这类股票, 目前我国股票市场大多数股票的日换手率在 1% ~ 2.5% 之间, 长期稳定的高换手率代表了企业在股票市场的稳定性, 投资者对其要求较低的风险溢价补偿. 但频繁变动的换手率 (STD_TURN) 如突然放大往往伴随着较大的波动风险和投机风险, 这类投机倾向严重的股票包含了高的风险预期.

(3) 销售 - 库存

销售和库存的变动代表了公司的周转能力, 同行业优良的企业具有将产品快速变现的能力, 周转期的缩短增强了企业抵御风险的能力, 降低了企业风险. 同时无论是销售库存比 (SI) 还是销售与库存增长额差值 (SMI) 均反应了企业的运营状态和其基本面价值, 根据绝对定价模型可以将企业现值分解为未来现金流和要求回报率的乘

积, 企业 β 值越大, 折现率越高, 其现值越低, 进一步反映在 SI 和 SMI 因子上解释了其动态 β 的预测能力.

为了进一步研究为什么动态 CAPM 模型能解释我国股市超额收益的经济机制, 本文将静态模型超额收益与动态模型超额收益的差值即风险定价偏误进行了解. Cederburg 和 O’Doherty^[5] 认为动态 β 主要受到时变市场风险水平和时变市场波动影响, 可以分解为如下两部分

$$a_i^U - a_i^C = \left(1 + \frac{\bar{R}_{m,t}^2}{\sigma_m^2} \right) Cov(\beta_{i,t}, R_{m,t}) - \frac{\bar{R}_{m,t}^2}{\sigma_m^2} Cov(\beta_{i,t}, R_{m,t}^2) \quad (17)$$

其中 $\beta_{i,t}$ 为模型动态 β , $R_{m,t}$ 为市场组合超额收益, $\bar{R}_{m,t}$ 和 σ_m^2 分别为市场组合超额收益的无条件均值和方差. 等式右边第一部分反映了市场组合超额收益变动对收益差的解释, 而第二部分反映了市场波动率变化的影响. 按照公式, 将随机森林模型计算得到的超额收益与静态模型的超额收益进行分解来研究两类影响的大小.

表 6 给出高 β 组合 (H)、低 β 组合 (L) 和多空组合 (HL) 中市场风险和波动风险对超额收益 $\alpha_i^U - \alpha_i^C$ 的影响, 其中市场组合超额收益变动对 HL 超额收益产生负向影响 (-1.78%), 市场波动变动风险同样产生负向影响 (-0.89%) 但小于前者. 举例来说, 考虑市场收益变动的影响, 对于 H 组合当市场整体收益上升时, 静态 CAPM 模型会高估组合收益 (-1.62%), 也可解释为市场在乐观情绪下对高风险类股票产生过高的期望, 而相反会低估 L 组合的收益 (0.16%).

表 6 动态模型对超额收益的分解

Table 6 Decomposition of excess return by conditional model

| | 收益水平变动 | | 波动率变动 | | | | | | |
|----|---|---|--|---|-------|---|--------------|---|--------------|
| 组合 | $\left(1 + \frac{\bar{R}_{m,t}^2}{\sigma_m^2} \right) Cov(\beta_{i,t}, R_{m,t})$ | — | $\frac{\bar{R}_{m,t}^2}{\sigma_m^2} Cov(\beta_{i,t}, R_{m,t}^2)$ | = | Total | = | α_i^U | - | α_i^C |
| H | -1.62 | | -0.95 | | -0.67 | | -0.49 | | 0.18 |
| L | 0.16 | | -0.06 | | 0.22 | | 0.37 | | 0.15 |
| HL | -1.78 | | -0.89 | | -0.89 | | -0.86 | | 0.03 |



图 6 随机森林模型中各因子重要度比较

Fig.6 Comparison of factor importance in random forest model

3 结束语

本文使用宏微观大数据和多种机器学习算法,开展了我国股市智能动态 CAPM 模型构建研究和时变系统性风险测度研究. 研究发现,相比传统的静态 CAPM 模型,本文提出的智能动态 CAPM 模型能显著改善风险定价能力,减少定价偏误,更好地体现风险-收益的正相关关系,并从时变风险角度解释了我国股市存在的高风险低收益这一低风险定价异象. 此外,时变市场风险水平是动态 CAPM 有效性的驱动因素,换手率和短期动量等市场类因子最重要,随机森林等非线性模型最有效.

股市系统性风险测度与风险定价是资本市场研究的核心问题,CAPM 模型是最早也是最重要的资产定价模型. 本文结合大数据和机器学习的前沿进展,提高了对股票时变系统性风险的测度精度,成功解释了低风险定价异象. 有理由相信本文构建的智能动态 CAPM 模型同样可以完全解

决或部分解决目前仍存在的其他各类股票异象. 结合金融大数据和机器学习,可以更精准的建模金融市场系统性风险,这一发现对维护我国金融安全和防范化解系统性金融风险也有重要的理论和实践指导意义.

本文的政策启示为:第一,我国股票市场显著存在着高风险低收益的状况,投资者对于风险的认知不足,对于以散户作为主要投资者的市场,监管机构要加强市场风险提示,尽可能杜绝市场炒作引发的过度投机现象;第二,在使用大数据构建系统风险指标时,市场类和运营基本面类因素对系统性风险影响最大,监管者在防范金融风险时要较多留意投机类股票,关注市场资金流向,减少操纵市场的行为,构建合适的退市机制,对于常年亏损的企业实行有效的监管;第三,相比于市场波动,市场组合收益的变动会导致更大的定价偏误,监管机构在考虑股票市场风险时可以更多的关注市场收益情况,对于大幅度的市场收益变动要实施相应的风险管控举措.

参考文献:

- [1] Fama E F, French K R. The cross-section of expected stock returns[J]. *The Journal of Finance*, 1992, 47(2): 427-465.
- [2] Fama E F, French K R. The value premium and the CAPM[J]. *Journal of Finance*, 2006, 61: 2163-2185.
- [3] Frazzini A, Pedersen L H. Betting against beta[J]. *Journal of Financial Economics*, 2014, 111(1): 1-25.
- [4] Oliver B, Murray C, Adlai F, et al. Horizon effects in average returns: The role of slow information diffusion[J]. *Review of Financial Studies*, 2016, (8): 2241-2281.
- [5] Cederburg S, O'Doherty M S. Does it pay to bet against beta? On the conditional performance of the beta anomaly[J]. *Journal of Finance*, 2016, 71(2): 737-774.
- [6] 丁志国, 苏治, 赵晶. 资产系统性风险跨期时变的内生性: 由理论证明到实证检验[J]. *中国社会科学*, 2012, (4): 83-102, 206-207.
Ding Zhiguo, Su Zhi, Zhao Jing. Endogeneity of intertemporal systemic risk: From theoretical proof to empirical test[J]. *Social Sciences in China*, 2012, (4): 83-102, 206-207. (in Chinese)
- [7] 邓可斌, 关子桓, 陈彬. 宏观经济政策与股市系统性风险——宏微观混合 β 估测方法的提出与检验[J]. *经济研究*, 2018, 53(8): 68-83.
Deng Kebin, Guan Zihuan, Chen Bin. Macroeconomic policies and systemic risk in China's stock market: An approach based on integrated hybrid betas[J]. *Economic Research Journal*, 2018, 53(8): 68-83. (in Chinese)
- [8] 丁慧, 吕长江, 陈运佳. 投资者信息能力: 意见分歧与股价崩盘风险——来自社交媒体“上证 e 互动”的证据[J]. *管理世界*, 2018, 34(9): 167-177.
Ding Hui, Lv Changjiang, Chen Yunjia. Investor's information ability: Divergence of opinion and risk of stock price crash

- [J]. *Management World*, 2018, 34(9): 167–177. (in Chinese)
- [9] 杨子晖, 李东承. 我国银行系统性金融风险研究——基于“去一法”的应用分析[J]. *经济研究*, 2018, 53(8): 36–51.
- Yang Zihui, Li Dongcheng. An investigation of the systemic risk of Chinese banks: An application based on leave-one-out [J]. *Economic Research Journal*, 2018, 53(8): 36–51. (in Chinese)
- [10] 李志生, 金 凌. “国家队”救市、股价波动与异质性风险[J]. *管理科学学报*, 2019, 22(9): 67–81.
- Li Zhisheng, Jin Ling. The “national team” bailout, stock price volatility and idiosyncratic risk [J]. *Journal of Management Sciences in China*, 2019, 22(9): 67–81. (in Chinese)
- [11] Ferson W E, Siegel A F. Testing portfolio efficiency with conditioning information [J]. *Review of Financial Studies*, 2009, (22): 2735–2758.
- [12] Hansen L P, Richard S F. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models [J]. *Econometrica*, 1987, (55): 587–613.
- [13] Ravi Jagannathan, Wang Zhenyu. The conditional CAPM and the cross-section of expected returns [J]. *Journal of Finance*, 1996, (51): 3–53.
- [14] Lewellen J, Nagel S. The conditional CAPM does not explain asset-pricing anomalies [J]. *Journal of Financial Economics*, 2006, 82(2): 289–314.
- [15] Cosemans M, Frehen R, Schotman P, et al. Estimating security betas using prior information based on firm fundamentals [J]. *Review of Financial Studies*, 2016, 29(4): 1072–1112.
- [16] 尹力博, 韦 亚, 韩复龄. 中国股市异象的时变特征及影响因素研究 [J]. *中国管理科学*, 2019, 27(8): 14–25.
- Yin Libo, Wei Ya, Han Fuling. Study on characteristics and influence factors of time-varying anomalies in China’s stock market [J]. *Chinese Journal of Management Science*, 2019, 27(8): 14–25. (in Chinese)
- [17] 蒋志强, 田婧雯, 周炜星. 中国股票市场收益率的可预测性研究 [J]. *管理科学学报*, 2019, 22(4): 92–109.
- Jiang Zhiqiang, Tian Jingwen, Zhou Weixing. Return predictability in the Chinese stock markets [J]. *Journal of Management Sciences in China*, 2019, 22(4): 92–109. (in Chinese)
- [18] Welch I, Goyal A. A comprehensive look at the empirical performance of equity premium prediction [J]. *Review of Financial Studies*, 2008, 21(4): 1455–1508.
- [19] Jiang F, Tang G, Zhou G. Firm characteristics and Chinese stocks [J]. *Journal of Management Science and Engineering*, 2018, 3(4): 259–283.
- [20] 姜富伟, 徐 俊, David E, 等. 中国股票市场可预测性的实证研究 [J]. *金融研究*, 2011, (9): 107–121.
- Jiang Fuwei, Tu Jun, David E, et al. Return predictability for Chinese stock market index [J]. *Journal of Financial Research*, 2011, (9): 107–121. (in Chinese)
- [21] 陈国进, 张润泽, 赵向琴. 经济政策不确定性与股票风险特征 [J]. *管理科学学报*, 2018, 21(4): 1–27.
- Chen Guojin, Zhang Runze, Zhao Xiangqin. Economic policy uncertainty and stock risk characteristics [J]. *Journal of Management Sciences in China*, 2018, 21(4): 1–27. (in Chinese)
- [22] Gu S, Kelly B T, Xiu D. Empirical asset pricing via machine learning [J]. *Review of Financial Studies*, 2020, (33): 2223–2273.
- [23] Rapach D E, Strauss J K, Zhou G. International stock return predictability: What is the role of the united states? [J]. *Journal of Finance*, 2013, 68(4): 1633–1662.
- [24] 苏 治, 卢 曼, 李德轩. 深度学习的金融实证应用: 动态、贡献与展望 [J]. *金融研究*, 2017, 443(5): 111–126.
- Su Zhi, Lu Man, Li Dexuan. Deep learning in financial empirical applications: Dynamics, contributions and prospects [J]. *Journal of Financial Research*, 2017, 443(5): 111–126. (in Chinese)

[25] 陈卫华, 徐国祥. 基于深度学习和股票论坛数据的股市波动率预测精度研究[J]. 管理世界, 2018, 34(1): 180 - 181.

Chen Weihua, Xu Guoxiang. Research on forecasting accuracy of stock market volatility based on deep learning and stock forum data[J]. Management World, 2018, 34(1): 180 - 181. (in Chinese)

High risk low return? Explanation from machine learning based conditional CAPM model

JIANG Fu-wei, MA Tian, ZHANG Hong-wei

School of Finance, Central University of Finance and Economics, Beijing 100081, China

Abstract: This paper proposes a risk-based explanation for the low risk anomalies in Chinese stock markets, i. e. , stocks with high beta risks tend to generate surprisingly lower expected returns. A novel conditional CAPM model is constructed with big data and machine learning, in which more than 600 conditioning variables from macroeconomy and firm characteristics are employed, together with several popular machine learning methods in order to precisely modelling the time-varying beta systematic risk measures. Empirical Results show that our conditional CAPM model can fully explain away the low risk anomalies and that market-related firms characteristics are the most important predictive variables. Further, random forest featuring nonlinearity is the most effective machine learning method in modelling time-varying beta. This study contributes to the time-varying risk modelling, conditional asset pricing, and application of machine learning in finance literature.

Key words: systematic risk; conditional CAPM; machine learning; finance big data

附录

附表 1 给出本文使用的 74 个企业特征指标的名称及定义.

附表 1 公司特征指标名称及定义

Appendix Table 1 Detailed description and definition of firm characteristics

| 缩写 | 名称 | 定义 |
|----------------|--|---------------------|
| 估值与成长类指标(15 个) | | |
| <i>AM</i> | 资产市值比(assets-to-market) | 财政年总资产除以财政年末市值 |
| <i>BM</i> | 账面市值比(book-to-market equity) | 财政年权益账面价值除以财政年市值 |
| <i>CFP</i> | 现金流股价比(cash flow-to-price) | 经营现金流量除以财政年末市值 |
| <i>DER</i> | 债务股本比(debt-to-equity ratio) | 总负债除以财政年末市值 |
| <i>DLME</i> | 长期债务股本比(long term debt-to-market equity) | 长期负债除以财政年末市值 |
| <i>DP</i> | 红利价格比(dividend-to-price ratio) | 总股息支出除以财政年末市值 |
| <i>EP</i> | 市盈率(earnings-to-price) | 特殊项目之前的年收入除以财政年末市值 |
| <i>LG</i> | 债务增长(liability growth) | 总负债的年末变化除以 1 年滞后总负债 |

续附表 1

Appendix Table 1 Continue

| 缩写 | 名称 | 定义 |
|--------------|---|------------------------------------|
| <i>OCFP</i> | 营运现金流价格比 (operating cash flow-to-price) | 经营现金流量除以财政年末市值 |
| <i>PY</i> | 股息股价比 (payout yield) | 年收入减去账面资产的变化除以财政年末市值 |
| <i>Rev1</i> | 反转 (reversal) | 从 $t-60$ 到 $t-13$ 月的累计回报 |
| <i>SG</i> | 可持续增长率 (sustainable growth) | 股票账面价值的年增长率 |
| <i>SMI</i> | 销量增长与存货增长差 (sales growth minus inventory growth) | 销售额的年增长减去存货的年增长率 |
| <i>SP</i> | 销量价格比 (sales-to-price) | 年营业收入除以财政年末市值 |
| <i>TG</i> | 纳税增长 (tax growth) | 应付税款的年变化除以上一年度税款 |
| 投资类指标 (9 个) | | |
| <i>ACC</i> | 应计收入 (accruals) | 年收入减去经营现金流除以平均总资产 |
| <i>PACC</i> | 百分比应计收入 (percent accruals) | 总利润减去经营现金流除以净利润 |
| <i>CAPXG</i> | 资本开销增长率 (capital expenditure growth) | 资本支出的年变化除以滞后一年的资本支出 |
| <i>dBe</i> | 股东权益变化 (change in shareholders' equity) | 账面资产的年变化除以滞后一年的总资产 |
| <i>dPIA</i> | 固定资产与存货变化率 (changes in PPE and inventory-to-assets) | 财产、厂房和资产总值的年变化设备加上库存的年变化,除以滞后一年总资产 |
| <i>IA</i> | 投资资产比 (investment-to-assets) | 总资产的年变化除以滞后一年的总资产 |
| <i>IVC</i> | 存货变化率 (inventory change) | 库存年变化除以总资产两年均值 |
| <i>IVG</i> | 存货增长率 (inventory growth) | 库存的年变化量除以滞后一年的库存量除以总资产 |
| <i>NOA</i> | 净运营资本 (net operating assets) | 营业资产减去营业负债除以总资产 |
| 盈利类指标 (16 个) | | |
| <i>ATO</i> | 资产换手率 (asset turnover) | 销售额除以滞后一年净资产 |
| <i>CFOA</i> | 现金流资产比 (cash flow over assets) | 经营现金流除以总资产 |
| <i>CP</i> | 现金生产率 (cash productivity) | 流通股市值加上长期负债除以总资产 |
| <i>CTA</i> | 现金资产比 (cash-to-assets) | 现金和现金等价物除以两年均值总资产 |
| <i>CTO</i> | 资本换手率 (capital turnover) | 销售额除以滞后一年总资产 |
| <i>EBIT</i> | 息税前收益 (eEarnings before interests and taxes) | 净利润加所得税费用和财务费用 |
| <i>EY</i> | 企业收益率 (earnings yield) | 利息和税前利润除以企业价值 |
| <i>GM</i> | 边际毛利 (gross margins) | 营业收入减去营业费用除以滞后一年营业收入 |
| <i>GP</i> | 毛利率 (gross profitability) | 季度营业收入减去季度营业费用除以当前和上季度总资产均值 |
| <i>NPOP</i> | 净利润 (net payout over profits) | 总净支出 (净收入减去账面资产变动)除以总利润 |

续附表 1

Appendix Table 1 Continue

| 缩写 | 名称 | 定义 |
|----------------|--|---|
| <i>RNA</i> | 净运营资产收益 (return on net operating assets) | 折旧后的营业收入除以滞后一年净运营资产 |
| <i>ROA</i> | 资产收益率 (return on assets) | 季度总营业利润除以当前季度和上季度总资产均值 |
| <i>ROE</i> | 股权收益率 (return on equity) | 季度净收入除以当前季度和上季度总股东权益均值 |
| <i>ROIC</i> | 投资型资产收益率 (return on invested capital) | 利息和税收前收益减去非经营性收入除以剔除现金的企业价值 |
| <i>TBI</i> | 应税所得与账面资产比 (taxable income-to-book income) | 税前收入除以净收入 |
| <i>Z</i> | <i>Z</i> 评分 (<i>Z-score</i>) | $z\text{-score} = 1.2 \times (\text{营运资本}/\text{总资产}) + 1.4 \times (\text{留存收益}/\text{总资产}) + 3.3 \times (\text{息税前收益}/\text{总资产}) + 0.6 \times (\text{股票市值}/\text{账面价值占负债总额}) + (\text{销售}/\text{资产总额})$ |
| 惯性类指标 (7 个) | | |
| <i>CHMOM</i> | 6 个月惯性变化 (change in 6-month momentum) | $t-6$ 月到 $t-1$ 月的累积回报减去 $t-12$ 至 $t-7$ 的累积回报 |
| <i>MOM1M</i> | 1 个月动量 | 当月的累积收益 |
| <i>MOM6M</i> | 6 个月动量 | $t-6$ 到 $t-1$ 月的累积收益 |
| <i>MOM12M</i> | 12 个月动量 | $t-12$ 到 $t-1$ 月的累积收益 |
| <i>MOM36M</i> | 36 个月动量 | $t-36$ 到 $t-13$ 月的累积收益率 |
| <i>VOLM</i> | 交易量惯性 (volume momentum) | 高交易量股票从 $t-6$ 到 $t-1$ 的累积收益 |
| <i>VOLT</i> | 交易量趋势 (volume trend) | 5 年月成交量趋势除以 5 年平均成交量 |
| 交易摩擦类指标 (19 个) | | |
| <i>B_DIM</i> | 蒂姆森系数 (The Dimson Beta) | 使用市场回报的领先和滞后一期以及当前市场回报来回归计算得到的 Beta |
| <i>B_DN</i> | 下行系数 (downside Beta) | 当市场超额收益低于市场的平均超额收益时计算股票超额收益和市场超额收益之间的条件协方差除以市场超额收益的条件方差 |
| <i>BETA</i> | 系统性风险系数 (market Beta) | 以周收益率计算得到的三年滚动窗口 Beta 值 |
| <i>BETASQ</i> | 系统性风险系数方差 (Beta squared) | 上述计算得到的 Beta 平方 |
| <i>B_FF</i> | 法马弗伦奇系数 (Fama and French (1992) Beta) | 五年滚动窗口的月收益 Beta 值, 市场因子包含一期滞后项 |
| <i>B_FP</i> | 弗兰兹尼与皮特森系数 (Frazzini and Pedersen (2014) Beta) | 股票的预期回报波动率除以市场回报波动率, 再乘以回报率的相关性 |

续附表 1

Appendix Table 1 Continue

| 缩写 | 名称 | 定义 |
|------------------|---|---------------------------------------|
| <i>B_HS</i> | 洪与斯拉尔系数(Hong and Sraer (2016) Beta) | 一年滚动窗口的日度收益 Beta 值 |
| <i>IVOL</i> | 异质性收益波动率(idiosyncratic return volatility) | 三年周收益和周市场组合收益回归残差的标准差 |
| <i>ILLIQ</i> | 非流动性(illiquidity) | 过去十二个月平均绝对收益除以交易量 |
| <i>MAXRET</i> | 最大日收益(maximum daily returns) | $t-1$ 月的最大日收益 |
| <i>PRC</i> | 股价(price) | $t-1$ 月末的股价 |
| <i>PRCDEL</i> | 延迟股价(price delay) | $t-37$ 到 $t-1$ 月周收益变化率除以包含四期滞后的市场组合收益 |
| <i>RVOL</i> | 人民币交易量(RMB trading volume) | $t-2$ 到 t 月交易量乘以每股价格的对数 |
| <i>SIZE</i> | 公司规模(firm size) | 股票市值 |
| <i>STD_RVOL</i> | 人民币交易量波动率(volatility of RMB trading volume) | 人民币交易量的月标准差 |
| <i>STD_TURN</i> | 换手率波动率(volatility of turnover) | 日股票成交量的月标准差 |
| <i>RETVOL</i> | 收益率波动率(return volatility) | 两个月日收益率的标准差 |
| <i>TURN</i> | 股票换手率(share turnover) | 最近三个月的平均月交易量除以当月流通股总量 |
| <i>ZEROTRADE</i> | 零交易天数(zero trading days) | 最近一月零股交易日的成交量的加权数 |
| 无形资产类指标(8个) | | |
| <i>AGE</i> | 公司年龄(firm age) | 公司上市年数 |
| <i>CFD</i> | 现金流负债比(cash flow-to-debt) | 折旧前和非经常性项目前的收益,除以流动负债总额和滞后一年年的负债总额均值 |
| <i>CR</i> | 流动比率(current ratio) | 流动资产除以流动负债 |
| <i>CRG</i> | 流动比率增长(current ratio growth) | 流动比率的年增长率 |
| <i>QR</i> | 速动比率(quick ratio) | 流动资产减去库存除以流动负债 |
| <i>QRG</i> | 速动比率增长(quick ratio growth) | 速动比率的年增长率 |
| <i>SC</i> | 销量现金比(sales-to-cash) | 销售额除以现金和现金等价物 |
| <i>SI</i> | 销量存货比(sales-to-inventory) | 销售额除以库存总量 |