

高风险低收益？ 基于机器学习的动态CAPM模型解释

姜富伟，马甜，张宏伟

管理科学学报，2021.1

吕漫妮
2022.2.21

目录

- 引言
- 研究背景和动机
- 研究问题和结论
- 贡献
- 理论模型
- 模型设计与实证研究
- 结论

研究背景和研究动机

- CAPM 模型用市场贝塔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，学者们发现实际情况并非如此，高贝塔股票呈现较低的超额收益，这类现象统称为资产定价的“低风险定价异象”
 - 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，忽视了大数据和机器学习技术的快速发展。
-
- 构建基于大数据和机器学习的智能动态 CAPM 模型？
 - 是否可以检验时变系统性风险对低风险定价异象的解释能力？

研究问题与结论

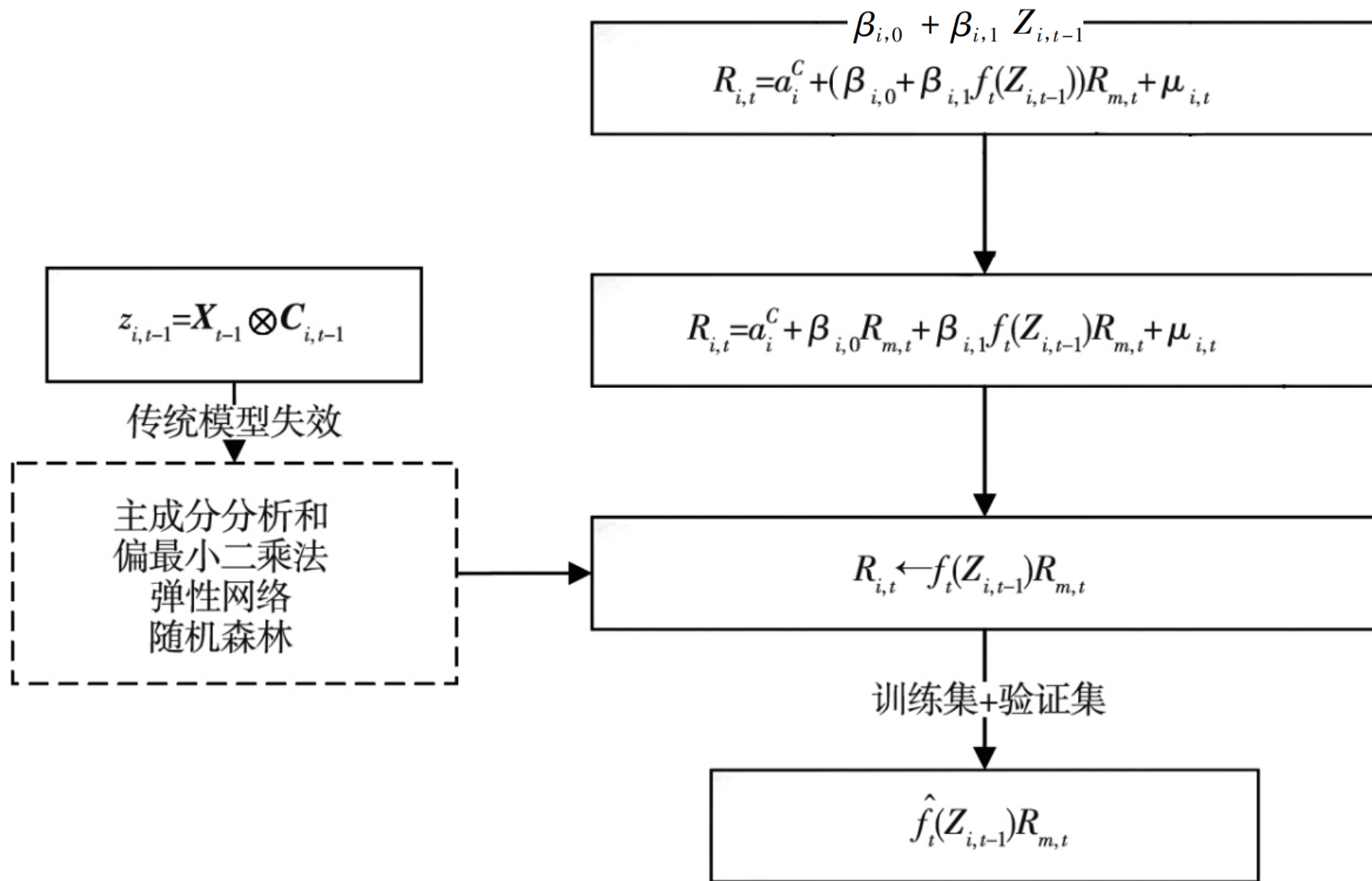
- 构建基于大数据和机器学习的智能动态 CAPM 模型
 - 使用宏微观大数据和多种机器学习算法建模
- 检验时变系统性风险对低风险定价异象的解释能力
 - 动态 CAPM 模型显著改善风险定价能力，更好地体现风险 - 收益的正相关关系，并从时变市场风险角度解释了中国股市存在的高风险低收益这一低风险定价异象
 - 模型复杂度和因子重要性

贡献

- 宏**微**观大数据估计股市系统性风险
- 使用前沿的**机器学习算法**改进股票时变贝塔建模
- 基于本文提出的智能动态CAPM 模型，**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。**

基于大数据和机器学习的动态 CAPM 模型

- 预测变量 $Z_{i,1}$ 为宏观微观变量的交叉乘积项
- 8 个宏观经济指标和 74 个微观企业特征，最后得到共 600 个宏微观混合预测变量大数据集



基于大数据和机器学习的动态 CAPM 模型

- 主成分回归：使用尽可能少的变量最大化提取变量里的公共信息

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

- 偏最小二乘法回归：关注特征变量与所预测的收益率之间的相关关系

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

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

基于大数据和机器学习的动态 CAPM 模型

求解最优参数时的目标损失函数为：
$$L(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (R_{i,t} - g(Z_{i,t-1} R_{m,t}; \theta))$$

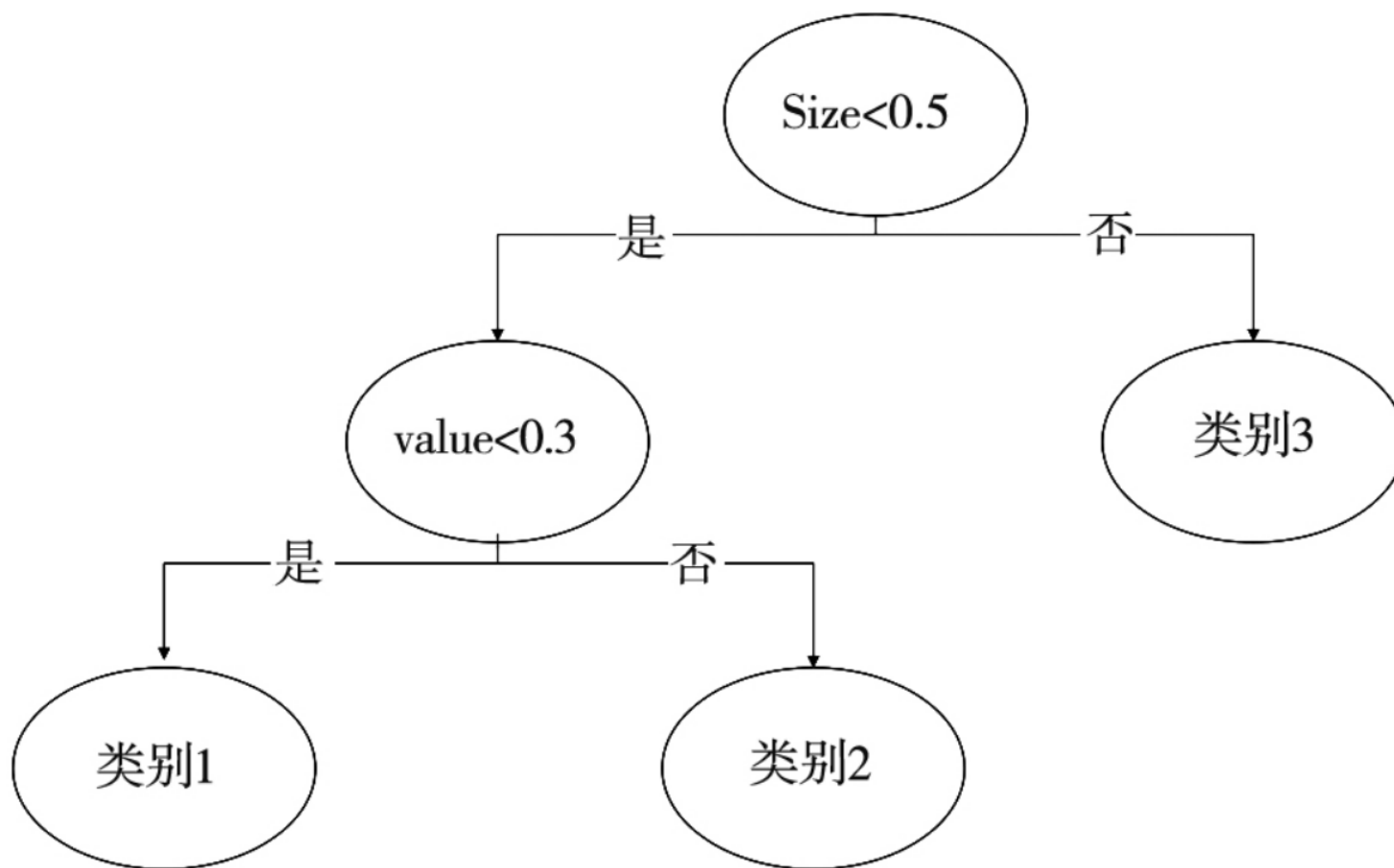
惩罚函数：
$$\phi(\theta) = \lambda(1-p) \sum_{j=1}^p |\theta_j| + \frac{1}{2} \lambda p \sum_{j=1}^p \theta_j^2$$

- LASSO回归： $p = 0$
- 岭回归： $p = 1$
- 弹性网络模型： $0 < p < 1$

λ 为惩罚因子，用来改变惩罚函数占总损失函数的比值

基于大数据和机器学习的动态 CAPM 模型

- 随机森林模型：非线性性



实证研究

样本池：2003 年 7 月 ~ 2017 年 6 月A 股市场股票，剔除 20% 市值最小股票，新发行股票前 3 个月的交易数据和金融类股票数据

• 静态CAPM模型

$$r_{i,j} = a_i + \beta_{i,0}r_{m,j} + \beta_{i,1}r_{m,j-1} + \beta_{i,2}\left[\frac{r_{m,j-2} + r_{m,j-3} + r_{m,j-4}}{3}\right] + \varepsilon_{i,j}$$
$$\hat{\beta}_{i,t} = \hat{\beta}_{i,0} + \hat{\beta}_{i,1} + \hat{\beta}_{i,2}$$

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

实证研究

- 静态CAPM模型

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

$$R_{i,t} = a_i + \beta_i R_m + \mu_{i,t}$$

- 高β组合相对低 β 组合有着明显的负超额收益α，在水平 5% 内显著，静态 CAPM 模型无法解释贝塔异象

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)		

实证研究

- **基于机器学习的动态 CAPM 模型**

条件信息包括：

Welch 和 Goyal 8 个宏观市场指标：波动率、分红率、盈利比、账面市值比、通货膨胀率、经济景气指数、消费者预期指数、市场换手率

Jiang 等 74 个企业特征指标：估值与成长类、投资类、盈利类、惯性类、交易摩擦类 and 无形资产类

起始训练期窗口：2003 年 7 月 ~ 2006 年 6 月共 36 个月

验证期：2006 年 7 月 ~ 2008 年 7 月共 25 个月

逐月向前滚动，样本外区间为 2008 年 7 月 ~ 2017 年 6 月共 108 个月

实证研究

- 机器学习动态CAPM模型

$$ret_{H,t} = \alpha_H + \beta_H R_{m,t} + \mu_{H,t}$$

$$ret_{L,t} = \alpha_L + \beta_L R_{m,t} + \mu_{L,t}$$

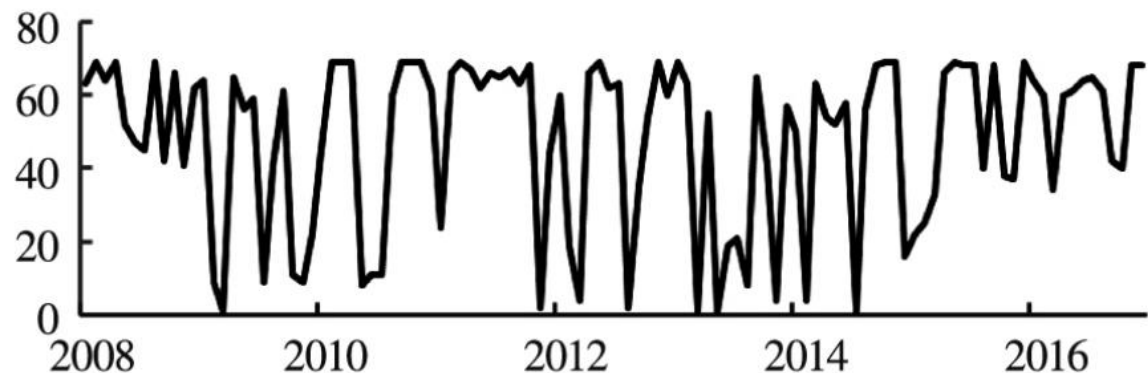
- 引入动态 CAPM 模型可以解释股票市场低风险定价异象，随机森林等非线性模型表现尤其好

模型		α (%)	p	$R_{m,t}$	$f_t \left(Z_{i,t-1} \right) R_{m,t}$	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			

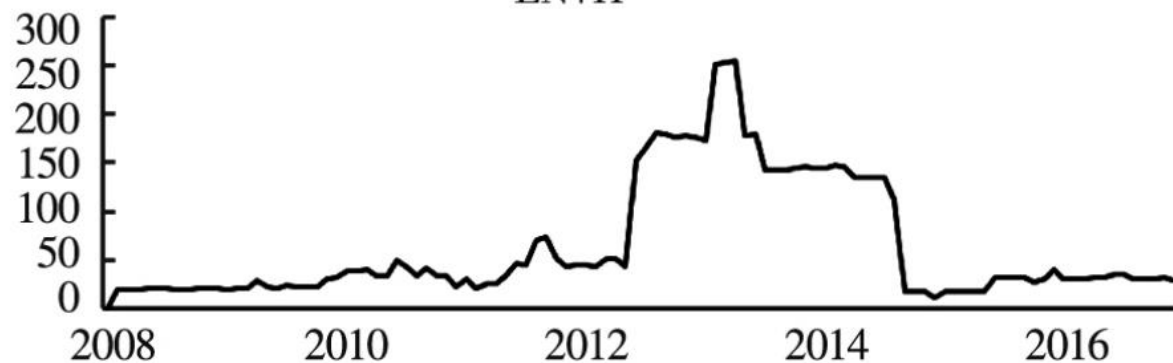
实证研究

- 动态 CAPM 模型 - 各类机器学习模型复杂度比较（维度数时序图）

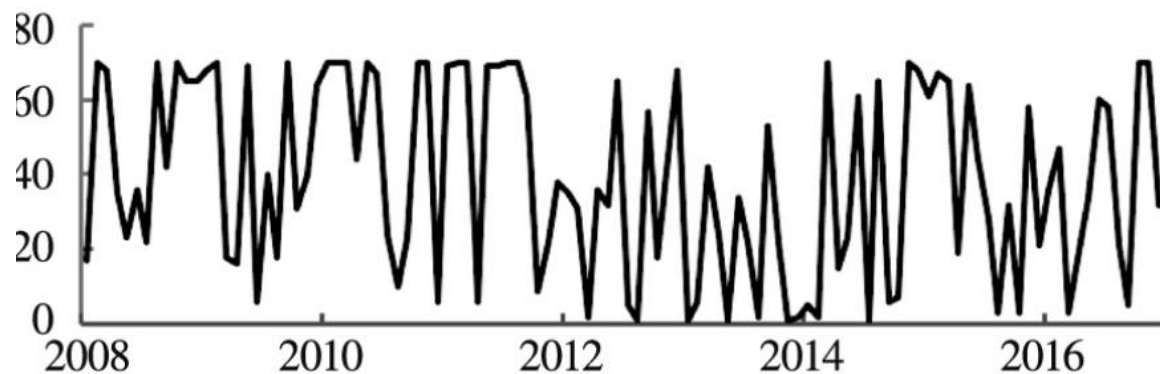
PCA



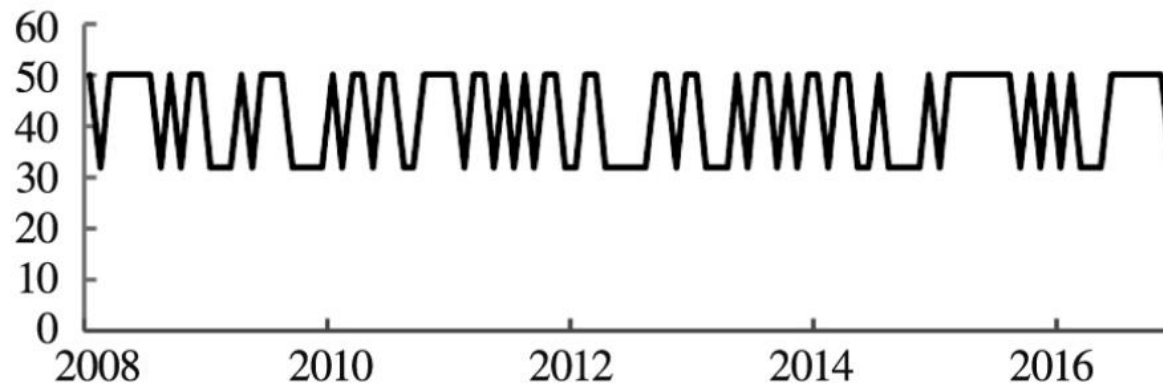
EN+H



PLS



RF



实证研究

- **动态 CAPM 模型 - 随机森林模型中各因子重要度比较 (Gu)**

最重要的 1 个宏观因子：市场换手率 TO

最重要的 5 个微观因子：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

实证研究

- 动态 CAPM 模型 - 动态模型对超额收益的分解

静态模型超额收益与动态模型超额收益的差值（Cederburg 和 O’ Doherty）

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

	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

- 市场组合超额收益变动对差值产生更大的影响. 对市场风险水平和股票收益的预期关系是导致静态模型中高风险低收益这一低风险定价异象的主要原因

结论

- 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.