现代金融与证券市场

政治关联对高新技术企业融资约束的影响

李 航

摘　要：本文的研究主要基于我国创业板和中小板高新技术企业的数据，探讨政治关联对我国创业板和中小板高新技术企业融资约束的影响及其影响机制。首先，采用现金-现金流模型进行实证研究，发现我国上市公司中的高新技术企业存在融资约束；然后将政治关联因子与企业自由现金流因子的交叉项加入到上述模型中中进行分析，实证结果表明，政治关联有助于高新技术企业缓解融资约束；最后进一步研究政治关联缓解我国中小板和创业板上市的高新技术企业的融资约束的机制，逐一分析可能的影响因素，我们发现：在我国创业板和中小板上市的高新技术企业中，相对于无政治关联的企业，有政治关联的高新技术企业更容易获得短期借款、权益融资和政府补助，并且可以通过这些渠道有效地缓解其融资约束。

关键词：高新技术企业；融资约束；政治关联

1、引言

融资约束已经成为制约中国经济转型和升级的重要瓶颈之一（邓可斌，2014），高新技术企业作为国家经济发展的重要引擎，发挥着创新先驱的作用。国内学者关于融资约束的研究一般都以主板市场为中心（饶华春，2009；卢馨，2013；王文华，2014 等），关于创业板和中小板的高新技术企业融资约束问题的研究还不是很充分。不少学者已经发现政治关联对企业融资约束的缓解作用（邓建平，2011；于蔚，2012 等），但是鲜少有学者研究政治关联是通过什么机制、什么途径来缓解融资约束的。本文运用创业板和中小板的数据研究高新技术企业是否存在融资约束，弥补了前人研究中样本的不足。同时，探讨政治关联与融资约束的关系，深入分析了政治关联缓解融资约束的途径和机制，发现政治关联企业能够通过获得更多的短期借款、权益融资、政府补助来缓解融资约束，这是本文的主要贡献。

2、文献综述与研究假设

饶华春（2009）以沪深两市非金融类上市公司作为研究对象，证明了中国上市公司存在融资约束，并分析了金融发展对融资约束的影响。卢馨、郑阳飞等（2013）以沪深 A 股市场的高新技术企业为研究对象，发现 A 股市场的高新技术企业存在融资约束、 R&D 投资强度和现金持有量之间呈显著正相关。王文华、夏丹丹等（2014）运用 A 股高新技术上市公司数据，对政府补贴缓解研发融资约束效应进行了实证研究表明，我国高新技术上市公司普遍存在研发融资约束现象，且表现出显著的研发投资现金流敏感性特征。马宏、李耿（2014）以创业板的数据研究发现我国高新技术企业融资约束问题普遍存在、社会资本对缓解企业融资约束起到了积极作用。所以本文做出假设：

假设 1：高新技术企业存在融资约束。

国外许多学者对政治关联的研究发现，许多政府官员利用职务之便为具有政治关联的企业提供方便，使这些企业受益(Claessens 等 , 2008 ; Bunkanwanicha 等 , 2008 ; Faccio , 2006 ; Fisman ,2001)。邓建平、曾勇（2011）以民营上市公司为研究对象，发现金融关联能够缓解融资约束，且政治关联较低

的民营企业中，金融关联缓解融资约束的作用更显著。罗党论、甄丽明(2008)的研究表明，有政治关联的民营企业受到的融资约束比无政治关联的民营企业更小；金融发展水平越低的地区，政治关联对民营企业缓解融资约束的帮助越明显。于蔚、汪淼军、金祥荣（2012）的研究认为政治关联确实能缓解企业融资约束，其核心机制在于信息效应和资源效应。我们参照学者们研究民营企业的经验，做出假设：

假设 2：政治关联能够缓解高新技术企业的融资约束。

于蔚、汪淼军、金祥荣（2012）的研究认为政治关联确实能缓解企业融资约束，其核心机制在于信息效应和资源效应。何镜清、李善民、周小春（2013）的研究发现民营企业家的政治关联使企业获得更高的银行贷款比率和更长的贷款期限, 民营企业家的政治关联程度越高, 政治关联在金融危机中的贷款效应越明显。王文华、夏丹丹等（2014）的研究发现政府补贴缓解研发融资的直接效应显著，间接效应不显著。借鉴这些研究，我们做出以下假设：

假设 3：有政治关联的高新技术企业通过获得更多的短期借款来缓解融资约束；

假设 4：有政治关联的高新技术企业通过获得更多的长期借款来缓解融资约束；

假设 5：有政治关联的高新技术企业通过获得更多的权益融资来缓解融资约束；

假设 6：有政治关联的高新技术企业通过获得更多的政府补助来缓解融资约束；

3、研究设计

3.1 研究变量

3.1.1 融资约束

早期 FHP（1988）以投资—现金流敏感度是否显著来评价是否存在融资约束， 且投资—现金流敏感度越大表明融资约束越高。然而 Kaplan 和 Zingales(1997)设计了 KZ 指数来衡量融资约束，得到了与 FHP 相反的结果。Alameida（2004）等在前人研究基础上提出了新的融资约束模型， 即现金-现金流模型来度量融资约束。Whited 和 Wu（2006）基于动态结构估计方法提出了度量融资约束的 WW 指数。本文采用现金-现金流模型来评价企业是否存在融资约束，因此，本文中融资约束变量指的是现金-现金流敏感度，通过本文模型中 CF 的系数来衡量，现金-现金流敏感度越高，CF 的系数越大，企业面临的融资约束越严重。

3.1.2 政治关联

本文借鉴刘慧龙、张敏等（2010）对政治关联变量的定义：若公司的董事长或总经理在中央或地方各级政府、人大、政协等部门任过职，则取值为 1，否则为 0。其中，根据职位的高低分为中央政治关联和地方政治关联；根据职位的类别分为代表型政治关联和政治型政治关联。中央政治关联代表着担任过全国人大或政协委员或者曾经在中央任职，地方政治关联代表着担任过省级人大或政协委员或者曾经在省级部门任职，代表型政治关联说明企业的董事长或 CEO 担任的是人大代表或政协委员，政治型政治关联说明董事长或 CEO 曾在政府部门任职。

3.2 研究模型

企业的融资约束程度取决于企业与资金供给者之间的信息不对称程度（郑江淮，2001），当企业面临的融资约束较大时，意味着企业难以获得外源融资，这时企业的投资将严重依赖于企业的内部资金(Hubbard, 1998)。融资约束越大，企业对内部资金的依赖越明显，企业的现金持有量对企业的现金流越敏感（连玉君，2008）。因此，本文借鉴 Alameida（2004）、姚耀军、董钢锋(2015)等的研究方法，采用现金-现金流敏感性作为衡量企业融资约束程度的指标构建了模型（1）如下：

$$DCash_{i,t}=\beta_0+\beta_1 CF_{i,t}+\beta_2 Size_{i,t}+\beta_3 DStd_{i,t}+\beta_4 DNwc_{i,t}+\beta_5 Grow_{i,t}+\beta_6 Expend_{i,t}+\beta_7 Lev_{i,t}+\beta_8 Astr_{i,t}+\beta_9 Soe_{i,t}+d_t+f_t+\varepsilon_t \quad (1)$$

为了检验政治关联是否能够缓解融资约束，我们需要把政治关联变量加入到模型中。具体做法是，在模型（1）的基础上加入政治关联和现金流的交互项 Pol*CF，用来检验有政治关联的企业对现金-现金流敏感性的影响，其余变量含义不变，作为模型（2）。

$$DCash_{i,t}=\beta_0+\beta_1 CF_{i,t}+\beta_2 POL_{i,t}\times CF_{i,t}+\beta_3 Size_{i,t}+\beta_4 DStd_{i,t}+\beta_5 DNwc_{i,t}+\beta_6 Grow_{i,t}+\beta_7 Expend_{i,t}+\beta_8 Lev_{i,t}+\beta_9 Astr_{i,t}+\beta_{10} Soe_{i,t}+d_t+f_t+\varepsilon_t \quad (2)$$

政治关联通过什么途径来缓解高新技术企业的融资约束？合理的推测是政治关联有助于企业获得外部融资（借款、权益融资、政府补助等），外部融资能力的增强毫无疑问有助于缓解融资约束，因此，我们构建了模型（3）。

$$y_{i,t}=\beta_0+\beta_1 POL_{i,t}+\beta_2 Size_{i,t-1}+\beta_3 Q_{i,t-1}+\beta_4 ROA_{i,t-1}+\beta_5 Liq_{i,t-1}+\beta_6 Lev_{i,t-1}+\beta_7 Soe_{I,t-1}+\beta_8 Tan_{i,t-1}+Industrydummy+d_t+f_t+\varepsilon_t \quad (3)$$

模型（3）也是一个双向固定效应模型，与前两个模型不同的是，它同时还控制了行业虚拟变量。被解释变量有 4 个，分别为：短期借款比率 Y1、长期借款比率 Y2、权益增加值 Y3、比率政府补助 Y4。其中，长期借款比率=（长期借款+应付债券）/年初总资产；权益增加值比率 Y3=（账面权益的净增加-留存收益的净增加值）/年初总资产。Industrydummy 是行业虚拟变量，表示用 robust 选项控制异方差。

表 1 变量定义

变量名称	变量定义
Dcash	现金持有量变动。现金及现金等价物增加额/期初总资产。
CF	现金流。经营活动现金流净额/期初总资产。
Size	企业规模。期末资产总额的自然对数。
DStd	短期负债变动。流动性负债增加额/期初总资产。
DNwc	净营运资本变动。净营运资本增加额/期初总资产。
Grow	企业成长性。主营业务收入增长率。
Expend	资本性支出。长期资产支出/期初总资产。
Lev	资产负债率。总负债/总资产。
Astr	资产结构。（固定资产+存货）/期初总资产
Soe	控股股东性质。哑变量，若控股股东为国有则取值 1，否则取 0
d	表示时间效应
f	表示企业个体效应
Pol	政治关联，公司的董事长或总经理是否担任过人大代表或者政协委员或者曾经在政府部门任职，担任过取 1，否则取 0
Pol_cen	中央政治关联，担任的职位是中央级的，担任过取 1，否则取 0
Pol_loc	地方政治关联，担任的职位是地方级的，担任过取 1，否则取 0
Q	等于公司股东权益的市场价值与净债务之和除以年初总资产
ROA	总资产报酬率或总资产收益率
Liq	流动比率，流动负债/流动资产
Tan	非流动资产/年初总资产

4、实证结果与分析

4.1 研究样本与描述性统计

本文的研究样本选取的是中小板和创业板 2004-2012 年所有的高新技术企业，其中剔除了：金融类企业，*ST、ST 与 PT 类公司，资产负债率大于 1 即资不抵债的公司，发行了 B 股或 H 股的公司，CF 值和 Expend 大于 1 的异常值，主营业务收入率超过 100%的公司，上市当年的数据，短期借款、长期借款、权益融资为负或大于总资产的数据。

样本的描述性统计和主要变量的 Pearson 相关系数见表 2 和表 3。

表 2 描述性统计

Panel A:高新技术企业政治关联状况的分年度描述性统计

年度	样本总数	政治关联企业数量	无政治关联企业数量	政治关联企业占比（%）
2007	21	14	7	66.67
2008	86	60	26	69.77
2009	121	91	30	75.21
2010	180	123	57	68.33
2011	402	273	129	67.91
2012	560	369	191	65.89

Panel B:主要变量的描述性统计

变量	样本数	平均值	中间值	标准差	最小值	最大值
DCash	1900	0.0140	-0.0346	0.315	-0.756	3.751
CF	1900	0.0446	0.0450	0.0887	-0.492	0.527
Size	1900	21.01	20.94	0.686	19.20	23.89
DStd	1900	0.0647	0.0396	0.123	-0.462	0.992
DNwc	1900	0.0544	-0.0154	0.380	-0.594	4.284
Grow	1900	0.183	0.168	0.246	-0.791	0.998
Expend	1900	0.109	0.0833	0.0905	0.00115	0.764
Lev	1900	0.293	0.271	0.176	0.00752	0.798
Astr	1900	0.406	0.396	0.193	0.0151	1.362
Soe	1900	0.0725	0	0.259	0	1

表 3 主要变量的 Pearson 相关系数

	DCash	CF	Size	DStd	DNwc	Grow	Expend	Lev	Astr	Soe
DCash	1									
CF	0.163	1								
Size	0.00220	-0.0178	1							
DStd	-0.0152	-0.126	0.228	1						
DNwc	0.926	0.0735	-0.0279	-0.161	1					
Grow	0.0358	0.0127	0.0720	0.335	0.0635	1				
Expend	0.134	0.0278	0.0355	0.247	0.1210	0.160	1			
Lev	-0.0621	-0.120	0.380	0.461	-0.153	0.0574	0.0530	1		

Astr	0.147	-0.0349	0.150	0.275	0.102	0.0683	0.308	0.516	1	
Soe	-0.0191	0.0291	0.0100	-0.0094	-0.0190	0.0190	-0.0160	0.0187	-0.0050	1

4.2 回归分析

表 4 是模型（1）的回归结果，检验高新技术企业是否存在融资约束。其中，列（1）是只保留了自变量 CF 的回归结果，列（2）是剔除了自变量 CF 的回归结果，列（3）是保留全部变量的回归结果。三个回归都同时控制了时间固定效应和企业个体固定效应。从表 4 可以看出，自变量 CF 的系数都是在 1%的水平上显著为正，因此，我们可以推断：在创业板和中小板中的高新技术企业存在融资约束。假设 1 得到验证。

表 4 高新技术企业是否存在融资约束的回归结果

因变量：现金持有量变动			
解释变量	(1)	(2)	(3)
CF	0.9459*** (7.46)		0.5618*** (14.39)
Size		-0.0272 (-1.62)	-0.0137 (-0.89)
DStd		0.5007*** (15.51)	0.5008*** (16.85)
DNwc		0.8436*** (84.49)	0.8251*** (88.90)
Grow		-0.0821*** (-5.92)	-0.0961*** (-7.50)
Expend		-0.0112 (-0.26)	-0.0377 (-0.94)
Lev		0.2287*** (4.17)	0.2034*** (4.02)
Astr		-0.1103*** (-3.14)	-0.0567* (-1.74)
Soe		0.0105 (0.57)	0.0039 (0.23)
_cons	-0.1200 (-1.11)	0.5169 (1.55)	0.2036 (0.66)
Year effect	control	control	control
N	1917	1917	1917
r2	0.0925	0.9021	0.9171

"* p<0.1 ** p<0.05 *** p<0.01"

表 5 是模型（2）的回归结果，CF*POL、CF*Pol_loc、CF*Pol_cen、CF*Pol_rep、CF*Pol_gov 分别为自变量 CF 与政治关联、地方性政治关联、中央性政治关联、代表型政治关联、政治型政治关联虚拟变量的交叉项。检验政治关联能否缓解高新技术企业融资约束。从表 5 可以看出，自变量 CF*POL 的系数都是在 1%的水平下显著为负，据此可以推断：具有政治关联的高新技术企业确实能够缓解融资

约束。进一步分析可以发现，自变量 CF*Pol_loc 的系数分别在 5%和 1%的水平下显著为负，而自变量 CF*Pol_cen 的系数虽然为负，但是不显著。因此，我们得出结论：具有政治关联的高新技术企业确实能够缓解融资约束，其中，地方政治关联对缓解融资约束起到了决定性的作用，中央政治关联对缓解融资约束的影响不明显。验证了假设 2。

表 5 政治关联能否缓解高新技术企业融资约束的回归结果

因变量：现金持有量变动									
解释变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CF	1.218*** (5.10)		0.723*** (9.49)	0.701*** (9.72)	0.557*** (12.23)	0.721*** (9.47)	0.564*** (11.66)	0.543*** (12.38)	0.554*** (11.27)
CF*POL	-0.752*** (-2.64)		-0.242*** (-2.69)						
CF*Pol_loc				-0.223** (-2.55)		-0.243*** (-2.67)			
CF*Pol_cen					-0.0248 (-0.11)	-0.185 (-0.82)			
CF*Pol_rep							-0.0261 (-0.28)		-0.0429 (-0.46)
CF*Pol_gov								0.191 (1.27)	0.201 (1.32)
Size		-0.0350 (-1.58)	-0.0232 (-1.15)	-0.0223 (-1.10)	-0.0204 (-1.01)	-0.0229 (-1.13)	-0.00836 (-0.46)	-0.00892 (-0.49)	-0.00870 (-0.48)
DStd		0.441*** (11.73)	0.448*** (13.15)	0.448*** (13.14)	0.447*** (13.07)	0.448*** (13.14)	0.531*** (15.92)	0.531*** (15.94)	0.531*** (15.93)
DNwc		0.816*** (62.87)	0.796*** (66.79)	0.795*** (66.69)	0.799*** (67.06)	0.795*** (66.68)	0.815*** (75.77)	0.817*** (75.68)	0.817*** (75.60)
Grow		-0.0915*** (-5.67)	-0.107*** (-7.25)	-0.107*** (-7.25)	-0.105*** (-7.12)	-0.107*** (-7.25)	-0.0992*** (-7.06)	-0.0991*** (-7.06)	-0.0991*** (-7.05)
Expend		0.0312 (0.59)	0.00396 (0.08)	0.00249 (0.05)	-0.000676 (-0.01)	0.00355 (0.07)	-0.0656 (-1.47)	-0.0668 (-1.50)	-0.0668 (-1.50)
Lev		0.191*** (2.92)	0.124** (2.08)	0.123** (2.07)	0.137** (2.29)	0.123** (2.07)	0.175*** (3.04)	0.177*** (3.08)	0.178*** (3.09)
Astr		-0.0736* (-1.72)	0.00493 (0.13)	0.00566 (0.14)	-0.00301 (-0.08)	0.00507 (0.13)	-0.0451 (-1.26)	-0.0486 (-1.35)	-0.0487 (-1.35)
Soe		0.0206 (0.88)	-0.000130 (-0.01)	-0.000276 (-0.01)	0.00286 (0.13)	-0.000250 (-0.01)	0.0149 (0.73)	0.0132 (0.64)	0.0132 (0.64)
Yeardummy	control	control	control	control	control	control	control	control	control
_cons	0.00123 (0.02)	0.609 (1.35)	0.350 (0.85)	0.331 (0.81)	0.294 (0.71)	0.345 (0.84)	0.0798 (0.21)	0.0928 (0.25)	0.0882 (0.24)

N	1370	1370	1370	1370	1370	1370	1760	1760	1760
r2_w	0.0777	0.892	0.911	0.911	0.910	0.911	0.903	0.903	0.903

"* p<0.1 ** p<0.05 *** p<0.01"

表 6、表 7、表 8、表 9 是模型（3）的回归结果，其因变量分别为：短期借款、长期借款、权益融资和政府补助。限于篇幅，没有把所有控制变量的回归结果列出来，只列出了我们关注的几个主要变量。

表 6 的列（4）、列（5）、列（6）中，自变量 POL、Pol_loc 的系数都是在 10%的水平上显著为正，说明有政治关联的高新技术企业能够获得更多的短期借款，从而更有能力缓解融资约束，验证了假设 3。Pol_loc 的系数显著、Pol_cen 的系数不显著，说明地方性政治关联对企业短期借款的影响比中央性政府关联更明显。另外，解释变量 Pol_rep、Pol_gov 的系数都不显著。

表 7 中，我们关注的主要变量 POL、Pol_cen、Pol_loc、Pol_rep、Pol_gov 都不显著，说明政治关联对企业长期借款的影响不明显，假设 4 并不能得到验证。

表 8 中，解释变量 POL 的系数显著为正，说明有政治关联的高新技术企业能够获得更多的权益融资，从而更好的缓解融资约束，假设 5 得到验证。列（3）中，Pol_cen 的系数不显著，列（4）中 Pol_loc 的系数在 10%的水平上显著为正，列（5）中 Pol_cen 的系数在 10%的水平上显著为正、Pol_loc 的系数在 5%的水平上显著为正，说明地方政治关联对高新技术企业权益融资的影响比中央政治关联更明显。Pol_rep 的系数显著为正，而 Pol_gov 的系数不显著为负，说明与代表型政治关联相比，政治型政治关联的高新技术企业能够获得更多的权益融资。

表 9 中，列（6）中解释变量 POL 的系数显著为正，说明有政治关联的高新技术企业能够获得更多的政府补助，有利于缓解融资约束，假设 6 得到验证。Pol_loc 的系数在 5%的水平上显著为正，Pol_cen 的系数不显著，说明有地方性政治关联的高新技术企业能够获得更多的政府补助；Pol_rep 的系数不显著，Pol_gov 的系数显著为正，说明有政治型政治关联的高新技术企业能够获得更多的政府补助。

表 6 政治关联与高新技术企业短期借款关系的回归结果

因变量：短期借款比率									
解释变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
POL	0.0422** (2.50)					0.0275* (1.95)			
Pol_cen			0.0197 (0.47)		0.0461 (1.04)				
Pol_loc				0.0229* (1.70)	0.0275* (1.94)				
Pol_rep							-0.000898 (-0.10)		-0.000931 (-0.10)
Pol_gov								0.0139 (1.19)	0.0139 (1.19)
Control									

表 7 政治关联与高新技术企业长期借款关系的回归结果

因变量：长期借款比率									
解释变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
POL	-0.00547 (-0.55)					-0.0155 (-1.53)			
Pol_cen			0.00103 (0.03)		-0.0139 (-0.44)				
Pol_loc				-0.0142 (-1.46)	-0.0155 (-1.53)				
Pol_rep							0.00589 (1.04)		0.00589 (1.04)
Pol_gov								-0.00347 (-0.48)	-0.00349 (-0.48)
Control									

表 8 政治关联与高新技术企业权益融资关系的回归结果

因变量：权益融资									
解释变量	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
POL	0.0459*** (3.50)					0.0254** (2.44)			
Pol_cen			0.0265 (1.17)		0.0433* (1.82)				
Pol_loc				0.0182* (1.82)	0.0241** (2.29)				
Pol_rep							0.0164* (1.88)		0.0163* (1.88)
Pol_gov								-0.0115 (-0.88)	-0.0115 (-0.88)
Control									

表 9 政治关联与高新技术企业政府补助关系的回归结果

因变量：政府补助									
解释变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
POL	0.00374 (1.62)					0.00542** (2.18)			
Pol_cen			0.00197 (0.26)		0.00718 (0.92)				
Pol_loc				0.00471** (1.99)	0.00542** (2.18)				

Pol_rep							-0.00194 (-1.26)		-0.00193 （-1.26）
Pol_gov								0.00740*** (3.98)	0.00739*** (3.98)
Control									

从表（6）、表（7）、表（8）、表（9）可以看出：有政治关联的高新技术企业能够获得更多的短期借款、权益融资和政府补助，从而更好地缓解融资约束；地方性政治关联对融资约束的缓解作用比中央性政治关联更明显；政治型政治关联比代表型政治关联对融资约束的缓解作用更大。

4.3 稳健性检验

4.3.1 运用 SA 指数重新检验模型（1）

SA 指数由 Hadlock & Pierce(2008)提出，用来定量测度企业的相对融资约束程度，本文借鉴这一方法，用 SA 指数代替模型（1）的自变量 DCash 作为融资约束的度量指标。本文中，SA 指数的构建方法如下：

$SA=-0.737*Size+0.043*Size^2-0.04*Age$

企业规模用总资产的对数值表示，企业年龄用观测年度与企业成立或注册年度的差值表示；SA 指数为负且绝对值越大，说明企业受到的融资约束程度越严重。由于 SA 指数包含了自变量 Size 及其二次项，为了防止出现自回归，在进行稳健性检验时从模型中剔除了自变量 Size。回归结果如表 10 所示，用 SA 指数代替 DCash 作为因变量后，列（3）中自变量 CF 的系数在百分之五的显著性水平下为负，说明高新技术企业的现金流对 SA 指数具有负作用，现金流越多，融资约束越大，从而企业将持有更多的现金，这与模型（1）的结论相符合。

表 10 模型（1）的稳健性检验

	(1)	**(2)**	**(3)**
	SA	**SA**	**SA**
CF	-0.0882 (-1.03)		-0.168** (-2.05)
DStd		0.265*** (4.26)	0.264*** (4.25)
DNwc		0.124*** (6.51)	0.129*** (6.73)
Grow		0.0123 (0.46)	0.0165 (0.61)
Expend		0.311*** (3.67)	0.318*** (3.75)
Lev		0.711*** (6.79)	0.716*** (6.85)
Astr		-0.377*** (-5.58)	-0.391*** (-5.78)
Soe		0.0327	0.0345

		(0.90)	(0.95)
_cons	2.378*** (32.30)	2.322*** (30.49)	2.336*** (30.59)
Year effect	control	control	control
N	1917	1917	1917
r2	0.522	0.575	0.576

4.3.2 运用一阶差分 GMM 对模型（2）进行重新估计

本文运用一阶差分 GMM 模型对模型（2）进行检验，除 CF、Grow 两个变量之外，其他自变量均假设为严格外生。回归结果如表 11 所示，L.dcash 为因变量 DCash 的一阶滞后项，L.CF 、 L.Grow 分别为自变量 CF、Grow 的一阶滞后项，通过回归结果可以看出，列（3）中 CF*Pol 的系数、列（4）和列（6）中 CF*Pol_loc 的系数在百分之一的显著性水平下为负，同时 CF*Pol_cen 的系数仍然不显著，回归结果与模型（2）的相同。可见，模型（2）的结果是稳健的。

表 11 模型（2）的稳健性检验

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	dcash	**dcash**	**dcash**	**dcash**	**dcash**	**dcash**
L.dcash	-0.0481* (-1.96)	-0.0267** (-2.49)	-0.0198* (-1.74)	-0.0196* (-1.72)	-0.0198* (-1.72)	-0.0197* (-1.73)
CF	0.747*** (3.11)		0.676*** (6.42)	0.649*** (6.87)	0.437*** (7.60)	0.672*** (6.39)
L.CF	0.0811 (0.69)		-0.0473 (-0.88)	-0.0480 (-0.89)	-0.0518 (-0.95)	-0.0475 (-0.88)
Grow		-0.0554*** (-3.32)	-0.0594*** (-3.42)	-0.0597*** (-3.44)	-0.0562*** (-3.21)	-0.0598*** (-3.44)
L.Grow		0.0213 (1.35)	0.0280* (1.72)	0.0254 (1.57)	0.0256 (1.55)	0.0264 (1.62)
CF*Pol	-0.259 (-0.99)		-0.294*** (-2.58)			
CF*Pol_loc				-0.279*** (-2.64)		-0.301*** (-2.62)
CF*Pol_cen					0.111 (0.53)	-0.121 (-0.54)
Size		-0.0401 (-1.62)	-0.0431 (-1.63)	-0.0429 (-1.62)	-0.0409 (-1.53)	-0.0429 (-1.62)
DStd		0.495*** (12.25)	0.474*** (11.66)	0.476*** (11.70)	0.470*** (11.44)	0.476*** (11.68)
DNwc		0.751*** (28.72)	0.762*** (27.61)	0.764*** (27.66)	0.761*** (27.28)	0.763*** (27.60)
Expend		0.0541 (0.97)	-0.0730 (-1.29)	-0.0715 (-1.26)	-0.0774 (-1.35)	-0.0719 (-1.27)

Lev		0.0409 (0.59)	0.0361 (0.52)	0.0333 (0.48)	0.0388 (0.55)	0.0340 (0.49)
Astr		-0.0797* (-1.68)	-0.0678 (-1.39)	-0.0716 (-1.47)	-0.0651 (-1.32)	-0.0707 (-1.45)
Soe		0.0185 (0.83)	0.0129 (0.54)	0.0126 (0.52)	0.0161 (0.66)	0.0125 (0.52)
Year effect	control	control	control	control	control	control
_cons	-0.0415***	0.780	0.871	0.870	0.826	0.870
	(-4.08)	(1.56)	(1.61)	(1.61)	(1.51)	(1.60)
N	390	572	390	390	390	390

5、结论

本文对模型（1）的实证分析表明高新技术企业存在融资约束，结论与假设 1 相符；模型（2）的实证分析表明政治关联能够缓解高新技术企业的融资约束，结论与假设 2 相符；模型（3）的回归结果表明政治关联有助于高新技术企业获得短期借款、权益融资和政府补助，但对企业获得长期借款的帮助不明显。短期借款、权益融资、政府补助的增加毫无疑问有助于企业缓解融资约束，因此，本文得出结论：高新技术企业存在融资约束，并且有政治关联的高新技术企业可以通过获得更多的短期借款、权益融资和政府补助来缓解融资约束。

参考文献：

[1]邓可斌,曾海舰. 中国企业的融资约束:特征现象与成因检验[J]. 经济研究,2014,02:47-60+140.

[2]饶华春. 中国金融发展与企业融资约束的缓解——基于系统广义矩估计的动态面板数据分析[J]. 金融研究,2009,09:156-164.

[3]何镜清,李善民,周小春. 民营企业家的政治关联、贷款融资与公司价值[J]. 财经科学,2013,01:83-91.

[4]卢馨,郑阳飞,李建明. 融资约束对企业 R&D 投资的影响研究——来自中国高新技术上市公司的经验证据[J]. 会计研究,2013,05:51-58+96.

[5]姚耀军,董钢锋. 中小银行发展与中小企业融资约束——新结构经济学最优金融结构理论视角下的经验研究[J]. 财经研究,2014,01:105-115.

[6]王文华,夏丹丹,朱佳翔. 政府补贴缓解研发融资约束效应实证研究——来自高新技术上市公司的经验证据[J]. 科技进步与对策,2014,08:22-26.

[7]于蔚,汪淼军,金祥荣. 政治关联和融资约束:信息效应与资源效应[J]. 经济研究,2012,09:125-139.

[8]罗党论,甄丽明. 民营控制、政治关系与企业融资约束——基于中国民营上市公司的经验证据[J]. 金融研究,2008,12:164-178.

[9]郑江淮,何旭强,王华. 上市公司投资的融资约束:从股权结构角度的实证分析[J]. 金融研究,2001,11:92-99.

[10]连玉君,彭方平,苏治. 融资约束与流动性管理行为[J]. 金融研究,2010,10:158-171.

[11]Claessens, S ., Feijen, E .and Laeven, L ., Political Connections and P referential Access to Finance:

the Role of Campaign Contributions .Journal of Financial Economics, Vol .88, 2008, pp . 554-580 .

[12]Fazzari, S., R. G. Hubbard, and B. C. Petersen, 1988, “Financing Constraints and Corporate Invest-ment,”Brookings Papers on Economic Activity, 1, 141–195.

[13]Kaplan, S., and L. Zingales, 1997, “Do Financing Constraints Explain Why Investment is Correlated with Cash Flow?,”Quarterly Journal of Economics, 112, 169–216.

[14]Whited,T.,and G.Wu,2006,“Financial Constraints Risk”,Review of Financial Studies,19,531-559.

[15]Hadlock,Charles J.,and Joshua R.Pierce, 2010,“New Evidence on Measuring Financial Constaints:Moving Beyond the KZ Index”,Review of Financial Studies,14,529-544.

基于机器学习对 ALPHA 套利的研究

李鹏辉

摘　要：量化投资策略实质上可以分为投机和套利两大类。本文主要研究 alpha 套利策略，alpha 套利是指筛选出能够跑赢指数的股票组合，然后在股指期货上进行对冲，从而只获得选股能力上的超额收益，理论上该策略成功与否的主要因素是能够筛选出优质股票的选股指标的新发现，由于选股指标的有效性是随着时间不断变化的，所以本文将基于机器学习的方法不断对选股指标进行优化从而得到未来预测性能比较可靠的指标。本文主要利用全部 A 股数据和沪深 300 股指期货对量化投资策略进行构建，在运用机器学习的方法选股的同时又加入了一些择时因素构建 alpha 测策略的股票池，在股指期货上运用协整理论发现指期货的由多向空的拐点进行套保。最后将该策略形成的时间序列对市场各个市场环境下和各个因子进行回归分析，得出策略在不同时期的不同表现，以归纳出策略取胜的科学依据。

关键词：机器学习；alpha 套利；协整；套期保值

1 引言

1.1 研究背景及意义

在国外，西蒙斯创办文艺复兴科技公司主要基于数量模型构建量化投资策略，在金融危机期间仍然能取得 85%的回报率。在我国，量化策略尚处于刚刚起步，以丁鹏为代表的学术大师开创了量化投资的先河之后市场上逐渐出现各种投资策略，第一，主要以 fama 的因子模型为代表的多因子选股模型，后来随着股指期货的推出发展为 alpha 套利策略；第二，主要是基于行为金融学理论的风格轮动策略以及动量策略的研究；第三，主要是基于计量经济学的协整理论发展而来的跨品种套利；第四，是 ETF 和成分股之间的高频交易和期现套利策略；第五，以技术指标和形态为主的趋势跟踪策略。

1.2 研究内容与方法

本文主要结合 alpha 套利策略和协整关系套利策略展开研究，以 alpha 套利策略中股票池的构建因子作为选股指标，以协整理论和技术分析方法构建择时因子主要用于股指期货的对冲。

第一部分，是对文章的学术背景和现实背景进行描述，指出本文的研究意义，接下来是介绍文章内容安排，最后是总结出文章的创新点。

第二部分，是运用机器学习对多因子选股的研究，讨论支撑压力位，并且根据协整理论讨论其合理性；

第三部分，实证分析，根据第二部分的理论，构建本文新型 alpha 量化投资策略，将得到的回测结果与传统策略结果进行对比。将本文的策略的时间序列做归因分析，分析本策略的收益来源，以判断其合理性。

本文作者：李鹏辉，江西财经大学金融学院研究生　研究方向：金融工程

第四部分，对研究结果进行总结，得到本文的研究结论。

1.3 相关理论及文献综述

国内外对量化投资的研究主要包含趋势跟踪、高频交易、量化选股三个方面。其中对高频交易的研究量要远远高于其他两个角度。

在高频交易的研究中，Vince Vella(2016)[1]，考虑了 Type-2 fuzzy 系统模型的可行性并将其引入高频交易，此模型与传统模型的区别在于引入了以提高风险调整收益为目标的智能算法，研究结果显示，按照此模型进行交易结果要优于传统模型的买入并持有策略;Robert Litzenberger (2012)[2]运用能够识别高频交易交易所提供的数据集，研究表明，传统手动的交易方式相对高频存在很大的局限性，而且高频交易有利于减小日内错误定价，增加市场流动性，最终使得做市商买入卖出价差收窄。

在趋势跟踪的研究中，刘冬烨（2014）[3]年将国外 R-Breaker 策略运用于国内，将其分为趋势跟踪和反转策略，趋势策略仍然基于技术分析，反转策略将天体物理学中的对数周期性幂律模型，最终在趋势跟踪策略的研究中得出了可观收益并得到适合在国内交易的 R-Breaker-Plus 模型；Ling-xiao Cui(2016)[4]运用 DMD 模型将中国 A 股市场看作复杂动态系统进行分解，并将其演变成 A 股市场的投资策略，研究结果表明，该策略捕捉趋势能力较强，特别适用于单边市场。

在量化选股研究方面，Hudson, Richsarod(2008)[5]指出会计信息的违规披露会造成投资者对该公司股票缺乏信心，并且运用量化模型分析了这种影响，度量出了二级市场投资者对会公司计信息的反应，这为利用公司基本面信息进行 alpha 选股构成有力依据；邹运(2012)[6]基于风格轮动理论，利用风格因子选股并且分别运用遗传算法、主观赋值方法对不同风格的股票配置权重，最后将各种策略方法和基准指数进行对比得出遗传算法模型这一客观方法是最优的配置算法。

1.4 本文的创新

第一，在 alpha 套利策略的股指期货对冲端引入择时方法，使得 alpha 套利不仅能够获得选股能力带来的超额收益，而且也保留了择时能力带来的收益。从而将 alpha 策略与协整策略相结合。

第二，不同于以往的神经网路模型，本文运用机器学习算法中的决策树分类准则对股票进行分类，决策树分类能够使得分类结果快速而准确。

第三，协整理论本身至少用于两种品种以上，本文试图将协整理论运用于单一品种用以对形态上的支撑压力位置进行研究，在股指期货的压力位做空进行套期保值。

2 理论基础

2.1 量化投资策略

量化投资策略来源于金融学无风险套利思想如期现套利，后来无风险套利思想逐渐复杂化一方面发展为风险较大的协整价差套利，另一方面结合选股指标模型相结合发展为 alpha 套利。两种套利策略各有优劣。

2.2 协整关系

协整理论模型：假设两个价格序列是一阶单整，而它们的差分分别是平稳的，而且两者存在如下的长期均衡关系：其中是平稳的残差序列，可在大于其序列标准差的一定倍数时，开仓操作，等待回归时即可盈利平仓。反之则相反。

2.3 机器学习理论

机器学习理论分为有监督学习和无监督学习，主要区别是，无监督的学习是根据输入特征向量自身特性设计分类器，而有监督的学习是，分类器设计中的参数依赖于某种结果，这种结果可以是方差

最小，错误率最小，风险最小等等。

本文主要采用有监督学习。运用决策树的方法对对样本分类，以信息增益来作为判断标准。这里所说的判断，它的判断依据也即股票本身是什么类型的，要想做有监督的学习，必须先弄清楚研究的对象本身是什么，然后才能根据它各方面的属性去分类。这里还需要用到无监督的学习，这里的无监督学习相对通过多种属性去判断对象类型必须更直观更容易理解。

夏普率本身就描述了股票本身是个有投资价值的股票还是垃圾股票，股票每期的概念板块信息虽多但变化很小。对于每只股票先根据夏普率对其进行无监督学习分类即 k 均值聚类的方法，再根据概念板块特征和聚类结果进行有监督学习即决策树的方法，然后就初步选出了下期表现较好的股票池，再根据会计指标进行学习进而得到最终股票池。

2.4 本文的模型

首先利用机器学习中的决策树的方法从基本面选股，选股指标有股票所属概念板块、证券类型、ROE、主营业务比率、EBITDA/营业收入、每股经营活动产生现金流、EPS、BPS、每股股利、前十大股东持股比例、每股收益预测机构数量、预测每股收益平均、预测每股现金流、预测每股股利、预测息税前利润、预测每股净资产、预测总资产收益率、预测净利润增长率。以上属性包含公司总体类型、会计信息和预测信息三大部分。其中公司总体类型包含所属概念板块和证券类型，概念板块是十分重要的信息，策略不能包含 ST 股票，概念板块中的其他属性还可以分为几类：按行业有农业、钢铁制造业、能源化工业、金融服务业、餐饮服务业、运输业其他。会计属性有 8 钟该信息比较完善；预期信息有 8 钟该类信息空值比较多。决策树的基本思想:任意选择属性作为决策起点进行建树，一个叶节点终止的标志是该节点上所有的的实例都属于同一类，否则选取其他属性继续细分，直到算法终止。属性的选择标准是信息增益（ID3）。它的计算方法如下：

$$I(N_1, N_2, \ldots, N_m) = -\sum_{i=1}^{M} p_i \log_2(p_i) \tag{1}$$

其中是样品属于的概率，用来估计。

$$E(A) = \sum_{j=1}^{k} \left[\left(\frac{N_{1j} + N_{2j} + \cdots + N_{Mj}}{N} \right) \times I(N_{1j} + N_{2j} + \cdots + N_{Mj}) \right] \tag{2}$$

其中是系统总熵

$$I(N_{1j}, N_{2j}, \ldots, N_{mj}) = -\sum_{i=1}^{M} p_{ij} \log_2(p_{ij}) \tag{3}$$

其中，表示中的样本属于概率；表示中的样品个数。

$$Gain(A) = I(N_1, N_2, \ldots, N_m) - E(A) \tag{4}$$

选出股票池子之后，要根据预测的高价、低价和现价之差预测建仓时机。

$$d_h = F_h - \beta_h \times p \tag{5}$$

$$d_l = f_l - \beta_h \times p \tag{6}$$

其中，和分别表示残差。很明显当价格超过预测的高价时就表示被高估，这时残差项小于 0；当现价小于预期低价时价格就相当于被低估，此时为正值。

3 实证研究

3.1 样本数据

全部原始数据皆从 2005 年 1 月 1 日开始。前面两期数据作为初始拟合而省去，后面有四个期间，股票数量不足 100 只，因此策略实际的回测期间是从 2006 年 6 月 30 日至 2016 年 7 月 27 日。

3.2 策略构建

选股模型利用决策树的方法筛选，原则上对筛选出的股票不限制股票数量，因为组合投资既可以增大资金容量也可以达到分散风险的目的：

概念板块是非常稳定的变量，根据股票所属概念板块变量将股票分成 N 维向量，N 是当期概念板块总数。股票本身走势的分类依据是夏普率，因为夏普率的分布并不是正态分布，所以需要运用 K 均值聚类法将股票分为高夏普率、偏正夏普率、偏负夏普率和负值很大的夏普率等不超过 10 类。由于最终结果用于预测，所以每期需要对滞后一期的概念板块和夏普分类作为目标分类值进行拟合，这样将会损失 3 个月数据: X(i-1),Y(i-1).

与概念板块相比较，会计信息虽然更加精确，但是具有滞后性，季度末只能得到上一季度的信息，这体现在股价波动上就是每逢季度报告公布时，业绩突出的公司股价出现连续涨停，出现较高的夏普率。因此，会计信息与板块信息相差一个季度: X(i-1), Y(i).

预测信息，是具有前瞻性的信心，可以根据当期对未来的预测进行投资，但是预测信息并不像会计信息那样准确: X(i-1),Y(i-1)

每个时期根据决策树模型选出股票后，仍需要依据形态学分析每只股票价格走势，这种择时方法即形态理论，对每只股票，判断其是否探底成功，若满足成功条件则买入股票，当股票到达顶部及时出局，等得下一个底部。顶部和底部的判断依据是：前 3 至 30 个交易日的最高价为顶部不合理价格，最低价为底部不合理价格。

3.3 策略分析

Figure 2 第二轮牛市股票组合与沪深 300 指数走势对比

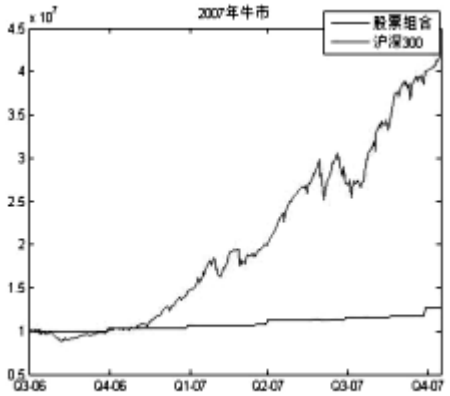


Figure 1 本区间是根据指数走势鉴定为牛市

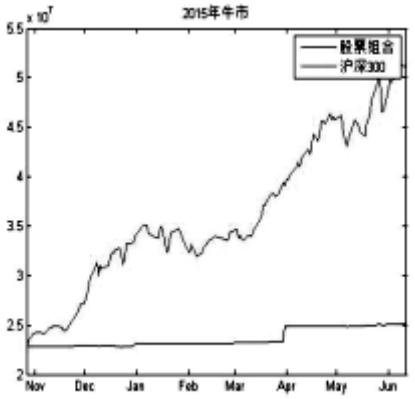


图 1 在 2007 年的牛市中本策略并未跑赢大盘，主要因为本策略不是趋势追踪策略，当价格达到很低的价位时，才考虑建仓。图 2 是从 2014 年 10 月 27 日到 2015 年 6 月 12 日的牛市，股票策略在牛市中的表现同第一轮牛市无异，虽然两个时期的牛市的信号非常强，但是策略要执行自身严格的买卖信号，否则就要产生未来数据。图 1 的牛市历时 15 个月、图 2 历时仅 8 个月，任何资产都不会永远上涨，任何商品也不会永远下跌，同时第二轮牛市的时间缩短了一半，也就是说策略的机会成本减少了一半，因为策略本身比较平稳。

图 3 是从 2007 年 12 月 16 日至 2008 年 10 月 28 日的熊市，在熊市中市场指数大幅下跌，而我们的策略依然能够保持高位，这就是反转策略的优势，无论牛市或者熊市反转总是有的，策略稳定的很大原因是 K 均值动态聚类的依据是夏普率。如图 4 所示的 2015 年 6 月 12 日至 2016 年 1 月 27 日所示股票的表现，与上个熊市的图形非常相似。由此可见当交易股票加入择时因素时，即使不通过股指期货对冲，股票的表现也可以很稳定。

图 5 是股票组合在所有时期的表现，其中绿线是沪深 300 指数对应的价格走势作为对照基准，另外三条线分别是按照相应承诺书交易股票所取得的权益曲线，即价格低于前 3 日至 30 日最低价作为定价基准，并且夏普率由负值转为正值作为买入条件。该策略，没有经过股指期货对冲但是取得了类似的结果。三条线策略走势都呈现上升趋势。

Figure 3　2008 年熊市股票组合的表现与沪深 300 指数的表现

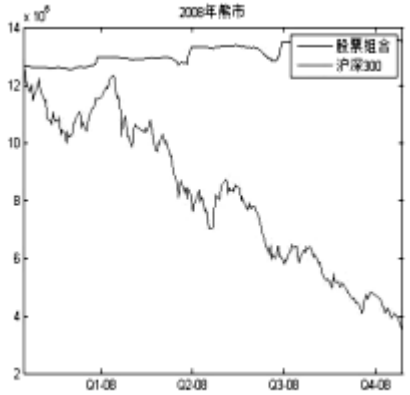


Figure 4 2015 年熊市中股票组合与沪深 300 指数表现对比图

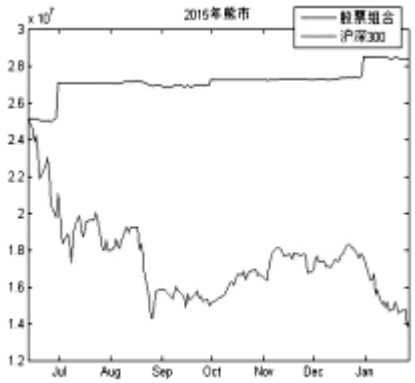


图 6 是期货市场择时做空策略在区间上的回报，沪深 300 股指期货是最早成立的，因此选此品种作为对冲工具。股指期货策略中，最敏感的 3 日压力线可以得到最优的回报，30 日压力线效果最差而且交易不多，从图形可以看到 30 日线还是可以战胜指数的。

图 7 中的蓝色线是综合期货现货的 alpha 策略表现，绿色的线是图 6 中股指期货的策略，黄色线是现货股票策略，可以看到在牛市股票策略表现的好，但是股指期货会有回撤，但在熊市，期货表现由于股票，但股票也几乎没有回撤。之所以和图一略有不同是因为，这里的基准是沪深股指期货当月合约从 2010 年 4 月开始的，图一中对应的基准是沪深 300 指数。所以从最终结果来看这样的策略是比较理想的。

Figure 6 股指期货空头择时策略

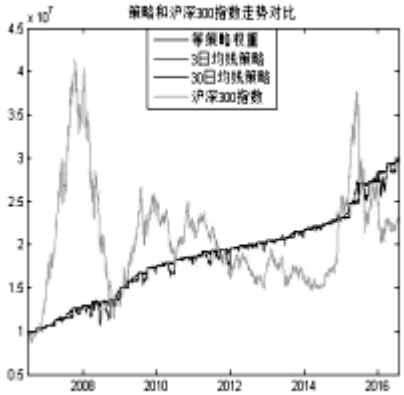


Figure 5 所有时期股票组合的表现

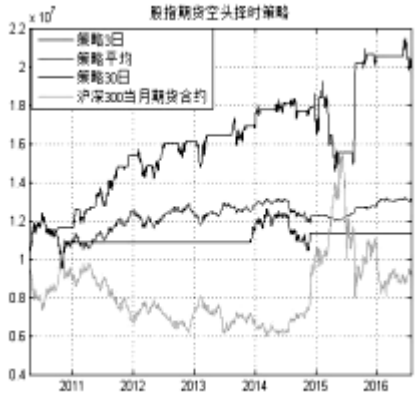
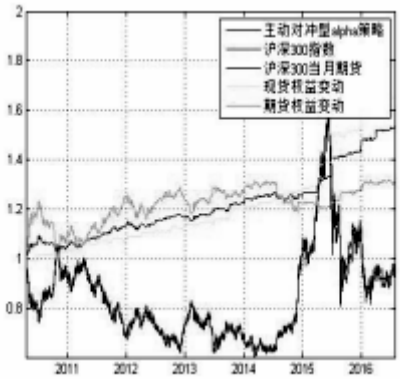


Figure 7 股票和股指期货合计的 alpha 策略



3.4 策略结果

表 1 策略收益率和回撤

策略名称	最大回撤	累计收益	年华收益率	年华夏普率
买入并持有现货	69.32%	1120.60%	25.51%	78.05%
被动持有期货空头	60.83%	5.28%	0.84%	3.22%
主动管理现货	3.97%	197.29%	11.11%	171.34%
主动管理期货	19.53%	31.96%	4.54%	50.21%
Alpha	60.83%	5.15%	0.64%	2.74%
主动 alpha	11.75%	114.63%	7.82%	110.77%

从表 1 来看，被动 alpha 策略的的现货具有高收益高回报的特点，而主动型的投资策略更加稳重。如果勉强得说被动类型的股票策略有收益率优势的话，那么主动型的期货策略就完全战胜被动型的期货策略。

表 2 策略夏普率对照表

策略名称	牛市年化夏普率	熊市年化夏普率	第二个牛市年华收益率	第二个熊市年华收益率
买入并持有现货	280.29%	-207.81%	194.95%	-62.38%
被动持有期货空头	0	0	-93.57%	189.37%
主动管理现货	188.83%	57.82%	20.94%	6.80%
主动管理期货	0	0	-4.83%	13.87%
Alpha	0.00%	0.00%	-93.57%	189.37%
主动 alpha	94.41%	28.91%	8.06%	10.33%

普通 alpha 收益只是获取了超额 alpha，而我们的策略通过协整理论还获得了残差波动的收益率。

4 结论与展望

4.1 研究结论

（1）alpha 套利的择时指标的选择即价格波动，价格波动超过前 3 天至 30 天最高价或者最低价时即可入市操作，前提是在公司基本面良好。

（2）本文通过协整理从实证的角度说明了价格回归的合理性。

（3）本文运用机器学习的方法通过基本面信息选股用来抵消基本面对本结论的影响，另一个方面说明了，股票价格波动即使不合理，但它受自身基本面特征影响较大，而股指期货往往波动较大经常出现不合理定价。

（4）普通的 alpha 模型选择避免承担系统性风险，而本文研究发现，系统性风险虽然值得回避，但是并不能完全回避。

参考文献：

[1][3]刘东烨. 股指期货日内量化投资策略[D],上海交通大学,2014.

[2][6]邹运.基于遗传算法的风格选股模型研究[D].东北财经大学,2012.

[3][1]Vince Vella. Improving risk-adjusted performance in high frequency trading using interval type-2 fuzzy logic[J]eswa,2016.01.056.

[4][2]Robert Litzenberger, Jeff Castura. The Impact of Automation and High Frequency Trading on Market Quality [J] Annual Review of Financial Economics, 2012, 4:59-98

[5][4]Ling-xiao Cui, Wen Long, Trading strategy based on dynamic mode decomposition: Tested in Chinese stock market [J] Elsevier, 2016.11.01:498-508

[6][5]Hudson, Richsarod. A qualitative study of stock market investors' perception of accounting practices and investment behavior [D] Phoenix, 2008.

基金经理投资能力与绩效的实证研究

罗 姗

摘 要：本文选取 2007 年第二季度至 2014 年第二季度的数据，采用 Fama-French 模型对基金经理的投资能力和基金绩效进行实证分析。采用 16 只具有代表性的基金，将其收益率与三因子模型进行回归分析。为了能够有效测度投资能力，也即选股能力和选时能力，故采用 TM 模型和 HM 模型对其进行测量。实证结果表明，只有为数不多的几只基金同时在选股能力和选时能力上突出，而大部分的基金则表现在选时能力上的弱处，说明这是今后基金经理需要更深入分析和探讨的。在测度出各个基金的投资能力后，可以有效的对基金投资提供参考。

关键字：Fama-French 模型；投资能力；基金绩效

1 引言

证券投资基金绩效评价的传统方法是通过测度基金的收益和风险，判断它相对已经确立的基准市场的组合是否实现了增值，是否跑赢大盘，这包括确立的适宜的基准，通过各种指标和方法计算基金的超额收益，这些方法包括夏普比率（Sharpe ratio,1966）、特雷诺比率（Treynor ratio,1965）、詹森比率（Jsense ratio,1969）、信息比率（information ratio）（Goodwin,1998）等。但是这些比率没有深入的考虑基金经理的投资能力，只是单纯的从基金的超额收益与系统性风险或者以样本方差度量的风险之间的比值而对基金的绩效下判断，这是不科学的。为了更好的测度基金经理的投资能力，应该要从选时能力和选股能力两个方面来测度，这样才能全方位的体现基金的投资价值。

2 文献回顾

基金经理的投资能力体现为能否为基金持有人获得超额资产收益，因此经理投资能力也界定为基金的选股、择时能力。在早期的研究中，Gruber(1996)认为基金经理不具备良好的择股择时能力，他以美国市场 1985 年至 1994 年间的基金样本实证地发现共同基金收益相较于市场指数平均低 65 个基点。另一方面，Grinblatt & Titman(1989)，Daniel & Titman（1997）等学者认为基金经理具有一定的择股择时能力，并且恰当择股能战胜市场。Ampomah & Clare 等(2015)发现利用传统的多因子模型会低估基金经理的管理技巧，在不同的投资风格中使用与投资风格相一致的比较基准进行分析表现出各自不同的技巧与运气。Rodríguez & Romero（2016）使用 T-M 与 H-M 模型对 1999 至 2010 期间的美国对外开放式基金进行择时分析认为一般情况下相应的地区市场不具有择时能力，而在 08 年 01 月至 10 年 12 月期间，样本基金不仅在相应的地理市场且在美国均表现出显著的择时技巧。本文的研究目的在于帮助广大投资者更好地依据基金经理投资能力，帮助其做出合理的投资决策。

本文作者：林剑修，江西财经大学信息管理学院研究生 研究方向：物流与供应链管理

3 数据与变量定义

3.1 样本及数据来源

本文选取了 16 只具有代表性的基金，包括华夏成长，博时价值，嘉实增长，长盛成长，大成价值，富国天源，易方达平稳，南方积极，金鹰中小盘，泰达宏利价值，中银中国精选，鹏华行业，招商安泰，景顺长城，广发聚富，东方龙混合型等。选取时间为 2007 年第 2 季度到 2014 年第 2 季度，由于采取日收益率时，搜集数据量较为庞大，且部分数据缺失，很难将准确的时间序列匹配，而若采取以时间为周期，则测量误差较大，不利于分析，所以选择以季度为单位，从第一次观测值到最后一次，大致有 29 个观测值，而选取了 16 只基金，所以共有 464 个观测值。采用的数据库为锐思数据库。

3.2 变量定义

1）R_i：每季度的收益率。

采集到每季度的基金单位净值 UNV 后，计算每个季度的实际收益率，计算公式如下：

$$R_{it} = \frac{UNV_t - UNV_{t\text{-}1} + D}{UNV_{t-1}} \qquad (3\text{-}1)$$

其中，UNV：基金单位资产净值

D：第 t 周的分红

2）R_f：无风险收益率

在各项理论研究中，无风险收益率通常采用 3 个月期的国库券收益率。但是在我国的国债市场的品种较少，并无短期国债发行，并且在发行国债时通常以同期银行存款利率为基准，加上一定的利率升水。因此本文采用一年期定期存款利率作为无风险利率。

3）R_m：市场基准组合

市场基准组合=（上证 A 股指数季度收益率+深成分 A 股指数季度收益率）×40%+国债季度收益率×20%

4）SMB：F-F 模型中的规模因子

规模因子即是在 F-F 模型三因子变量中的一个，它是市值小的公司的组合投资回报与市值大的公司之间的收益之差。

$$SMB = \frac{r_{sl} + r_{sm} + r_{sh}}{3} - \frac{r_{bl} + r_{bm} + r_{bh}}{3} \qquad (3\text{-}2)$$

5）HML：账面价值与规模比率高的公司与账面价值与规模比率低的公司收益率之间的差值。

$$HML = \frac{r_{\mathrm{sh}} + r_{bh}}{2} - \frac{r_{sl} + r_{bl}}{2} \qquad (3\text{-}3)$$

测量方式：将中国股票市场 A 股市值排序，规模较大的 50%的公司分配到 B 组合中，规模较小的 50%公司分配到 S 组合。后将所有上市公司按其账面价值与市值比例（book-to-market）进行排序，将比例最高的 30%放置在 H 组合中，较低的 30%放在 L 组合中，中间的 40%放在 M 组合中。根据两次排序，共建立 6 个具有不同特征的投资组合，即 B/H,B/M,B/L,S/H,S/M,S/L 六个组合。

6）MKT：超额收益

$$MKT = R_{mt} - R_{ft} \quad (3\text{-}4)$$

也就是市场资产组合的收益率与无风险利率之间的差额。

4 实证分析

为了考虑基金绩效与基金经理投资能力之间的关系，我们设定如下回归方程：

$$R_{mt}=R_{ft}+\beta_1\times mk_t+\beta_2\times smb_t+\beta_3\times hml_t+\varepsilon_{ti} \quad (4\text{-}1)$$

将 R_{ft} 左移，构成（R_{mt}-R_{ft}），也即 R_{mt}^1

$$R_{mt}^1=\alpha_p+\beta_1\times mk_t+\beta_2\times smb_t+\beta_3\times hml_t+\varepsilon_{ti} \quad (4\text{-}2)$$

故，将 16 只基金的基金收益与各个解释变量之间进行多元回归，回归结果如下：

表 1　基金回归结果

证券简称	α_p	β_1	β_2	β_3	R^2	$\overline{R^R}$	F
华夏成长	0.043	1.09	-0.64	-1.25	0.400	0.328	5.562
嘉实增长	0.038	1.037	-0.580	-0.806	0.507	0.448	8.579
长盛成长	0.031	0.904	-0.621	-1.276	0.411	0.340	5.803
大成价值	0.040	1.225	-0.483	-1.213	0.499	0.439	8.309
富国天源	0.027	0.878	-0.090	-0.262	0.441	0.374	6.568
易方达平稳	0.022	0.795	-0.564	-0.716	0.374	0.299	4.983
南方积极	0.030	1.099	-1.231	-1.475	0.472	0.409	7.46
金鹰中小盘	0.042	0.903	1.002	-0.958	0.411	0.341	5.82
泰达宏利价值	0.045	0.797	-0.015	-1.259	0.358	0.281	4.649
博时价值	0.04	1.15	-0.92	-0.81	0.554	0.501	10.368
中银中国精选	0.051	1.006	-0.508	-1.020	0.411	0.340	5.806
鹏华行业	0.045	1.055	-0.407	-0.864	0.473	0.410	7.478
招商安泰	0.028	1.062	-0.386	-0.918	0.409	0.338	5.769
景顺长城	0.033	1.196	-0.635	-0.513	0.482	0.419	7.741
广发聚富	0.041	1.049	-0.567	-1.134	0.384	0.311	5.205
东方龙混合型	0.031	1.062	-0.446	-0.242	0.409	0.338	5.774

从上述实证结果可以看出，β_1 的值一般为正，数值上大约为 1，说明市场的超额收益与基金的收益率是呈正相关的，当超额收益增长 1 个百分点时，收益率亦增长 1 个百分点左右，而其中 α_p 最大的是中银中国精选。除了金鹰中小盘之外，β_2 的系数基本上为负数，说明规模因子与收益率之间是呈负相关的，在基金投资中，需要将基金的规模因子考虑在内。在 16 只基金中，景顺成长的规模因子对基金收益率的影响最大。β_3 的值为负数，说明账面价值与规模的比率与基金收益之间呈负相关，南方积极的 β_3 最大，说明 HML 因子的影响最大。从判定系数的角度来看，判定系数基本处于 50%的水平，也

就是说收益率的一半的变异程度可以由三因子来解释，其中，最大的判定系数是博时价值为 51%，广发聚富的判定系数最小为 31.1%，一般来说，诸如基金、股票类证券市场的资产，判定系数在 50%左右已经算是很好的解释程度了。从 F-test 的检验来看，F 统计量很小，基本上不能拒绝原假设（一般取得显著性水平为 5%），说明三因子不能很好的组成该回归方程。这从 t-test 的检验来看，只有 MKT 因子显著，而 SMB、HML 因子的不能通过检验。如果一旦将后二者因子剔除出方程，那即是经典的 CAPM 模型了，而 α_p 值表达的意思为基金是否跑赢大盘。所以，从这个角度来看，给予中国国内市场的因素，F-F 三因子模型并不能很好的解释中国基金业收益率的因素，这是存在缺陷的。而从更加全面的观点来看，国外实证分析一般会将 car-hart 四因子模型予以考虑，即在 F-F 模型的基础上，加上动量模型，即是股票收益率延续原来的运动方向的趋势。由于篇幅原因，故不将包含动量模型的 car-hart 四因子模型予以实证分析。

由于博时价值的判定系数最高，故以其单独列出进行相关检验：

4.1 多重共线性检验

第一种方法是通过 Variance Inflation Factors（VIF）来检验，检验表明，当数值超过 10 时，那么就会存在多重共线性，而 MKT 的数值为 1.011，SMB 的数值为 1.085，HML 的数值为 1.092，说明三者并不存在多重共线性。

第二种方法是通过相关系数来检验。

表 2　相关系数

MKT	**SMB**	**HML**	
MKT	1.0000	0.0367	0.0842
SMB	1.0000	-0.2738	
HML	1.0000		

从相关系数的数值来看，这很明显，各因子之间是不存在多重共线性的。

4.2 异方差检验

使用怀特（white）检验：

表 3　检验结果

	coefficient	**std. error**	**t-ratio**	**p-value**	
const	0.00875801	0.00309949	2.826	0.0108	**
v3	0.0272987	0.0329215	0.8292	0.4173	
v4	0.000997317	0.0713300	0.01398	0.9890	
v5	0.0749142	0.109116	0.6866	0.5007	
sq_v3	0.298661	0.208151	0.435	0.1676	
X2_X3	0.781966	0.658645	1.187	0.2498	
X2_X4	2.27089	1.36441	1.664	0.1124	
sq_v4	1.17201	1.14277	1.026	0.3180	
X3_X4	0.890428	1.67503	0.5316	0.6012	

sq_v5	1.24930	3.42813	0.3644	0.7196

R^2= 0.543207　$N\times R^2$=15.75 >$\chi_{0.025}(29)$

说明是拒绝原假设，存在异方差。因为在存在异方差的基础上，需要对原方程进行修正，如果不加以修正，那么 OLS 的估计量将会有偏，准确度会降低。因此采用修正后的 WLS 的方程进行修正。

通过先将 OLS 残差进行平方，然后再取自然对数 $\log(\hat{u}^2)$。再将该值与观测值及其平方进行回归并得到拟合值 $\hat{g}$，取 $\hat{h}=\exp(\hat{g})$作为权数，加入至 WLS 方程中，得到修正后的方程。

表 4　WLS 的检验结果

	Coefficient	std. error	t-ratio	p-value	
const	0.0433552	0.00996112	4.352	0.0002	***
v3	1.17743	0.0815396	14.44	1.23e-013	***
v4	0.834882	0.221256	0.773	0.0009	***
v5	0.477008	0.404192	0.180	0.2490	

Sum squared resid	68.28525	S.E. of regression	1.652698
R-squared	0.983543	Adjusted R-squared	0.981568
F(3, 25)	498.0455	P-value(F)	2.06e-22
Log-likelihood	-53.56699	Akaike criterion	115.1340
Schwarz criterion	120.6032	Hannan-Quinn	116.8469
rho	0.172487	Durbin-Watson	1.590300

从上述修正后的方程来看，R^2 达到了较高的水平 0.9835，而 t-test 检验的显著性上，MKT 和 SMB 都达到了显著性水平，而 HML 仍旧不能通过。

4.3 序列相关性

第一种方式是通过 autocorrelation 来测试，首先去滞后项的项数为 4 项，得到的结果并不能够体现相关性，且从 $\hat{u}_1$ 到 $\hat{u}_4$ 的显著性都是不能通过的。因此，逐次检验滞后项，但是最终得到的项数为 0，也即纵然是在只有一个滞后项时，仍旧不能通过，也即是说不存在序列相关性，$E(U_tU_{t-1})=0$。

第二中方式是通过杜宾检验（D-W）来测试。

d_l=1.27，d_u=1.56，其中，N=29，K=3

D-W 检验的结果显示，D-W=1.60918> d_u，接近于 2，所以不存在序列相关性。

4.4 投资能力检验

测度投资能力，应该是要从选时能力和选股能力的角度上来看，而需要对此进行有效的测度，所以我们在此通过 T-M 模型和 H-M 模型来检验(仍旧选择的是博时价值)：

T-M 检验：

$$R_{pt}-R_{ft}=\alpha_p+\beta_1\times(R_{mt}-R_{ft})+\beta_2\times(R_{mt}-R_{ft})^2+\varepsilon_{pt} \quad (4\text{-}3)$$

其中，$R_{mt}-R_{ft}$：MKT 因子

表 5　T-M 检验

	coefficient	std. error	t-ratio	p-value
const	0.0340879	0.0254010	1.342	0.1912
MKT	1.11720	0.231502	4.826	5.32e-05 ***
MKT^2	0.821408	1.52016	0.5403	0.5936
Mean dependent var	0.045528		S.D. dependent var	0.154859
Sum squared resid	0.352559		S.E. of regression	0.116447
R-squared	0.474949		Adjusted R-squared	0.434561
F(2, 26)	11.75951		P-value(F)	0.000230
Log-likelihood	22.79337		Akaike criterion	-39.58674
Schwarz criterion	-35.48485		Hannan-Quinn	-38.30208
rho	0.245858		Durbin-Watson	1.470670

$\alpha_p > 0$，代表的是选股能力较强　　$\beta > 0$，代表的是选时能力较强

所以由此可见，博时价值的基金经理的选股能力还是比较强的，并且选时能力也是比较强的。综合来看，经理的投资能力还是较强的。

H-M 模型：

$$R_{pt}-R_{ft}=\alpha_p+\beta_1\times(R_{mt}-R_{ft})+\beta_2\times(R_{mt}-R_{ft})\times D+\varepsilon_{pt} \qquad (4\text{-}4)$$

其中，D：表示虚拟变量，当市场出现超额收益时，虚拟变量 D 的值为 1，否则，虚拟变量的值为 0。

按照上述方式运行，结果仍旧可以得到相关分析。

表 6　投资能力结果

证券简称	α_p	选股能力	β_2	选时能力
华夏成长	0.0431136	强	0.532273	强
嘉实增长	0.0449675	强	-0.369388	弱
长盛成长	0.0465520	强	-1.19386	弱
大成价值	0.0449258	强	-0.138537	弱
富国天源	0.0330346	强	-0.596093	弱
易方达平稳	0.0306957	强	-0.611186	弱
南方积极	0.0351009	强	0.374576	强
金鹰中小盘	0.0534243	强	-1.75222	弱
泰达宏利价值	0.0622980	强	-1.76269	弱
博时价值	0.0340879	强	0.821408	强

中银中国精选	0.0588839	强	-0.451332	弱
鹏华行业	0.0552858	强	-0.776147	弱
招商安泰	0.0274963	强	0.376059	强
景顺长城	0.0363449	强	0.106074	强
广发聚富	0.0524848	强	-0.799847	弱
东方龙混合型	0.0431703	强	-0.12323	弱

从上述结果可以看出，选股能力和选时能力同时兼备的基金有华夏成长、南方积极、博时价值、招商安泰、景顺长城，其他可测的基金都是在选股能力上较强，而选时能力较弱。并且可以看出基本上具有代表性的基金都可以满足选股能力，而在选时能力上还是比较欠缺的，也就是说，基金经理对时间的选择上较弱，不能在合理的时间区间内选择，所以这是基金经理的短处所在。

5、结论

从先前的 16 只基金的绩效测量，可以看出 Fama-French 三因子模型并不能很好的适用于中国的基金市场，相反的是，如果去掉规模因子 SMB 和账面价值与规模比率 HML 两个因子，检验的解释程度更高，但是这样的结果正是 CAPM 模型的很好解释。当然，按照国际的理论检验，如果加上动量因子后的 car-hart 模型后，可能可以更好的解释基金市场的收益率，在此笔者不再赘述。其次，从投资能力的角度出发，我们从选股能力和选时能力两个方面来看待基金经理的投资能力，从实证结果看，只有为数不多的几只基金同时在选股能力和选时能力上突出，而大部分的基金则表现在选时能力上的弱处，说明这是今后基金经理需要更深入分析和探讨的。

参考文献：

[1]Gruber, M. Another Puzzle: The Growth in Actively Managed Mutual Funds[J] . The Journal of Finance, 1996,51(3):783-810.

[2]Grinblatt, M., & Titman, S. Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings[J]. The Journal of Business, 1989,62(3):393-416.

[3]Daniel, K., Grinblatt, M., & Titman, S., et al. Measuring Mutual Fund Performance with Characteristic-Based Benchmarks[J]. Journal of Finance,1997,53(2)1035-1058.

[4]Agyei-Ampomah, S., Clare, S., & Mason, S. On luck versus skill when performance benchmarks are style-consistent[J]. Journal of Banking & Finance, 2015,59:127-145.

[5]Rodríguez, J., & Romero, H. Assessing foreign funds geographical focus timing skill[J]. Studies in Economics and Finance, 2016,33(2):209-221.

[6] 江萍，田澍，Cheung Yan-Leung:《基金管理公司股权结构与基金绩效研究》,《金融研究》, 2011；123-134

[7] 徐琼，赵旭:《我国基金经理投资行为实证研究》,《金融研究》, 2008:145-155

[8] 赖淑懿:《中国股票型开放式基金绩效综合评价研究》,《西南财经大学博士论文》, 2010：1-130

基于 CiteSpace 的国内风险收益研究可视化分析

朱红珊

摘　要：本文结合 CiteSpace 软件对国内关于风险收益的文献进行统计和可视化分析，揭示该领域的研究现状、研究前沿及学科基础，为风险收益研究领域的研究人员提供参考与借鉴。同时本文尝试将可视化分析方法引入金融研究领域，通过创新性的分析思路，完善该领域可视化分析方法。首先导出 CSSCI 中国社会引文检索数据库所收录的研究风险收益主题的引文，并结合 CiteSpace 软件，对其进行统计和可视化分析，得出国内关于风险收益研究期刊文献的历年发表量、关键词及关键词聚类、施引作者及作者共被引、机构及机构合作、高被引文献及文献共被引可视化分析。由分析结果可知，目前我国风险收益领域的文献以微创新为主，缺少实质创新内容，该领域还有很大的探索空间，量化投资是近期的一个研究的热点。

关键词：CiteSpace；风险收益；关键词聚类；可视化分析

0 引言

马克维茨（Markowitz,1952）是研究投资风险回报替换关系的先驱之一，后来夏普(Sharpe,1964)及其他人以马克维茨理论为基础，进一步发展了资本资产定价模型。资本资产定价模型显示了预期回报和系统性风险的关系，罗斯（Ross，1976）发表了风险套利定价理论，这一理论将资本资产定价理论中单一系统风险假设推广到多系统风险。这些颇具洞察力的理论研究成果对资产组合管理以及分析风险回报关系产生了巨大影响。鲍曼（Bowman，1980）在 1980 年得出的鲍曼悖论理论，为风险收益的研究提供了新的认识。我国自 1978 年进行市场化改革开始，学界对于风险收益领域的研究逐步加深。风险收益在金融学、管理学、农业科技、基础科学等都有涉及，在先后经历 1997 年的亚洲金融危机以及 2007 年的美国次贷危机后，国内对风险收益领域的研究已经形成了一定的规模，并且形成了不同的研究前沿和学科基础。

1 数据来源及研究方法

对于本文研究的期刊文献来源，经综合考虑选择了 CSSCI 中国社会科学引文索引数据库。我们运用的分析软件是 CiteSpace 软件，它主要基于共引分析理论和寻径网络等算法，对特定领域文献（集合）进行计量，探寻出学科领域演化的关键路径及知识转折点，并通过一系列可视化图谱的绘制来形成对学科领域演化潜在动力机制的分析和学科发展前沿的探测。该软件能提供聚类视图、时间线视图和时区视图，能够将复杂的学术研究领域的发展进程以可视化的方式显示。我们将通过合理设置阈值，并依据得出的结果来对风险收益进行可视化分析。

1.1 文献检索条件设置

在 CSSCI 中国社会科学引文索引页面，我们以“风险”与“收益”为关键词，文献类型、学科类别、

本文作者：朱红珊，江西财经大学金融学院研究生　研究方向：金融工程

学位分类和基金类别使用默认设置，时间范围选择“1998 年—2014 年”，进行来源文献检索，共检索出 451 篇引文，将文献下载并转换格式，使用这 451 篇引文进行可视化分析。

1.2 软件参数设计

首先设置 CiteSpace 支持中文编码，并将下载的文献分别进行存储路径、文件名称、格式转换等一系列的处理，时间跨度都选择 1998-2014 年，时间切面都为 1 年，然后分别选择 Keyword、Institute、Author 等，设置相应的阈值，运行 CiteSpace 并对生成的可视化图进行分析。

2 可视化分析与结果讨论

本文将借助 CiteSpace 软件，对国内关于风险收益的研究文献可视化分析。

2.1 文献年代分布

亚洲金融危机在 1997 年夏爆发，国内也是此时逐渐开始增强风险收益的研究，如 1998 年蒲宇飞发表的《谁将第一个摆脱危机——关于亚洲金融危机的几点认识》中探讨亚洲金融危机两方面的问题，一是关于亚洲模式的评价 ，二是关于摆脱金融危机时机与方式的预测。作者认为决定一个国家模式优劣的，最重要的是一个国家的自我矫正能力。王国平发表的《经济体制“改革攻坚”的目标及途径》中提出了，完善分散投资主体是推进改革的现实着眼点。正是基于这样的大背景，1998 年开始国内学术界对风险收益的研究愈加重视。

2.2 关键词及关键词聚类分析

2.2.1 关键词分析

论文的关键词是整篇论文的核心词和词组，它们在一定程度上反映了论文研究的内容。同时论文的关键词是对论文主题的高度归纳和总结，因此对论文的关键词进行分析和研究，频次高的关键词能反映一个领域的研究热点。我们选择关键词（Keywords）作为生成可视化聚类视图的依据。按（2，2，20）、（2，2，20）、（2，3，20）的节点阈值选择节点数据，导入文献运行 Citespace 软件。

生成的关键词共现聚类视图，共选择出 62 个关键词节点及 52 条连线。图中节点表示关键词，一个年轮的厚度与某个时间分区内的关键词出现频次成正比，节点之间的连线代表关键词的共现。国内研究文献中前 14 位高频关键词，剔除前两名，分别为投资风险、投资收益、风险收益、收益率、风险投资等，结合关键词的引用历史分析该研究主要集中在金融投资研究领域。

图 1 关键词聚类图谱

2.2.2 关键词聚类分析

接着依照 TF*IDF 加权算法（TF*IDF）从选取文献的“索引词条”中提取聚类标签词，运行软件生成如图 1，导出关键词聚类表 1。结合关键词似然率算法挑出的关键词和聚类推荐的重点文献，国内风险收益研究研究有 6 类研究前沿。

聚类 0#市场风险，研究结合相应指标，对风险收益理论在我国的实证研究。聚类 1#投资风险，研究我国证券公司以及基金公司在不同环境下对我国二级市场进行的组合投资。聚类 2#道德风险，研究保险行业以及银行业面对市场风险时，相应的对策。聚类 3#风险收益，是基于投资者情绪的风险收益实证研究。聚类 4#理论收益率，结合我国股市特色，探寻合理证券投资组合的，探寻合理报酬的实证研究，黄秀海、滕秀清和李杰于 2010 年提出并证明了非参数核估计能够很好地估计股市理论收益率指标，准确地计量与监管我国股市风险动态变化情况。聚类 5#”双边道德风险”，分析基于 CVaR 风险度量准则下，有效消除双边道德风险从而提高收益的方法，代建生、孟卫东、马国旺于 2013 年，基于 CVaR 风险度量准则，在收益共享契约下，证明了供应链联合促销博弈存在一个纳什均衡。殷林森于 2010 年，提出通过间接控制私人收益来激励双方的努力，能部分消除双边道德风险。

表 1 国内风险收益文献可视化研究的关键词聚类（TF*DF 算法）

聚类 ID	聚类名	平均年份	聚类关键词(TF*DF)
0	市场风险	2003	“市场风险”、“收益率”、“风险分析”、“风险态度”、“风险投资”
1	投资风险	2000	“投资风险”、“风险投资”、“风险投资”、“投资收益”、“证券市场”
2	道德风险	2006	“道德风险”、“次贷危机”、“预期收益”、“收益风险”
3	风险收益	2001	“风险收益”、“证券市场”、“投资风险”
4	理论收益率	2010	“理论收益率”、“置信区间”、“非参数核估计”、“区域监管机制”、“股市风险”
5	双边道德风险	2011	“双边道德风险”、“私人收益”、“股权契约”、“私人收益”

2.3 来源作者与作者共被引分析

有学术影响力的作者群体的状况集中体现了学科科研实践的趋向，作者的学术影响力可以通过发文量、合著情况以及被引频次来衡量，发文量从数量角度反映作者的科研生产力，合著情况及被引频次从质量角度反映作者的学术影响力。

2.3.1 来源作者分析

我们选择作者（author）作为生成可视化聚类视图的依据，节点阈值选取 top=30，其他设置为默认值不变。生成的作者合作科学知识图谱，共选出 448 个节点以及 148 条连线，表明国内作者间合作相对频繁。发文量排前十四的作者都与他人合作撰写了风险收益研究领域的文献。分析通过导出的作者间合作列表，国内作者跨区域以及跨机构间的合作相对较少。

2.3.2 作者共被引分析

作者共被引分析着重从作者共被引方面分析作者的影响力，在每一个领域，被引次数最多的作者一般是那些做了奠基性研究工作的学者，这些学者对该领域的发展及演变往往产生了根本又深远的影响。

我们选择被引作者（Cited Author）作为生成可视化聚类视图的依据，节点阈值选取 top=30，算法路径选择 pathfinder 算法，其他设置为默认值不变，得到图 2。由作者共被引图谱在风险收益研究领域

被引用作者外国作者将近一半。统计导出的被引作者频次表，被引国外作者占比接近 48.72%，在国内文献中出现这种比例的被引情况，也从侧面说明在风险收益这一研究领域我国与国外的研究成熟度还是有一定差距。国内被引作者中陈浪南、吴世农、张维迎三人被引次数居于前三。王春峰的被引频次以及发文量都居于前列，他是我国金融工程科学的主要创始人之一。从被引作者统计情况可知，在风险收益这一研究领域我国与国外的研究成熟度还是有一定差距。

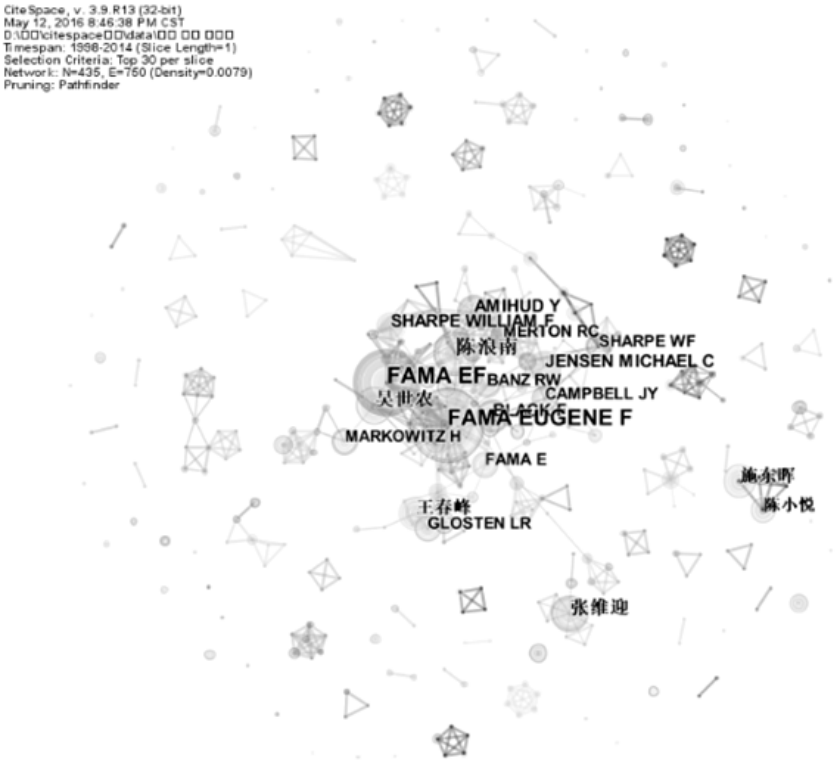


图 2 作者共被引图谱

2.4 机构及机构合作分析

对研究领域发文机构的研究可以看出该研究领域中的研究力量布局。我们选择机构（institution）作为生成可视化聚类视图的依据，共选择出 321 所机构及之间的 78 条连线。从分析结果可知，虽然机构之间会有合作，但合作的机构多集中在同一区域，如黑龙江省诚荣建设开发有限公司、哈尔滨工业大学市政环境工程学院以及松原市顺达监理公司合作撰写《基础设施资产证券化融资研究》，并且机构之间的合作相对较少。大学是研究风险收益的主要机构，像国务院发展研究中心金融研究所等研究机构，以及像普益财富、信诚基金等金融机构都是风险收益研究领域的重要力量。跨机构之间的合作能充分利用各自的优势，实现在风险研究领域的突破创新。如中国国有经济研究中心、哈尔滨商业大学.财政与公共管理学院和中国银行.长春分行三机构共同完成的《中国商业银行风险、收益对效率影响的实证分析》。在风险收益研究领域机构之间加强合作，并且加大跨行业、跨区域之间的机构合作能更好的拓展风险收益领域的研究。

2.5 文献共被引分析

2.5.1 高被引文献分析

一个研究领域的知识基础与其研究前沿将随着时间的变迁而不断延伸和拓展，最终形成一个学科的核心文献集合，并进而表现为该领域的高被引文献簇。图 3 中的节点显示的是被引次数较多的几篇

重要被引文献，可以看到被引度较高的大多都是国外文献，统计导出的文献高被引频次表，高被引文献中国外文献占比接近 3/4。我国风险收益领域高被引文献以微创新为主，通过借鉴国外创新性的研究思路和研究方法可以较大地拓展我国在这一领域的深度和广度，我国风险收益领域具有巨大拓展前景。

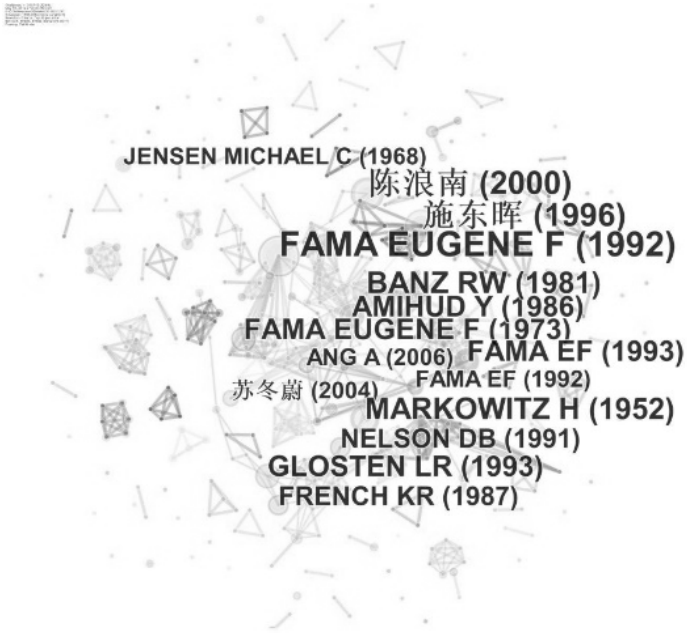


图 3 高被引文献图谱图

2.5.2 文献共被引聚类图谱

图 4 是文献共被引的时间序列图谱，它由 57 个聚类组成，年代分布从 1952 年到 2009。图中显示 2020 年是因为下载下来的数据中部分引文文献发表时间缺失，在用刘盛博的 java 程序进行转换时将其设为 2020 年。文献共被引的时间序列图谱不仅能展现被引率最高的重要文献还能清晰的给出该领域的知识基础和发展脉络。依据选出的被引文献，由文献共被引根据似然率算法，进行聚类，共得出 28 类聚类。聚类反映的是风险收益研究领域不同热点的学科基础，结合时间线视图对聚类进行分析，可以归纳出目前国内风险收益研究领域的 28 个研究前沿。

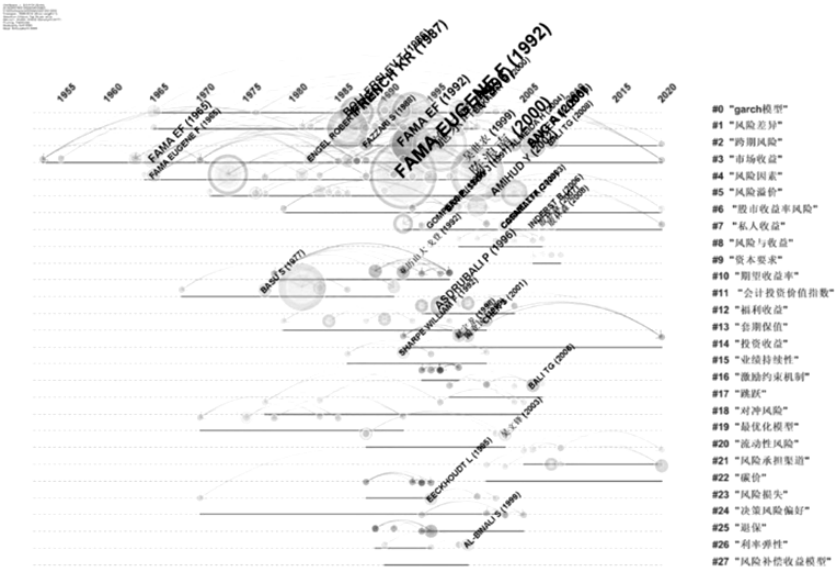


图 4 共被引文献图谱

2.5.3 文献聚类原理

关键词聚类分析的是学科的研究前沿，文献共被引聚类分析的是学科基础，他们都是基于共现网络理论，将共词分析和共引分析相结合，结合科学计量学主流的引文分析图谱基本原理，通过似然率算法形成聚类，如图 5。本文 CiteSpace 软件选择 TF*IDF 加权选词排序算法，从文献的标题、文摘、关键词中抽取聚类标签词，作为共现聚类的依据，再通过 Modularity Q 指标和 Silhouette 指标对抽取的词进行聚类结果运算，展示学科基础与隐含的研究前沿的对应关系。

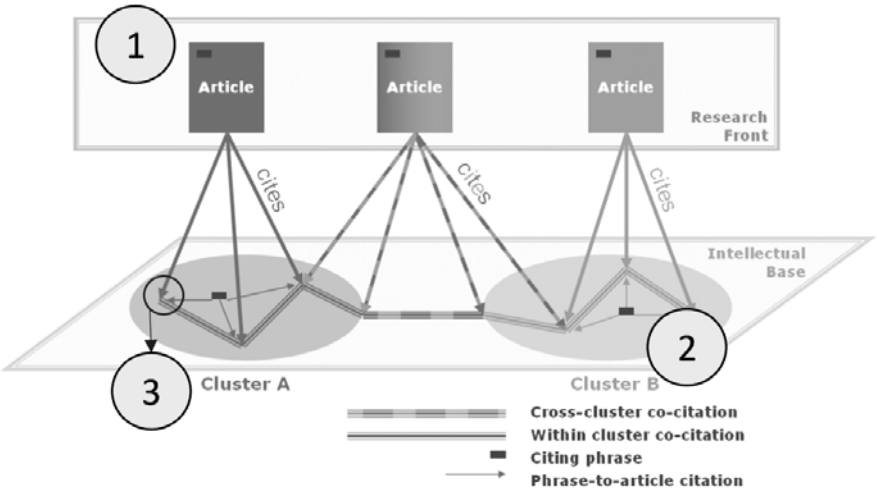


图 5 CiteSpace 的聚类概念模型

3 结论与不足

经历 1997 年亚洲金融危机后，国内风险收益领域发文量逐年增加，在 2008 年次贷危机后又呈现增加趋势。通过关键词聚类生成六大聚类，并且根据聚类文献找到六大前沿方向。国内作者跨区域以及跨机构之间的合作较少。在风险收益领域，量化投资是一个研究的热点。国内被引作者相对国外作者较少。我国风险收益领域高被引文献以微创新为主，缺少实质创新内容。根据施引文献聚类得出 28 类学科基础。大学是研究领域的主力军，机构与机构之间的跨领域合作相对不足，量化投资是风险收益领域的一大研究热点。

从目前的分析结果来看，本文的数据来源、研究方法、研究结果都有一定的现实意义，但总体而言，本研究还存在许多不足之处：如数据文献完整性方面，CSSCI 中国社会科学引文索引库并不能穷尽国内风险收益研究文献；在软件阀值确定方面，我也只是根据经验来进行阀值选择，因此具有一定的主观性。在以后的研究中，我们将对该领域进行更深入的分析。

参考文献：

[1]陈悦、陈超美、刘泽渊等，2015：《CiteSpace 知识图谱的方法论功能》,《科学学研究》，第 2 期：P242 ~ 253。

[2]胡志刚、陈超美、刘则渊等，2013：《从基于引文到基于引用——一种统计引文总被引次数的新方法》,《图书情报工作》，第 21 期：P5 ~ 10。

[3]陈超美、陈悦，2014：《科学发现的结构与时间属性》,《科学与管理》，第 3 期：P27 ~ 33。

[4]顾洪梅，谢淑萍，温秀玲, 2014：《中国商业银行风险、收益对效率影响的实证分析》,《社会科学战线》，第 5 期：P248 ~ 250。