

中国股票市场可预测性的实证研究

姜富伟¹ 涂俊¹ David E. Rapach² Jack K. Strauss² 周国富^{3,4}

(1. 新加坡管理大学李光前商学院; 2. 圣路易斯大学经济系;

3. 华盛顿大学奥林商学院, 美国 263130; 4. 上海交通大学中国金融研究院)

摘要: 我们研究了中国市场投资组合和根据公司行业、规模、面值市值比和股权集中度等划分的各种成分投资组合的股票收益的可预测性。选取各种经济变量作为预测变量, 中国市场投资组合和各种成分投资组合都存在显著的样本内和样本外可预测性。不同成分投资组合的可预测性存在显著差异, 其中金融与保险业、房地产业和制造业等行业投资组合的可预测性特别强, 小市值、低面值市值比和低股权集中度的投资组合也非常容易预测。对于成分投资组合间的可预测性差异, 我们给出了两个经济解释: (1) 基于样本外可预测性分解, 我们发现条件 CAPM 模型捕捉的时变系统性风险溢价可预测性可以解释成分投资组合的大部分样本外可预测性, 高系统性风险暴露的投资组合有较高的样本外可预测性; (2) 基于 Hong, Torous, and Valkanov (2007) 的信息流动摩擦理论, 我们发现行业集中度可以显著解释行业投资组合间的可预测性差异。

关键词: 成分投资组合; 样本内预测; 样本外预测; 条件 CAPM; 信息流动摩擦

JEL分类号: C22; C53; G12; G17 **文献标识码:** A

一、引言

股票收益预测是金融研究的核心问题之一, 例如, 它是 2011 年美国金融学会会长 Cochrane 演讲论文的主题 (Cochrane, 2011)。它还和很多重要的金融学问题密切相关, 比如投资组合管理、资金成本和市场有效性等。大量文献围绕股票收益可预测性展开了研究, 众多经济变量被发现可以作为股票收益的预测变量^①。Campbell (2000) 指出, 学术界对股票收益有显著的样本内可预测性已基本达成共识。然而, 股票收益的样本外可预测性至今还有争议, 比如, Welch and Goyal (2008) 发现许多流行的预测变量并没有在样本外预测中优于历史平均收益这一简单预测。但 Campbell and Thompson (2008) 提出, 通过附加基于金融理论参数约束, 股票收益在样本外也是可预测的。

虽然已经有大量文献对股票收益的可预测性进行了研究, 但几乎所有的文献都是针对美国市场。与现有文献不同, 本文研究了中国股票市场的可预测性^②。研究中国股票市场收益的可预

作者简介: 姜富伟, 金融学博士研究生, 新加坡管理大学李光前商学院, Email: jfuwei@gmail.com。

涂俊, 金融学博士, 新加坡管理大学李光前商学院助理教授, Email: tujun@smu.edu.sg。

David E. Rapach, 经济学博士, 圣路易斯大学经济系副教授, Email: rapachde@slu.edu。

Jack K. Strauss, 经济学博士, 圣路易斯大学经济系 Simon 讲席教授, Email: Strausjk@slu.edu。

周国富, 经济学博士, 华盛顿大学奥林商学院 Frederick Bierman and James E. Spears 金融学讲席教授和上海交通大学中国金融研究院研究员, Email: zhou@wustl.edu。

本文曾获 The Chinese Finance Association (TCFA) 中国资本市场研究最佳论文奖。作者感谢邹恒甫、孔爱国、田澍、吴有昌、邱劲、范龙振、徐冶晓、龚关、周晓岚、梁丹、胡晓娟、路磊、宿成建、熊和平、Jerry Cao、Jeremy Goh、Jun Wang、Jason Chenyang Wei、Joe Zhang 及参加 2010 年中国国际金融年会 (CICF)、2010 年第 16 届 TCFA Annual Conference 和新加坡管理大学学术报告会的与会学者的宝贵意见和建议, 但文责自负。作者涂俊感谢 Sing Lun Fellowship 的资助。

^① 文献发现的预测变量包括股息价格比、盈余价格比、面值市值比、通货膨胀率、股票发行量、波动率和货币政策等。关于股票收益预测的文献综述可参看 Campbell (2000) 和 Welch and Goyal (2008)。

^② 只有很少几项研究检验了中国股票市场收益的可预测性, 而且本文与这些研究有显著区别。例如, Chen,

测性有如下重要意义：首先，研究中国股票市场收益的可预测性对中国股票横截面收益的资产定价模型检验有潜在重要意义；其次，研究中国股票市场收益的可预测性对中国资金成本测度也有重要意义；第三，研究中国股票市场收益的可预测性有助于为投资中国股票市场的证券投资基金设定合理的业绩评价基准；第四，研究中国股票市场收益的可预测性加深了对除美国外的全球股票市场可预测性的理解。

本文对中国股票市场的可预测性进行了如下分析：首先，我们既研究了市场投资组合的可预测性，又研究了各种成分投资组合的可预测性，包括 13 个行业投资组合、10 个公司规模投资组合、10 个面值市值比投资组合以及 10 个股权集中度投资组合。我们借鉴 Welch and Goyal (2008) 选取 12 个经济变量作为预测变量。其次，我们分别从样本内和样本外可预测性两个角度研究了中国股票市场可预测性。我们还重点研究了主成分预测方法的样本外预测能力。Ludvigson and Ng (2009) 和 Rapach, Strauss, Tu and Zhou (2011) 发现主成分预测方法可以显著预测股票和债券收益。第三，我们利用样本外可预测性分解和 Hong, Torous, and Valkanov (2007) 的信息流动摩擦理论分析了成分投资组合间可预测性差异的经济原因。

样本内检验发现，股息率、换手率、货币供给增长率等经济变量可以显著预测中国市场投资组合以及绝大多数成分投资组合收益。不同成分投资组合的可预测性存在显著差异，比如房地产业、金融与保险业和制造业等行业投资组合有相当强的样本内可预测性，而采掘业、信息技术业和传播与文化产业等行业投资组合的样本内可预测性则小很多。

样本外检验发现，所有的行业投资组合都有显著的样本外可预测性，房地产业、金融与保险业、制造业和批发与零售贸易业等行业的样本外可预测性尤其强。规模、面值市值比和股权集中度等成分投资组合也有显著的样本外可预测性，其中小市值、低面值市值比和低股权集中度的投资组合的样本外可预测性尤其强。比较发现，中国股票市场在泡沫较大的 2007 年—2009 年比在相对比较平稳的 2002 年—2006 年呈现出相对更强的样本外可预测性。总之，样本内检验和样本外检验都证明，中国股票市场有非常强的可预测性，并且不同成分投资组合的可预测性存在显著差异。

我们对成分投资组合间的可预测性差异给出了两个经济解释。首先，我们用条件 CAPM 模型把成分投资组合的样本外可预测性分解成时变系统性风险溢价可预测性和 α 可预测性两部分。条件 CAPM 模型捕捉的时变系统性风险溢价可预测性可以解释投资组合的绝大部分样本外可预测性。高系统性风险暴露的投资组合有较高的样本外可预测性。接着，我们检验了 Hong, Torous, and Valkanov (2007) 的信息流动摩擦理论对不同行业投资组合间可预测性差异的解释能力。Hong, Torous, and Valkanov (2007) 提出，信息在股票市场中逐步扩散，部分行业对经济信息的反应存在时滞。本文的实证研究发现行业集中度与行业投资组合的样本外可预测性显著负相关，支持了信息流动摩擦理论。总之，时变系统性风险溢价可预测性和信息流动摩擦都有助于解释中国股票收益的可预测性。

本文的其余部分组织如下：第二部分对中国股票收益的样本内可预测性做实证检验；第三部分对中国股票收益的样本外可预测性做实证检验；第四部分讨论不同投资组合间的可预测性差异的经济原因；第五部分是本文的结论。

二、样本内可预测性检验

这一部分首先介绍了预测性回归模型，然后说明了使用的数据，最后报告了样本内可预测性检验的结果。

(一) 计量方法

Kim, Yang, and Yu (2010) 检验了中国企业层面股票收益可预测性。Wang and Cheng (2004) 检验了换手率对中国横截面股票收益的预测能力。

与相关文献一致，我们使用单变量预测性回归模型分析中国股票收益的可预测性：

$$r_{i,t} = a_i + b_{i,j}x_{j,t-1} + e_{i,t} \quad (1)$$

其中， $r_{i,t}$ 是投资组合 i 的收益与无风险利率之差， $x_{j,t}$ 是潜在预测变量， $e_{j,t}$ 是残差项。在本文中， $r_{i,t}$ 代表市场投资组合以及 13 个行业、10 个规模、10 个面值市值比和 10 个股权集中度成分投资组合的超额收益。我们一般用公式 (1) 中 $b_{i,j}$ 的最小二乘法 (OLS) 估计 $\hat{b}_{i,j}$ 对应的 t 统计量来判断潜在预测变量 $x_{j,t}$ 对 $r_{i,t}$ 是否有显著预测能力。原假设是 $b_{i,j} = 0$ ，即没有预测能力；备择假设是 $b_{i,j}$ 不等于零，即 $x_{j,t}$ 包含的信息可以用来预测 $r_{i,t}$ ，故而应使用公式 (1) 对应的时变预期收益模型预测股票超额收益。本文使用 bootstrap 方法进行统计推断以避免假设检验中的小样本水平扭曲问题 (Size Distortion)。

(二) 数据描述

本文研究了中国市场投资组合以及 13 个行业、10 个规模、10 个面值市值比和 10 个股权集中度成分投资组合超额收益的可预测性。股票收益数据来自锐思数据库，包括在上海和深圳上市的所有正常 (没有被 “PT” 或 “ST”) A 股上市公司。首先，我们用所有正常 A 股上市公司构造了 1996 年 7 月—2009 年 6 月的市值加权市场投资组合。用于计算超额收益率的无风险利率也来自锐思数据库。第二，在每年 6 月底，按照中国证监会行业分类，我们构造 13 个市值加权的行业投资组合：农、林、牧、渔业 (AGR)、采掘业 (MNS)、制造业 (MAN)、电力、煤气及水的生产和供应业 (UTL)、建筑业 (CNT)、交通运输、仓储业 (TRS)、信息技术业 (INF)、批发和零售贸易业 (WRS)、金融、保险业 (FIN)、房地产业 (PRT)、社会服务业 (SVC)、传播与文化产业 (MED)、综合类 (MUT)。第三，在每年 6 月底，按照当年 6 月底的股票市值构造 10 个含相同数量上市公司的市值加权的规模投资组合，并按升序依次命名为 S1,...,S10。第四，在每年 6 月底，按照当年 6 月底的股票面值市值比构造 10 个含相同数量上市公司的市值加权的面值市值比投资组合，并按升序依次命名为 BM1,...,BM10。其中，第 t 年 6 月的股价乘以股数得到的第 t 年 6 月的市值，第 t 年 6 月的市值和第 $t-1$ 财年末的账面价值被用于计算 t 年 6 月底的面值市值比。最后，在每年 6 月底，按照当年 6 月底的股权集中度，构造了 10 个含相同数量上市公司的股权集中度投资组合，并按升序依次命名为 OC1,...,OC10。其中，第 $t-1$ 年末最大股东持股比例被用于计算第 t 年 6 月的股权集中度。

我们参考 Welch and Goyal (2008) 选取了 12 个经济变量作为中国股票超额收益的潜在预测变量。其中，股利支付率 (D/E) 是在上海或深圳上市的所有 A 股上市公司的总股利的对数与总盈余的对数之差，总股利和总盈余分别是通过把过去 12 个月的股利和盈余移动加总得到。股利价格比 (D/P) 是所有 A 股上市公司的总股利的对数与总市值的对数之差。股息率 (D/Y) 是所有 A 股上市公司的总股利的对数与滞后总市值的对数之差。盈余价格比 (E/P) 是所有 A 股上市公司的总盈余的对数与总市值的对数之差。面值市值比 (B/M) 是所有 A 股上市公司的总账面价值除以总市值。账面价值来自上市公司年报和中报。每年 1 月到 3 月的面值市值比等于上一年 6 月的账面价值除以当月的市值，4 月到 9 月的面值市值比等于上一年 12 月的账面价值除以当月的市值，10 月到 12 月的面值市值比等于当年 6 月的账面价值除以当月的市值。股票方差 (SVR) 是 A 股市值加权市场投资组合日收益率平方和。通货膨胀 (INF) 根据国家统计局公布的居民消费价格指数 (CPI) 算出，由于当月通货膨胀数据一般在下一个月公布，因此我们使用 $x_{j,t-2}$ 作为公式 (1) 中的通货膨胀数据 (Welch and Goyal, 2008)。净权益增加 (NTIS) 是过去 12 个月上海和深圳交易所新股发行量的移动加总除以当月总市值，新股发行数据来自中国证监会。换手率 (TO) 是上海和深圳交易所 A 股总交易量除以总市值，总交易量数据来自 CEIC。M0 增长率 (M0G) 根据 M0 数据算出，流通现金 M0 定义为当月流通中现金总量。M1 增长率变动量 (M1G) 是当月 M1 增长率的与上月 M1 增长率之差即未预期的 M1 货币冲击，狭义货币 M1 定义为当月 M0 加企事业单位活期存款。M2 增长率 (M2G) 利用 M2 数据算出，广义货币 M2 定义为当月 M1 加居民的存折储蓄和定期存款，M0、M1 和 M2 数据都来自中国

人民银行。

描述性统计研究发现，行业投资组合的月平均收益在 0.65%（CNT）到 2.33%（MNS）之间，标准差在 9.25%（UTL）到 12.11%（MED）之间。同时，小市值和高面值市值比的股票的平均收益和波动率都比较高。

表 1 行业投资组合样本内预测结果

| 组合 | D/P | D/Y | D/E | SVR | E/P | B/M | INF | NTIS | TO | M0G | M1G | M2G | \bar{R}^2 |
|-------------|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|-------|-------------|
| MKT | 2.02 | 2.28* | 1.25 | 0.80 | 1.64 | 1.38 | 1.70* | 1.28 | 2.98* | 1.61 | 1.72* | 2.58* | 2.19 |
| | 2.58 | 3.28 | 1.01 | 0.41 | 1.72 | 1.23 | 1.85 | 1.06 | 5.46 | 1.65 | 1.88 | 4.13 | |
| AGR | 1.17 | 1.27 | 0.20 | 2.02* | 1.31 | 0.59 | 1.09 | 1.40 | 3.08* | 1.30 | 1.07 | 2.16* | 1.54 |
| | 0.88 | 1.04 | 0.03 | 2.59 | 1.10 | 0.23 | 0.76 | 1.26 | 5.80 | 1.09 | 0.74 | 2.93 | |
| MNS | 0.85 | 1.09 | 0.45 | 0.22 | 0.74 | 0.90 | 1.68* | 1.54 | 1.95* | 1.09 | 1.36 | 1.99* | 1.04 |
| | 0.46 | 0.76 | 0.13 | 0.03 | 0.36 | 0.52 | 1.80 | 1.52 | 2.40 | 0.77 | 1.18 | 2.50 | |
| MAN | 2.18 | 2.38* | 1.16 | 0.95 | 1.90 | 1.61 | 1.31 | 1.32 | 2.73* | 1.52 | 1.71* | 2.28* | 2.11 |
| | 2.98 | 3.55 | 0.86 | 0.58 | 2.28 | 1.65 | 1.10 | 1.12 | 4.61 | 1.47* | 1.87 | 3.26 | |
| UTL | 1.33 | 1.44 | 1.07 | 1.31 | 0.92 | 0.73 | 1.77* | 1.18 | 2.55* | 1.68* | 1.35 | 1.71* | 1.41 |
| | 1.13 | 1.33 | 0.74 | 1.10 | 0.54 | 0.34 | 1.99 | 0.90 | 4.04 | 1.81 | 1.18 | 1.86 | |
| CNT | 1.85 | 1.84* | 1.17 | 1.36 | 1.49 | 1.39 | 1.73* | 2.28* | 1.41 | 1.54 | 1.08 | 0.45 | 1.49 |
| | 2.18 | 2.16 | 0.89 | 1.19 | 1.43 | 1.23 | 1.90 | 3.26 | 1.27 | 1.51 | 0.76 | 0.13 | |
| TRS | 1.21 | 1.53 | 0.99 | 0.99 | 0.83 | 0.55 | 2.46* | 0.71 | 3.29* | 1.09 | 1.21 | 2.47* | 1.71 |
| | 0.94 | 1.50 | 0.63 | 0.64 | 0.44 | 0.19 | 3.79 | 0.32 | 6.56 | 0.76 | 0.95 | 3.83 | |
| INF | 1.18 | 1.40 | 0.07 | 1.42 | 1.41 | 1.00 | 0.95 | 0.86 | 2.32* | 0.73 | 1.24 | 2.48* | 1.25 |
| | 0.90 | 1.26 | 0.00 | 1.29 | 1.28 | 0.64 | 0.59 | 0.47 | 3.39 | 0.34 | 0.99 | 3.83 | |
| WRS | 1.93 | 2.14* | 0.66 | 1.04 | 1.94 | 1.54 | 0.85 | 1.60 | 2.65* | 1.54 | 1.50 | 1.87* | 1.82 |
| | 2.37 | 2.88 | 0.28 | 0.70 | 2.38 | 1.52 | 0.47 | 1.64 | 4.37 | 1.52 | 1.43 | 2.22 | |
| FIN | 2.62* | 2.76* | 1.99* | 0.59 | 1.88 | 1.89 | 1.81* | 1.08 | 2.52* | 0.78 | 0.83 | 2.48* | 2.31 |
| | 4.27 | 4.70 | 2.50 | 0.22 | 2.24 | 2.28 | 2.08 | 0.76 | 3.96 | 0.39 | 0.45 | 3.84 | |
| PRT | 2.70* | 3.03* | 1.31 | 1.08 | 2.42* | 1.96 | 1.51 | 1.20 | 3.69* | 1.12 | 2.00* | 2.54* | 3.00 |
| | 4.50 | 5.62 | 1.11 | 0.75 | 3.67 | 2.42 | 1.45 | 0.93 | 8.14 | 0.80 | 2.53 | 4.01 | |
| SVC | 2.08 | 2.37* | 0.62 | 1.11 | 2.14 | 1.76 | 1.19 | 1.40 | 2.75* | 1.50 | 1.32 | 1.59 | 1.93 |
| | 2.73 | 3.51 | 0.25 | 0.79 | 2.90 | 1.97 | 0.91 | 1.25 | 4.68 | 1.43 | 1.13 | 1.61 | |
| MED | 1.67 | 1.69 | 0.42 | 0.95 | 1.78 | 1.56 | -0.32 | 1.50 | 1.51 | 1.22 | 1.32 | 1.33 | 1.17 |
| | 1.77 | 1.81 | 0.11 | 0.59 | 2.01 | 1.56 | 0.07 | 1.44 | 1.46 | 0.96 | 1.12 | 1.14 | |
| MUT | 2.00 | 2.15* | 0.53 | 1.29 | 2.11 | 1.61 | 1.26 | 1.09 | 2.58* | 1.43 | 1.67* | 2.24* | 1.95 |
| | 2.53 | 2.90 | 0.18 | 1.07 | 2.82 | 1.65 | 1.01 | 0.77 | 4.15 | 1.32 | 1.78 | 3.17 | |
| # Sig | 2 | 7 | 1 | 1 | 1 | 0 | 5 | 1 | 11 | 2 | 3 | 10 | |
| \bar{R}^2 | 2.13 | 2.54 | 0.59 | 0.89 | 1.80 | 1.25 | 1.38 | 1.20 | 4.22 | 1.09 | 1.24 | 2.64 | |

注：本表报告了行业投资组合超额收益样本内预测性回归的回归系数 $b_{i,j}$ 的 t -统计量（上面的数据）和 R^2 统计量（百分比，下面的数据）。MKT 行是 A 股市场投资组合超额收益的预测结果。样本期是 1996 年 7 月—2009 年 6 月。“*” 代表在 5% 水平下显著。# Sig 行是每一列中在 5% 水平显著的行业数量。 \bar{R}^2 列或行是每一列或行的 R^2 统计量的平均值，列平均排除了 MKT。

（三）市场投资组合

表 1 的 MKT 行是当公式（1）的 $r_{i,t}$ 是市场投资组合超额收益、 $x_{j,t}$ 是 12 个经济变量中的某一个时的样本内预测结果。样本期是 1996 年 7 月—2009 年 6 月。对于每个投资组合/预测变量组合，表内数据包括公式（1）中 $b_{i,j}$ 的 t 统计量（上面的数据）和 R^2 统计量（下面的数据）。最后一列是预测变量 R^2 的均值。预测性回归模型的 R^2 统计量一般比较小，Campbell and Thompson（2008）认为月股票收益的预测性回归 R^2 只要约大于 0.5% 在经济上就是显著的，也可参见 Xu（2004）。D/Y、INF、TO、M1G 和 M2G 等 5 个经济变量在 5% 水平下显著预测市场投资组合超额收益。同时，这些变量对行业投资组合的预测能力也是最强的。

（四）行业投资组合

表 1 从第二行到最后一行是行业投资组合的样本内预测结果。最后一行是不同行业的平均 R^2 。倒数第二行是各经济变量在 5% 的显著水平下可以显著预测的行业的数量。能显著预测市场投资组合的 D/Y、INF、TO 和 M2G 等经济变量分别可以显著预测 7 个、5 个、11 个和 10 个行业投资组合。因此，行业投资组合预测和市场投资组合预测是紧密相关的，两者都有很强的样本内可预测性。从最后一列可见，不同行业投资组合的样本内可预测性存在显著差异，MAN、FIN 和 PRT 等行业最容易预测，平均预测性回归 R^2 大于 2%；而 MNS、INF 和 MED 等行业的可预测性差很多，平均 R^2 约为 1%。

(五) 其他成分投资组合^③

我们发现规模、面值市值比和股权集中度等成分投资组合也有很强的样本内可预测性。预测能力最强的 D/Y、TO、M1G 和 M2G 等 4 个经济变量分别可以预测 6 (9; 9) 个、10 (10; 10) 个、7 (5; 6) 个和 10 (10; 10) 个规模 (面值市值比; 股权集中度) 成分投资组合。在第三部分我们可以发现，这三类成分投资组合间也有很大的可预测性差异。

三、样本外可预测性检验

相对于样本内预测，样本外预测对于现实世界投资者的实时投资决策更有意义。但针对美国市场的实证研究发现，股票收益的样本外可预测性弱且不稳定 (Welch and Goyal, 2008)。Ludvigson and Ng (2009) 等发现主成分预测方法可以在样本外显著预测美国股票和债券市场。本文这一部分使用主成分预测方法研究了中国股票市场的样本外可预测性，并依次介绍了样本外预测回归模型、样本外预测能力检验方法和样本外预测的实证发现。

(一) 计量方法

我们采用扩展估计法 (Expanding Estimation) 来估计样本外股票超额收益。具体讲，我们把 $r_{i,t}$ 和 $x_{j,t}$ 的 T 个总样本观测值分成由 n_1 个观测值组成的样本内估计期和由 n_2 个观测值组成的样本外检验期两部分。预测变量 $x_{j,t}$ 对股票超额收益 $r_{i,t+1}$ 的初始样本外预测是：

$$\hat{r}_{i,n_1+1} = \hat{a}_{i,j,n_1} + \hat{b}_{i,j,n_1} x_{j,n_1} \quad (2)$$

把 $\{r_{i,t+1}\}_{t=1}^{n_1-1}$ 对常数项和 $\{x_{j,t}\}_{t=1}^{n_1-1}$ 做线性回归可以得到的公式 (2) 中的 $a_{i,j}$ 和 $b_{i,j}$ 的最小二乘法 (OLS) 估计 \hat{a}_{i,j,n_1} 和 \hat{b}_{i,j,n_1} 。下一期的样本外预测是：

$$\hat{r}_{i,n_1+2} = \hat{a}_{i,j,n_1+1} + \hat{b}_{i,j,n_1+1} x_{j,n_1+1} \quad (3)$$

把 $\{r_{i,t+1}\}_{t=1}^{n_1}$ 对常数项和 $\{x_{j,t}\}_{t=1}^{n_1}$ 做回归可以得到 \hat{a}_{i,j,n_1+1} 和 \hat{b}_{i,j,n_1+1} 。以此类推，一共得到 n_2 个 $x_{j,t}$ 对 $r_{i,t+1}$ 的样本外预测值： $\{\hat{r}_{i,t+1}\}_{t=n_1}^{T-1}$ 。样本外预测模拟了现实世界的实时预测。

除了单变量预测性回归模型外，我们还研究了主成分预测性回归模型对中国股票超额收益的样本外预测能力。主成分预测方法能用少数几个主成分概括大量关联潜在预测变量的主要信息。首先，我们估计出 $\{x_s\}_{s=1}^t$ 的主成分 $\{\hat{z}_{s,t}\}_{s=1}^t$ ，其中 $x_s = (x_{1,s}, \dots, x_{J,s})'$ 是 J 维预测变量， $\hat{z}_{s,t} = (\hat{z}_{1,s,t}, \hat{z}_{2,s,t})'$ 是用到 t 期为止的信息估计的第 s 期的前两个主成分 (Rapach, Strauss, Tu and Zhou, 2011)。接着，我们使用前两个主成分对 $r_{i,t+1}$ 进行样本外预测：

$$\hat{r}_{i,t+1} = \hat{a}_{i,t} + \hat{b}'_{i,t} \hat{z}_{t,t} \quad (4)$$

$\hat{a}_{i,t}$ 和 $\hat{b}_{i,t}$ 是 $\{r_{i,s}\}_{s=2}^t$ 对常数项和 $\{\hat{z}_{s,t}\}_{s=1}^{t-1}$ 做线性回归得到的 OLS 估计。我们最终得到 n_2 个股票超额收益的样本外主成分预测： $\{\hat{r}_{i,t+1}\}_{t=n_1}^{T-1}$ 。

在中国股票市场没有可预测性的原假设下，可以用带漂移项的随机游走模型刻画中国股票市场超额收益。我们因此选择历史平均收益作为样本外预测基准： $\bar{r}_{i,t+1} = \frac{1}{t} \sum_{s=1}^t r_{i,s}$ (Welch and Goyal, 2008)。如果潜在预测变量包含的信息能预测中国股票市场超额收益，那么使用该变量的单变量预测性回归模型和主成分预测模型应该比历史平均收益有更好的样本外预测能力，因为后者忽略了有用的预测信息。

^③ 需要规模、面值市值比和股权集中度等成分投资组合的样本内预测详细结果的读者可来函索取。

我们使用 Campbell and Thompson (2008) 的 R_{OS}^2 统计量来检验 $\hat{r}_{i,t+1}$ 的样本外预测表现。 R_{OS}^2 统计量测度了使用预测性回归模型替代历史平均收益带来的均方预测误差减少量:

$$R_{OS}^2 = 1 - \frac{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \hat{r}_{i,n_1+k})^2}{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \bar{r}_{i,n_1+k})^2} \quad (5)$$

当 $R_{OS}^2 > 0$ 时, 按照均方预测误差 (MSPE) 标准, $\hat{r}_{i,t+1}$ 预测优于 $\bar{r}_{i,t+1}$ 预测。

我们使用 Clark and West (2007) *MSPE-adjusted* 统计量对 R_{OS}^2 统计量进行假设检验。原假设是预测性回归模型预测的均方预测误差和历史平均收益的均方预测误差相等, 即 $R_{OS}^2 = 0$; 备择假设 (单侧) 是预测性回归模型预测的均方预测误差比历史平均收益预测的均方预测误差小, 即 $R_{OS}^2 > 0$ 。我们选择 1996 年 7 月—2001 年 12 月作为样本内估计期, 并选取 2002 年 1 月—2009 年 6 月作为样本外检验期, 样本外检验期又被划分为市场比较平稳的 2002 年—2006 年和市场泡沫比较大的 2007 年—2009 年两段子时期。

表 2 行业投资组合样本外预测结果

| 组合 | D/P | D/Y | D/E | SVR | E/P | B/M | INF | NTI | TO | M0G | M1G | M2G | PCF |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| MKT | 0.30 | 0.98 | 1.84 | -2.01 | -4.78 | -7.13 | 0.76 | 1.23 | 7.79* | 2.05* | 2.61* | 2.97* | 9.30* |
| AGR | -2.04 | -1.88 | 0.45 | -1.22 | -6.48 | -7.54 | 0.43 | -0.23 | 4.90 | 1.40* | 0.62 | -0.82 | 4.00* |
| MNS | -1.07 | -0.97 | -2.99 | -1.88 | -4.34 | -7.60 | 0.77 | 2.08 | 3.48* | 1.01 | 1.85* | 3.26* | 3.67* |
| MAN | 1.34 | 2.08 | 1.69 | -1.05 | -2.78 | -4.48 | 0.04 | 1.41 | 6.11* | 1.76* | 2.41* | 1.96* | 8.16* |
| UTL | -1.19 | -1.37 | 0.77 | -1.07 | -3.84 | -4.89 | 1.01 | 0.29 | 4.91* | 1.96* | 1.46* | 0.66 | 4.41* |
| CNT | 2.88* | 3.04* | 1.53* | 0.74 | 0.60 | -0.04 | -0.55 | 5.49* | -0.49 | 2.08* | 1.29* | -1.87 | 4.17* |
| TRS | -3.44 | -3.71 | 0.45 | -3.46 | -8.68 | -10.9 | 1.40 | -1.07 | 8.95* | 0.85 | 1.26* | 1.03 | 6.16* |
| INF | -5.30 | -6.14 | 0.77 | -2.02 | -13.3 | -19.3 | -2.22 | -1.20 | 6.99* | 0.36 | 2.14* | 1.01 | 4.70* |
| WRS | 1.35 | 2.07 | 0.17 | -0.81 | -2.06 | -3.96 | -0.33 | 2.64 | 5.63* | 1.73* | 1.74* | 1.56* | 7.57* |
| FIN | 1.39 | 2.87* | 2.37 | -4.52 | -2.50 | -4.38 | 1.99 | -2.65 | 5.69* | 0.62 | 0.58* | 2.59 | 6.84* |
| PRT | 3.00* | 3.74* | 1.34 | -1.19 | 0.06 | -1.77 | -1.27 | 0.56 | 9.44* | 0.94 | 2.88* | 2.69* | 11.37* |
| SVC | 3.17* | 3.55* | -0.09 | -1.69 | -0.04 | -2.24 | -1.09 | 2.38 | 6.11* | 1.44 | 1.32* | 1.25 | 8.53* |
| MED | -0.01 | 0.37 | 0.08 | -0.22 | -4.33 | -8.75 | -0.15 | 2.41* | 0.98 | 0.55 | 1.34* | 0.73 | 2.84* |
| MUT | 0.14 | 0.26 | 0.38 | -0.72 | -4.62 | -8.46 | -0.65 | 0.42 | 5.74* | 1.47 | 2.43* | 1.38* | 8.09* |
| #Sig | 3 | 4 | 1 | 0 | 0 | 0 | 0 | 2 | 10 | 5 | 12 | 5 | 13 |
| \bar{R}_{OS}^2 | 0.02 | 0.30 | 0.53 | -1.47 | -4.03 | -6.50 | -0.05 | 0.96 | 5.26 | 1.24 | 1.64 | 1.19 | 6.19 |

注: 本表报告了行业投资组合超额收益的预测性回归 Campbell and Thompson (2008) 样本外 R^2 统计量 R_{OS}^2 (百分比)。行变量是行业投资组合超额收益, 列变量是经济变量。MKT 行是市场投资组合的样本外预测 R_{OS}^2 统计量。PCF 列是基于 12 个经济变量的主成分预测的 R_{OS}^2 统计量。“*”代表 R_{OS}^2 根据 Clark and West (2007) *MSPE-adjusted* 统计量在 5% 水平下显著。# Sig 行是每一列中在 5% 水平显著的行业数量。 \bar{R}_{OS}^2 行是每一列除了 MKT 外的 R_{OS}^2 统计量的平均值。

(二) 市场投资组合

表 2 的 MKT 行是使用 12 个经济变量对市场投资组合超额收益进行样本外预测的 R_{OS}^2 统计量 (百分比)。TO、M0G、M1G 和 M2G 等 4 个经济变量以及主成分预测 (PCF 列) 的 R_{OS}^2 在 5% 水平显著, 对市场投资组合有显著的样本外预测能力。主成分预测对市场投资组合的样本外预测 R_{OS}^2 高达 9.30%, 在经济上是相当大的。从表 3 的 A 部分可见, 市场投资组合在泡沫较大期更容易预测, 泡沫期 2007 年—2009 年的主成分预测 R_{OS}^2 约为平稳期 2002 年—2006 年的 3 倍。

TO、M0G、M1G 和 M2G 等 4 个经济变量对中国股票市场有显著且稳定的预测能力。考虑到中国股票市场可能并不是非常有效, 换手率 (TO) 有很强的预测能力是很容易理解的。Miller (1977) 指出, 当存在卖空约束时, 投资者的异质信念会导致股票高估。由于中国股票市场发展时间短、不允许卖空和市场主体主要由缺乏投资经验的非理性个人投资者构成, 这些因素使得 Miller (1977) 的理论在中国股票市场尤其适用^④。M0G、M1G 和 M2G 等货币政策变量对中国

^④ Baker and Stein (2004)、Wang (1994)、Wang and Cheng (2004)、陈国进、张贻军 (2009) 和郑方镰、

股票市场的预测能力可能源于货币政策传导的信贷渠道（Bernanke and Gertler, 1995）^⑤。银行信贷在中国融资渠道中起着重要地位，宽松货币政策导致廉价银行信贷过多，继而影响股票收益。

（三）行业投资组合

表 2 的第二行到最后一行报告了行业投资组合超额收益的样本外预测结果。和市场投资组合类似，TO、MOG、M1G 和 M2G 等 4 个经济变量对行业投资组合也有很强的样本外预测能力，其中，TO 对 PRT 和 TRS 这两个行业的 R_{OS}^2 分别高达 9.44% 和 8.95%。主成分预测（PCF 列）进一步提高了单个经济变量的预测能力，对所有的行业都有显著的样本外预测能力， R_{OS}^2 最高达到 11.37%。不同行业的可预测性差异很大，其中 MAN、TRS、WRS、FIN、PRT、SVC 和 MUT 等 7 个行业的主成分预测 R_{OS}^2 统计量大于 6.00%，经济上是相当可观的；而 MNS（3.67%）和 MED（2.84%）等行业的可预测性则小很多。类似市场投资组合，行业投资组合也在泡沫较大期更容易预测，泡沫期 2007 年—2009 年的平均主成分预测 R_{OS}^2 约为平稳期 2002 年—2006 年的 2.5 倍（表 3 的 B 部分）。

表 3 相对平稳期 2002 年—2006 年和泡沫较大期 2007 年—2009 年的 R_{OS}^2 统计量

| 组合 | 02-06 | 07-09 | 02-09 | 组合 | 02-06 | 07-09 | 02-09 | 组合 | 02-06 | 07-09 | 02-09 |
|------------------|-------|-------|-------|-----|-------|-------|-------|------|-------|-------|-------|
| A: 市场投资组合超额收益 | | | | | | | | | | | |
| MKT | 4.31 | 11.82 | 9.30 | | | | | | | | |
| B: 行业投资组合超额收益 | | | | | | | | | | | |
| AGR | 5.10 | 3.48 | 4.00 | TRS | -1.37 | 8.47 | 6.16 | SVC | 5.81 | 9.59 | 8.53 |
| MNS | 1.26 | 5.10 | 3.67 | INF | -0.49 | 8.39 | 4.70 | MED | -1.64 | 5.98 | 2.84 |
| MAN | 3.72 | 10.11 | 8.16 | WRS | 4.66 | 9.17 | 7.57 | MUT | 2.55 | 10.37 | 8.09 |
| UTL | 0.04 | 6.12 | 4.41 | FIN | 4.06 | 9.95 | 6.84 | 平均 | 2.41 | 8.17 | 6.19 |
| CNT | 2.16 | 5.03 | 4.17 | PRT | 5.49 | 14.38 | 11.37 | | | | |
| C: 规模投资组合超额收益 | | | | | | | | | | | |
| S1 | 8.55 | 9.49 | 9.15 | S5 | 3.69 | 9.57 | 7.79 | S9 | 2.83 | 11.06 | 8.70 |
| S2 | 5.73 | 9.71 | 8.41 | S6 | 5.19 | 10.95 | 9.19 | S10 | 5.20 | 10.26 | 8.38 |
| S3 | 5.49 | 10.17 | 8.67 | S7 | 4.03 | 10.41 | 8.33 | 平均 | 4.91 | 10.13 | 8.49 |
| S4 | 4.75 | 10.01 | 8.39 | S8 | 3.62 | 9.62 | 7.92 | | | | |
| D: 面值市值比投资组合超额收益 | | | | | | | | | | | |
| BM1 | 3.91 | 10.75 | 8.36 | BM5 | 3.56 | 10.47 | 8.31 | BM9 | 2.71 | 9.76 | 7.48 |
| BM2 | 4.65 | 10.90 | 8.86 | BM6 | 4.80 | 9.68 | 7.76 | BM10 | -0.21 | 11.64 | 7.55 |
| BM3 | 5.80 | 12.21 | 10.30 | BM7 | 0.12 | 10.57 | 7.70 | 平均 | 3.56 | 10.72 | 8.42 |
| BM4 | 6.61 | 9.95 | 8.63 | BM8 | 3.66 | 11.31 | 9.23 | | | | |
| E: 股权集中度投资组合超额收益 | | | | | | | | | | | |
| OC1 | 6.38 | 13.64 | 11.01 | OC5 | 3.83 | 12.73 | 9.39 | OC9 | 3.13 | 10.09 | 8.13 |
| OC2 | 2.12 | 11.51 | 8.54 | OC6 | 3.27 | 7.58 | 6.21 | OC10 | 5.58 | 8.21 | 7.21 |
| OC3 | 1.63 | 12.71 | 8.83 | OC7 | 4.94 | 10.59 | 9.03 | 平均 | 3.55 | 10.91 | 8.51 |
| OC4 | 3.30 | 11.60 | 8.95 | OC8 | 1.36 | 10.40 | 7.79 | | | | |

注：本表报告了市场（A 部分）、行业（B 部分）、规模（C 部分）、面值市值比（D 部分）和股权集中度（E 部分）等投资组合分别在相对平稳期 2002 年—2006 年和泡沫较大期 2007 年—2009 年的主成分预测 Campbell and Thompson（2008）样本外 R^2 统计量 R_{OS}^2 。02-06 列是相对平稳期 2002 年—2006 年的主成分预测 R_{OS}^2 统计量；07-09 列是泡沫较大期 2007 年—2009 年的主成分预测 R_{OS}^2 统计量；02-09 列是全样本的主成分预测 R_{OS}^2 统计量。平均行是每一部分 R_{OS}^2 统计量的平均值。

（四）其他成分投资组合

我们发现规模、面值市值比和股权集中度等成分投资组合也都有很强的样本外可预测性（表

吴超鹏和吴世农（2007）等也从不同的角度研究了中国股票市场交易量和股票收益之间的关系。

^⑤ 王晓明（2010）和易纲、王召（2002）等分析了中国货币、信贷和资产价格之间的关系。

3 的 C、D 和 E 部分)^⑥。TO、M0G、M1G 和 M2G 等 4 个经济变量对这三类成分投资组合都有统计上和经济上显著的样本外预测能力。主成分预测的 R_{OS}^2 统计量基本都在 7% 以上, 比几乎所有的单个经济变量的 R_{OS}^2 都要高, 主成分预测进一步改进了成分投资组合的样本外可预测性。不同规模、面值市值比和股权集中度投资组合的样本外可预测性存在显著差异, 小市值、低面值市值比和低股权集中度的投资组合的可预测性最强。最后, 规模、面值市值比和股权集中度等成分投资组合也在泡沫期更容易预测, 泡沫较大期 2007 年—2009 年的平均主成分预测 R_{OS}^2 约为平稳期 2002 年—2006 年的 3 倍。

四、成分投资组合可预测性的经济解释

我们使用条件 CAPM 模型把第三部分得到的样本外 R_{OS}^2 统计量分解成条件资产定价模型可预测性 $R_{OS,R}^2$ 和 alpha 可预测性 $R_{OS,\alpha}^2$ 两部分。我们还研究了信息流动摩擦对行业投资组合样本外可预测性差异的解释能力。

(一) 样本外可预测性分解

我们用 Ferson and Korajczyk (1995) 的研究框架得到被条件资产定价模型捕捉的时变系统性风险溢价可预测性, 余下的可预测性属于 alpha 可预测性, alpha 可预测性反应了资产定价模型的定价误差。本文使用条件 CAPM 模型把公式 (5) 的 R_{OS}^2 统计量分解为条件资产定价模型可预测性 $R_{OS,R}^2$ 和 alpha 可预测性 $R_{OS,\alpha}^2$ 两部分。

假设成分投资组合 i 的超额收益服从如下模型:

$$r_{i,t} = \alpha_i(x_{t-1}) + \beta_i' f_t + \epsilon_{i,t} \quad (6)$$

其中, x_{t-1} 是 J 维滞后预测变量, f_t 是 K 维基于投资组合的系统性风险因子, β_i 是 K 维投资组合 beta 系数, $\epsilon_{i,t}$ 是扰动项。进一步假设:

$$f_t = \lambda(x_{t-1}) + u_t \quad (7)$$

其中, $\lambda(x_{t-1})$ 是 K 维基于投资组合的风险因子的条件预期收益, u_t 是扰动项 (和 $\epsilon_{i,t}$ 独立)。系统性风险溢价 λ 受 x_{t-1} 影响是时变的。我们从条件资产定价模型得到 $r_{i,t}$ 的条件期望^⑦:

$$E(r_{i,t}|x_{t-1}) = \beta_i' E(f_t|x_{t-1}) = \beta_i' \lambda(x_{t-1}) \quad (8)$$

根据条件资产定价模型, $r_{i,t}$ 的可预测性源于风险因子可预测性和 $r_{i,t}$ 的风险暴露, 即 $\beta_i' \lambda(x_{t-1})$, 故而 $\alpha_i(x_{t-1}) = 0$ 。 $r_{i,t}$ 的可预测性中不能被 $\beta_i \lambda(x_{t-1})$ 解释的部分属于 alpha 可预测性, 因此 $\alpha_i(x_{t-1}) \neq 0$ 。如果公式 (8) 充分捕捉了系统风险, 则 $\alpha_i(x_{t-1}) \neq 0$ 就源于成分投资组合 i 的定价误差。

为了解 R_{OS}^2 , 我们需要计算出 $r_{i,t}$ 基于条件资产定价模型约束的主成分预测 $\hat{r}_{i,t}^R$ 。先用 $\{r_{i,s}\}_{s=1}^{t-1}$ 对 $\{f_s\}_{s=1}^{t-1}$ 做无截距项线性回归来估计出 β_i , 记为 $\hat{\beta}_{i,t-1}$ 。基于第三部分介绍的样本外主成分预测方法, 用预测变量 x_{t-1} 对 f_t 的主成分预测估计出 $\lambda(x_{t-1})$, 记为 $\hat{\lambda}(x_{t-1})$ 。最后把 $\hat{\beta}_{i,t-1}$ 和 $\hat{\lambda}(x_{t-1})$ 插入公式 (8) 就得到 $r_{i,t}$ 条件资产定价模型约束的主成分预测:

$$\hat{r}_{i,t}^R = \hat{\beta}_{i,t-1}' \hat{\lambda}(x_{t-1}) \quad (9)$$

我们把第三部分得到的 $r_{i,t}$ 的主成分预测标记为 $\hat{r}_{i,t}$ 。 $\hat{r}_{i,t}$ 是没有附加条件资产定价模型约束的无约束的主成分预测, 它既包含条件资产定价模型可预测性也包含 alpha 可预测性。

我们把 R_{OS}^2 统计量分解成 $R_{OS,R}^2$ 和 $R_{OS,\alpha}^2$ 两个子统计量。第一个子统计量 $R_{OS,R}^2$ 测度了条件资产定价模型约束的主成分预测相对于历史平均收益的均方预测误差减少量:

$$R_{OS,R}^2 = 1 - \frac{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \hat{r}_{i,n_1+k}^R)^2}{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \bar{r}_{i,n_1+k})^2} \quad (10)$$

^⑥ 需要规模、面值市值比和股权集中度等成分投资组合的样本外预测检验详细结果的读者可来函索取。

^⑦ 实证研究发现风险溢价 λ 的时变性显著大于 β_i 的时变性, 可参考 Ferson and Korajczyk (1995) 等, 故而本文模型假定 β_i 是常数。不过, 本文使用的递归估计允许 β_i 有一定的时变性。

$R_{OS,R}^2$ 反应了成分投资组合 i 的条件资产定价模型可预测性。第二个子统计量 $R_{OS,\alpha}^2$ 测度了无约束主成分预测相对于条件资产定价模型约束的主成分预测的均方预测误差减少量:

$$R_{OS,\alpha}^2 = 1 - \frac{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \hat{r}_{i,n_1+k})^2}{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \hat{r}_{i,n_1+k}^R)^2} \quad (11)$$

$R_{OS,\alpha}^2$ 代表了成分投资组合 i 的 alpha 可预测性, 即没有被条件资产定价模型捕捉的那部分可预测性。从公式 (5)、(10) 和 (11) 可以得到:

$$R_{OS,\alpha}^2 = 1 - \left[\frac{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \hat{r}_{i,n_1+k})^2}{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \bar{r}_{i,n_1+k})^2} \right] \left[\frac{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \bar{r}_{i,n_1+k})^2}{\sum_{k=1}^{n_2} (r_{i,n_1+k} - \hat{r}_{i,n_1+k}^R)^2} \right] = 1 - \left(\frac{1 - R_{OS}^2}{1 - R_{OS,R}^2} \right) \quad (12)$$

从公式 (12) 我们得到:

$$R_{OS}^2 = R_{OS,R}^2 + R_{OS,\alpha}^2 - R_{OS,R}^2 R_{OS,\alpha}^2 \quad (13)$$

当 $R_{OS,R}^2$ 和 $R_{OS,\alpha}^2$ 都很小, 交叉乘积项约等于零, 故而

$$R_{OS}^2 \approx R_{OS,R}^2 + R_{OS,\alpha}^2 \quad (14)$$

投资组合的可预测性 R_{OS}^2 被近似分解成条件资产定价模型可预测性 $R_{OS,R}^2$ 和 alpha 可预测性 $R_{OS,\alpha}^2$ 两部分。

样本外检验更适用于资产定价模型检验。样本内检验存在著名的“过度拟合”的问题[®], 我们总是可以通过增加解释因子的数量的方法来提高模型的样本内解释能力。因此, 某些缺乏经济学依据的新增因子在样本内也可能是统计上显著的。使用样本外检验可以降低“过度拟合”风险。因此, 样本外分解比样本内分解更适合用于推断资产定价模型对成分投资组合可预测性的解释能力。

我们检验了条件 CAPM 模型对成分投资组合样本外可预测性的解释能力。从公式 (9) 可以得到条件 CAPM 模型约束的主成分预测 $\hat{r}_{i,t}^{CAPM} = \hat{\beta}_{i,t-1}^{CAPM} \hat{\lambda}^{CAPM}(x_{t-1})$, 其中 $\hat{\beta}_{i,t-1}^{CAPM}$ 是成分投资组合 i 的 beta 系数, $\hat{\lambda}^{CAPM}(x_{t-1}) = \hat{r}_{MKT,t}$ 是市场投资组合的主成分预测。

表 4 基于条件 CAPM 模型的 $R_{OS,R}^2$ 和 $R_{OS,\alpha}^2$ 统计量

| 组合 | $R_{OS,R}^2$ | $R_{OS,\alpha}^2$ | 组合 | $R_{OS,R}^2$ | $R_{OS,\alpha}^2$ | 组合 | $R_{OS,R}^2$ | $R_{OS,\alpha}^2$ | 组合 | $R_{OS,R}^2$ | $R_{OS,\alpha}^2$ |
|------------------|--------------|-------------------|-----|--------------|-------------------|-----|--------------|-------------------|------|--------------|-------------------|
| A: 行业投资组合超额收益 | | | | | | | | | | | |
| AGR | 4.25* | -0.25 | CNT | 7.53* | -3.64 | FIN | 5.88* | 1.02 | WRS | 7.98* | -0.44 |
| MNS | 6.23* | -2.72 | TRS | 8.20* | -2.23 | PRT | 9.56* | 2.00* | MED | 4.70* | -1.94 |
| MAN | 8.24* | -0.10 | INF | 7.05* | -2.53 | SVC | 8.12* | 0.45 | MUT | 8.46* | -0.41 |
| UTL | 6.32* | -2.04 | | | | | | | | | |
| B: 规模投资组合超额收益 | | | | | | | | | | | |
| S1 | 11.93* | -3.16 | S4 | 9.40* | -1.11 | S7 | 7.92* | 0.44 | S9 | 8.30* | 0.45 |
| S2 | 10.25* | -2.05 | S5 | 8.76* | -1.06 | S8 | 7.74* | 0.19 | S10 | 8.82* | -0.48 |
| S3 | 9.95* | -1.42 | S6 | 9.13* | 0.07 | | | | | | |
| C: 面值市值比投资组合超额收益 | | | | | | | | | | | |
| BM1 | 7.67* | 0.74 | BM4 | 8.50* | 0.14 | BM7 | 8.19* | -0.54 | BM9 | 7.98* | -0.54 |
| BM2 | 8.22* | 0.70* | BM5 | 9.36* | -1.16 | BM8 | 9.37* | -0.16 | BM10 | 8.48* | -1.01 |
| BM3 | 9.51* | 0.88* | BM6 | 9.03* | -1.39 | | | | | | |
| D: 股权集中度投资组合超额收益 | | | | | | | | | | | |
| OC1 | 10.04* | 1.08* | OC4 | 8.54* | 0.45 | OC7 | 8.78* | 0.28 | OC9 | 9.09* | -1.06 |
| OC2 | 8.89* | -0.38 | OC5 | 9.29* | 0.11 | OC8 | 8.23* | -0.48 | OC10 | 7.29* | -0.09 |
| OC3 | 9.78* | -1.05 | OC6 | 7.90* | -1.83 | | | | | | |

注: 本表报告了行业 (A 部分)、规模 (B 部分)、面值市值比 (C 部分) 和股权集中度 (D 部分) 投资组合的基于条件 CAPM 模型的 $R_{OS,R}^2$ 和 $R_{OS,\alpha}^2$ 统计量。 $R_{OS,R}^2$ ($R_{OS,\alpha}^2$) 测度了条件 CAPM 模型约束的主成分预测相对历史平均收益预测 (无约束的主成分预测相

[®] 比如 Lo and MacKinlay (1990) 对资产定价模型中的“Data-Snooping”问题的研究。

对条件 CAPM 模型约束主成分预测)均方预测误差减少量。“*”代表根据 Clark and West (2007) *MSPE-adjusted* 统计量在 5%水平下显著。

表 4 报告了条件 CAPM 模型约束的主成分预测的 $R_{OS,R}^2$ 统计量和 $R_{OS,\alpha}^2$ 统计量。13 个行业、10 个规模、10 个面值市值比和 10 个股权集中度投资组合的 $R_{OS,R}^2$ 统计量都显著为正。这说明条件 CAPM 模型约束的主成分预测相对于历史平均收益可以在样本外显著预测所有的成分投资组合。 $R_{OS,\alpha}^2$ 统计量进一步检验了条件 CAPM 模型约束的主成分预测能否充分解释各成分投资组合样本外可预测性。从表 4 可见, 13 个 (10 个; 10 个; 10 个) 个行业 (规模; 面值市值比; 股权集中度) 投资组合中只有 1 个 (0 个; 2 个; 1 个) 投资组合的 $R_{OS,\alpha}^2$ 显著为正的, 有 alpha 可预测性。因此, 我们认为条件 CAPM 模型可以充分解释各成分投资组合的样本外可预测性。

房地产业 (PRT) 是 13 个行业投资组合中唯一一个有 alpha 可预测性的行业, 这可能和房地产泡沫有关。中国城市住房市场在 2002 年—2009 年的样本外检验期的中后期持续存在严重的泡沫, 可能导致房地产业股票价格非理性波动, 故而房地产业有 alpha 可预测性。低股权集中度股票的 alpha 可预测可能源于中国股票市场目前似乎仍或多或少是一个政策市, 政府政策是重要的系统性风险因素 (鲁臻、邹恒甫, 2007)。中国大部分上市公司是国有控股企业, 它们更有可能去追求政治目标, 故而股权集中度高的公司运营更有可能受政策影响 (Sun and Tong, 2003)。Morck, Yeung, and Yu (2000) 也指出, 不重视保护外部股东利益的公司股价主要仅受市场因素影响, 而重视保护外部股东利益的公司股价还会受到公司基本面的影响。故而, 股权集中度高的公司 (国有企业为主) 倾向于只有系统性风险 (政策) 可预测性; 而股权集中度最低的企业 (私营企业为主) 更重视商业目标, 除了有系统性风险可预测性还有 alpha 可预测性。

根据条件 CAPM 模型, 高市场风险暴露投资组合的条件预期收益的波动也会比较大。因此高市场风险暴露投资组合有更强的条件资产定价模型可预测性, 即高 $R_{OS,R}^2$ 。表 5 分别报告了行业、规模、面值市值比和股权集中度等成分投资组合的 $R_{OS,R}^2$ 对 $\hat{\beta}_i^{CAPM}$ 的平均值 $\bar{\beta}_i^{CAPM}$ 做线性回归的回归系数、White (1980) 异方差调整的 t 统计量和回归 R^2 统计量。从表 5 可见, 这四组成分投资组合的 $R_{OS,R}^2$ 统计量都和市场风险暴露 $\hat{\beta}^{CAPM}$ 正相关, 规模和股权集中度投资组合的回归系数分别在 1% 和 5% 的水平下显著为正, 且线性回归的 R^2 统计量最高达到 50.34%。

表 5 $R_{OS,R}^2$ 对 $\bar{\beta}^{CAPM}$ 回归结果

| 投资组合 | 截距 | $\bar{\beta}^{CAPM}$ | R^2 |
|-------|-------|----------------------|--------|
| 行业 | -2.12 | 8.99 | 18.96% |
| | -0.34 | 1.45 | |
| 规模 | -3.49 | 11.86 | 50.34% |
| | -0.78 | 2.85 | |
| 面值市值比 | 3.28 | 5.15 | 11.82% |
| | 0.61 | 0.97 | |
| 股权集中度 | 1.65 | 6.98 | 41.25% |
| | 0.55 | 2.37 | |

总之, 样本外可预测性分解证明, 条件 CAPM 模型捕捉的时变系统风险溢价可以解释行业、规模、面值市值比和股权集中度等成分投资组合的绝大部分样本外可预测性^⑨。回归分析发现, 高系统性风险暴露的成分投资投资组合有较高的样本外可预测性。

(二) 样本外可预测性与行业集中度

受 Hong, Torous, and Valkanov (2007) 的信息流动摩擦理论启发, 我们还研究了行业投资

^⑨ 我们还分析了 Fama-French 三因子模型约束的主成分预测的解释能力, 我们发现时变市场风险溢价仍然可以显著解释各成分投资组合的样本外可预测性, 但规模和面值市值比因子的解释能力并不强。

组合的样本外可预测性与行业集中度的关系。根据信息流动摩擦理论，投资者更容易获得那些由有限几个大公司构成的行业的信息，因此高集中度行业的可预测性差；主要由中小企业构成的行业的信息流动摩擦高，投资者难以获得这类行业的信息，故而低集中度的行业有较强的可预测性。公式（15）报告了行业投资组合主成分预测 R_{OS}^2 对行业集中度 IC 做线性回归的估计结果：

$$R_{OS}^2 = 11.15 - 0.07IC, \quad R^2 = 40.78\% \quad (15)$$

[5.92] [-2.75]

其中，行业集中度 IC 是行业前8大公司的收益占该行业总收益的百分比的平均值。行业投资组合的样本外可预测性和行业集中度显著负相关，而且回归 R^2 统计量高达40.78%。故而，信息流动摩擦可以解释中国行业投资组合的样本外可预测性差异。

五、结论

我们检验了文献中提出的众多预测变量对中国市场投资组合以及行业、规模、面值市值比和股权集中度等成分投资组合的预测能力。我们发现，市场投资组合和各种成分投资组合都有显著的样本内和样本外可预测性，不同的成分投资组合的可预测性存在明显的差异。具体说，部分行业、小市值、低面值市值比和低股权集中度的股票有更强的可预测性。

我们还研究了中国成分投资组合间的样本外可预测性差异的经济原因。我们使用基于条件CAPM模型的样本外可预测性分解把成分投资组合的样本外可预测性分解为条件资产定价模型可预测性和 α 可预测性两部分。本文发现，条件CAPM模型约束的主成分预测可以充分解释成分投资组合的大部分样本外可预测性。高系统性风险暴露的成分投资组合有较高的样本外可预测性。并且，行业投资组合的样本外可预测性和行业集中度显著负相关，这与Hong, Torous, and Valkanov (2007)的信息流动摩擦理论是一致的。总之，本文发现系统性风险溢价和信息流动摩擦都有助于解释中国成分投资组合间的样本外可预测性差异。

本文还可以在多个方面进一步展开。比如，本文考察了大量经济变量对中国股票市场的预测能力，但这些变量主要来源于针对美国股票市场的文献。相对美国等成熟股票市场，中国个人投资者直接参与股票投资的比例更高、交易更频繁。因此，研究比如新增活跃交易账户和个人投资者情绪等有中国特色的经济变量的预测能力也是非常意义的，我们将在后续研究中解决这些问题。

参考文献

- [1] 陈国进、张贻军，2009：《异质信念、卖空限制与我国股市的暴跌现象研究》，《金融研究》第4期，80-91。
- [2] 鲁臻、邹恒甫，2007：《中国股市的惯性与反转效应研究》，《经济研究》第9期，145-155。
- [3] 王晓明，2010：《银行信贷与资产价格的顺周期关系研究》，《金融研究》第3期，45-55。
- [4] 易纲、王召，2002：《货币政策与金融资产价格》，《经济研究》第3期，13-20。
- [5] 郑方镛、吴超鹏和吴世农，2007：《股票成交量与收益率序列相关性研究——来自中国股市的实证证据》，《金融研究》第3期，140-150。
- [6] Baker, M., and J.C. Stein, 2004, "Market Liquidity as A Sentiment Indicator", *Journal of Financial Markets* 7, 271-299.
- [7] Bernanke, B.S., and M. Gertler, 1995, "Inside the Black Box: The Credit Channel of Monetary Policy Transmission", *Journal of Economic Perspectives* 9, 27-48.
- [8] Campbell, J.Y., 2000, "Asset Pricing at the Millennium", *Journal of Finance* 55, 1515-1567.
- [9] Campbell, J.Y., and S.B. Thompson, 2008, "Predicting the Equity Premium Out Of Sample: Can Anything Beat the Historical Average?", *Review of Financial Studies* 21, 1509-1531.

- [10] Chen, X., K.A. Kim, T. Yao, T. Yu, 2010, “On the Predictability of Chinese Stock Returns”, *Pacific-Basin Finance Journal* 18, 403–25.
- [11] Clark, T.E., and K.D. West, 2007, “Approximately Normal Tests for Equal Predictive Accuracy in Nested Models”, *Journal of Econometrics* 138, 291-311.
- [12] Cochrane, J.H., 2011, “Presidential Address: Discount Rates”, *Journal of Finance* 66, 1047-1108.
- [13] Ferson, W. E., and R. A. Korajczyk, 1995, “Do Arbitrage Pricing Models Explain the Predictability of Stock Returns”, *Journal of Business* 68, 309-349.
- [14] Hong, H., W. Torous, and R. Valkanov, 2007, “Do Industries Lead Stock Markets?”, *Journal of Financial Economics* 83, 367-396.
- [15] Lo, A., A.C. MacKinlay, 1990, “Data-snooping Biases in Tests of Financial Asset Pricing Models”, *Review of Financial Studies* 3, 431-467.
- [16] Ludvigson, S. C., and S. Ng, 2009, “Macro factors in bond risk premia”, *Review of Financial Studies* 22, 5027-5067.
- [17] Miller, E. M., 1977, “Risk, Uncertainty, and Divergence of opinion”, *Journal of Finance* 32, 1151-1168.
- [18] Morck, R., B. Yeung, and W. Yu, 2000, “The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements?”, *Journal of Financial Economics* 58, 215-260.
- [19] Rapach, D. E., J. K. Strauss, J. Tu and G. Zhou, 2011, “Are Industry Returns Predictable? An Out-of-sample Analysis with a Large Number of Predictors”, Working Paper, Singapore Management University.
- [20] Sun, Q., W. H.S. Tong, 2003, “China Share Issue Privatization: The Extent of Its Success”, *Journal of Financial Economics* 70, 183-222.
- [21] Wang, C., and N.S. Cheng, 2004, “Extreme Volumes and Expected Stock Returns: Evidence from China's Stock Market”, *Pacific-Basin Finance Journal* 12, 577-597.
- [22] Wang, J., 1994, “A Model of Competitive Stock Trading Volume”, *Journal of Political Economy* 102, 127-168.
- [23] Welch, I., and A. Goyal, 2008, “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction”, *Review of Financial Studies* 21, 1455-1508.
- [24] White, H. L., 1980, “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”, *Econometrica* 48, 817-838.
- [25] Xu, Y., 2004, “Small Levels of Predictability and Large Economic Gains”, *Journal of Empirical Finance* 11, 247-275.

How Predictable Is the Chinese Stock Market?

Abstract: We analyze return predictability for the Chinese stock market, including the aggregate market portfolio and the components of the aggregate market, such as portfolios sorted on industry, size, book-to-market and ownership concentration. Considering a variety of economic variables as predictors, both in-sample and out-of-sample tests highlight significant predictability in the aggregate market portfolio and substantial differences in return predictability across components. Among industry portfolios, Finance and insurance, Real estate, and Manufacturing exhibit the most predictability, while portfolios of small-cap, low book-to-market ratio and low ownership concentration firms also display considerable predictability. Two key findings provide economic explanations for component predictability: (i) time-varying systematic risk premiums captured by the conditional CAPM model largely account for component predictability; (ii) industry concentration significantly explain differences in return predictability across industries, consistent with the information-flow frictions

emphasized by Hong, Torous, and Valkanov (2007).

Keywords: Component Portfolios; In-Sample Return Predictability; Out-of-Sample Return Predictability; Conditional CAPM; Information-Flow Frictions