

资本市场收益可预测性研究进展^{*}

张春玲 姜富伟 唐国豪

摘要:准确预测资本市场收益对资产管理和金融学术研究都非常重要。从资产管理的角度看,资产配置需要实时进行资产回报预测;从学术研究角度看,深入理解资产收益预测的本质有助于研究者们提出更符合现实、更准确的资产定价模型和理论。如何根据经济或市场所处的不同时期选取合适的预测方法和指标,并研究其影响资产价格变动的机制已经成为近年来的研究热点。本文从收益预测检验与市场有效性的争论出发,总结了资本市场收益(特别是股票收益)可预测性的研究进展。首先,本文介绍了近期研究中发现的多种收益预测指标,包括基本面指标、技术面指标、情绪指标和波动率指标。然后,本文整理了提高收益预测准确性的主要方法,包括加入经济机制约束法、潜在因子法、组合预测法、状态转换法和机器学习法。最后,本文还在比较不同预测方法与指标的基础上说明了如何科学地度量收益预测的准确性,以及在预测中应注意避免的数据过度挖掘等问题。

关键词:资本市场 收益预测 预测指标 资产定价

一、引言

资本市场收益预测是一个经久不衰的经典金融研究问题。正如彩票购买者急切地想提前获知下一期的中奖号码,投资者们也想通过多种方法预知资产收益从而获利。从资产管理的角度看,在资产配置中需要实时地对股票回报进行预测,准确地预测资产收益将使得投资者更易获得高回报。从学术研究角度看,可预测性对于市场有效性的检验具有重要意义,理解收益可预测性的本质也有助于研究者们做出更符合现实、能更好解释数据变化规律的资产定价模型。

当代著名的金融经济学家 John H. Cochrane(2008)指出资产收益(特别是股票收益)预测的应用几乎贯穿了金融学的所有重要研究领域。但因为资产收益内在地包含了大量不能预测的部分,目前的实证和理论研究还处于起步阶段。而且,投资者之间的激烈竞争也意味着当成功的预测模型被广泛采用之后,资产价格会朝着减弱可预测性的方向动态调整,这进一步提高了检测收益预测的难度。因此,目前即使是最好的预测模型也仅能预测未来资产回报的一小部分,如何准确地对收益进行预测是非常具有挑战性的研究课题。

收益可预测性首先是一个实证问题。研究发现,许多经济变量可以在样本内显著预测股票收益,常见的变量包括估值比率、利率利差和通胀等(Campbell, 2000)。随着对收益预测性来源研究的不断深入,大量新的预测指标被发现。有的指标基于宏观经济形势或公司运营状况的基本面信息,如工业原材料价格(Jacobsen et al, 2016)和做空(Rapach et al, 2016);有的指标基于公司股票价格的变化趋势,如趋势因子(Han et al, 2016);有的指标基于非理性的市场情绪,如投资者情绪

^{*} 张春玲,山东工商学院经济学院,邮政编码:264005,电子邮箱:zhangcl3790@163.com;姜富伟,中央财经大学金融学院,邮政编码:100081,电子邮箱:jfuwei@gmail.com;唐国豪,湖南大学金融与统计学院,邮政编码:410006,电子邮箱:ghtag@hnu.edu.cn。基金项目:国家自然科学基金项目(71872195,71602198),北京市自然科学基金项目(9174045)。感谢匿名审稿人的宝贵意见,文责自负。

(Huang et al, 2015)、经理人情绪(Jiang et al, 2018);有的指标基于衡量市场风险的波动率,如投资组合的收益波动率(Atilgan et al, 2015),时变尾部风险(Kelly & Jiang, 2014)。研究者们除了关心不同指标对于收益预测的准确程度外,更关注背后的经济机制与解释。

然而,近期文献发现众多具有样本内预测能力的预测变量不能在样本外预测中起作用(Goyal & Welch, 2008)。这个发现使得研究者对收益可预测性产生了怀疑,但该问题可能是计量方法本身造成的。许多实证研究建模忽视了预测中的模型不确定性和参数不稳定性问题,也就是一个预测者既不知道最好的模型设定也不知道与它相关的唯一的参数值,最好的模型会随时间变化。

因此,围绕资产回报数据生成过程的模型不确定性和参数不稳定性开展收益预测建模正成为实证金融研究的热门问题。许多研究也尝试用新的方法提高收益预测的准确性。Campbell & Thompson(2008)将经济机制约束加入预测回归中,降低了参数的不稳定性。Neely et al(2014)采用潜在因子法提取了基本面和技术面变量中能有效预测收益的公共部分,从而在月度层面能很好地预测股票市场收益。Rapach et al(2010)使用组合预测法,赋予不同模型的估计收益不同权重,使得那些在近期能够更准确预测的部分占比更高。Green et al(2016)使用了机器学习的相关方法从上百个金融变量中构造了不同维度的收益预测信号。Zhang et al(2017)借鉴机器学习文献提出了一组新的考虑参数不确定的贝叶斯模型来预测股票收益。Gu et al(2018)在资产定价领域广泛使用了随机森林、神经网络等机器学习方法,利用公司特征、行业特征和宏观经济变量对股票收益进行了较准确的估计。

总之,准确地进行收益预测需要使用恰当的方法,并根据不同的市场状态和时期不断更新预测指标,在参数估计时应选择合适的估计区间,避免数据的过度挖掘。准确地进行收益预测不仅可以加深对于资本市场各个部分的相互关系、信息传导途径的理解,还有助于提出新的更广泛的资产定价理论。

二、收益预测检验与市场有效性

资本市场的资产收益是否可预测?这一问题吸引了很多金融经济学者的关注,也曾长期存在争议。较早的文献可以追溯到 Colwes(1933),他调查了 45 家专业机构对于股票收益的预测能力,发现股票市场不能被有效预测。尽管如此,诸如道氏理论、技术分析、趋势分析等对股票价格的预测方法依然很受市场欢迎。但是,Fama & Blume(1966)、Jensen & Bennington(1970)等的研究显示,在考虑交易成本后,多数技术指标构建的投资组合不能比买入并持有大盘指数的投资组合表现更好。在此基础上,Fama(1970)提出了有效市场假说,他们将股票市场分为了强、半强以及弱有效市场三种状态,并认为有效市场中的当前股票价格已经反映了市场的全部信息。Malkeil(1973)提出了随机游走假说,认为股票价格波动是随机的,因此在短期内预测股票价格变得不可行。

20 世纪 80 年代以后,众多学者开始关注收益可预测性的内在经济机制研究,并发现了大量的资本市场收益可预测性的实证证据。Lo & MacKinlay(1988)分析了随机游走假说中价格波动的残差项分布特征,提出了一个简单的方法来检验该假说是否成立,并发现随机游走假说并不成立。Rozeff(1984)、Campbell & Shiller(1988)、Fama & French(1988, 1989)分析了宏观经济或公司财务信息与股票市场收益之间的关系,发现名义利率、通胀率、市盈率、账面市值比等多个指标能够显著预测未来股票收益。Campbell & Shiller(1988)将股息价格比的波动分解为期望股息波动和期望折现率波动,发现由市场总风险的变动产生的时变折现率会带来股票收益可预测性。Campbell(2000)认为,在目前的研究中已经有十分确凿的证据来证明股票收益能够在样本内可预测。所谓的样本内预测是指根据估计的模型在已知的样本内进行预测,并检验预测值与实际值的差异。

但是,Goyal & Welch(2008)批评现有的实证研究多基于样本内检验而忽视了样本外检验的重要性。样本外预测则指利用已估计的模型对于未知的未来数据进行预测。Stambaugh(1999)也指出,由于预测指标长期不变且与股票回报率滞后相关,因此样本内估计系数和 t 值可能有小样本偏

误。Goyal & Welch(2008)选取了文献中被证明能够在样本内显著预测股票收益的 12 个经济变量,使用单变量预测回归研究了指标在样本外对股票收益溢价的预测能力。他们发现样本内预测的回归参数并不稳定,在样本外,这些变量的预测精度并不比历史平均收益值更准确。

在中国金融市场收益预测的研究中,姜富伟等(2011)选取估值、货币发行、交易量等 9 个经济变量作为预测变量,研究了我国股票市场以及根据行业、规模、账面市值比和股权集中度等划分的成分组合的收益可预测性,发现我国股票市场组合和各种成分投资组合都具有显著的样本内和样本外可预测性。后续研究进一步指出,美国宏观经济状况、国际波动率风险、经济政策不确定性和公司盈利等变量都对我国股票市场有显著的预测能力(Goh et al, 2013; Chen et al, 2017; Jiang et al, 2018)。

总之,随着对收益是否可预测问题的深入探讨,学者们逐步发现并深刻地揭示了股票价格与宏观经济以及公司经营状况之间的内在联系,并已经基本认同股票收益的确存在可预测性。但是,可预测性并不意味着无成本的套利,也并不一定违背有效市场假说。Rapach & Zhou(2013)指出,只有当经过风险调整后的期望收益在考虑了交易成本以及其他交易摩擦后依然不为零时,才能说明与有效市场假说相违背。因此,认为收益可预测性天然地与有效市场假说相矛盾的说法是一种误解。在有效市场中,收益依然可以预测,只是在通过预测来获取额外收益时需要付出成本。另外,投资者对风险的厌恶程度也会随着经济周期的变化而变化。若想确定某一指标是否能真正预测未来收益,需要从样本内和样本外的多个角度分析。

三、收益预测的主要指标

基于经济理论模型,学者们发现了众多变量可以预测资产收益。本文重点关注资产数量和类别较多的股票市场,并将近年来被发现的收益预测指标分为四个类别,包括基本面指标、技术面指标、情绪指标以及波动率指标。其中,情绪指标和波动率指标是最近研究中常关注的。我们分别介绍各类收益预测指标,并探讨市场上不同类型信息与资产回报之间的关系。

(一) 基本面指标

基本面指标是指包含宏观经济信息或公司运营状况的预测指标。投资者可以通过了解宏观经济指标(诸如 GDP、失业率、原油价格等)对投资的可能性和经济的总体健康状况进行判断。投资者也可以通过关注公司的收入、支出、资产、负债等信息,对公司未来现金流与股票收益进行合理预期。由于股票市场收益的可预测性与经济周期变化紧密相关,因此了解宏观经济信息有助于正确把握经济周期的波动。

早期研究发现,基本面指标包括与金融市场相关的名义利率(Fama & Schwert, 1977; Ang et al, 2007)、利率差(Fama & French, 1989)、公司发行活动(Baker & Wurgler, 2000)等。还有与宏观经济密切相关的劳动收入(Santos & Veronesi, 2005)、总产出(Rangvid, 2006)和产出缺口(Cooper & Priestly, 2009)等。近期 Lettau & Ludvigson(2002)通过构造非常有影响力的总消费财富比率指标进行研究发现,总消费财富比率对于实际的股票回报和国债利率超额回报都具有很强的预测性,而且预测能力主要集中在短期和中期范围内。姜富伟等(2011)指出股息率、通货膨胀、换手率、货币供给增长率等宏观基本面指标对中国股票市场也有很强的预测能力。样本内检验结果显示,宏观基本面指标可以显著预测下一个月的市场收益以及绝大多数基于上市公司行业、规模、面值市值比和股权集中度等所构建的成分投资组合收益。Chen et al(2017)研究指出,我国宏观经济政策的不确定性与股票市场收益存在反向关系。他们通过衡量与经济政策不确定性相关的新闻数量构造了经济政策不确定性指数,并发现当指数较高时往往意味着未来较低的股票市场收益。进一步研究显示,经济政策的不确定性放大了投资者的行为偏差,进而加深了市场错误定价程度,最终通过上市公司实际较低的现金流现状导致了股票收益的下降。

不少研究也从其他市场行情出发来预测股票收益。Jacobsen et al(2016)研究发现,铜和铝等工业材料的价格变动能够在十个工业国家中预测股票回报。工业材料的价格变化不仅包含经济周期变

动信息,还可以通过企业购买原材料进行生产的渠道影响企业的折现率和资金流,继而扩散到股票市场。Kilian & Park(2009)研究发现,美国股票市场对于石油价格冲击的反应比较敏感,而股票实际回报很大程度上与油价的变化息息相关。需求冲击和供给冲击共同驱动的国际原油市场解释了美国实际股票价格中22%的长期波动。针对中国股票市场,Goh et al(2013)和Chen et al(2017)研究发现,中国股票市场预期收益对美国宏观经济状况和国际波动率风险等跨国、跨市场信息都有显著的反应。

此外,还有不少研究从公司运营状况等角度对收益进行预测。Sadka & Sadka(2009)研究了个体公司盈利情况对于市场总体盈利和股票回报的预测效果,并发现个体公司盈利能在总体层面上更好地预测盈利增长和股票收益。Li et al(2013)采用资产的机会成本与国债收益率之差构造了一个隐性风险溢价指标,研究发现,资本的机会成本是衡量投资者时变期望回报的重要代理变量。Rappach et al(2016)的研究指出,卖空交易和公司基本面恶化可以在月度、季度、半年度和年度层面显著预测未来股票市场收益。Jiang et al(2018)研究发现,包括上市公司的毛利率、资产收益率和股权收益率等多个盈利指标可以显著预测中国股票市场预期收益。同时,盈利能力与股票预期收益之间的正向关系会随着公司所面临的投资摩擦程度的降低而增强。

(二)技术面指标

技术面指标通常依据股票价格的变化趋势而不是其内在价值来预测未来收益。由于股价随机游走假说曾被广泛接受,因此在文献中技术指标往往处于边缘位置。同时,许多技术分析方法的批评缺少合理的经济传导机制,并存在对数据过度挖掘的问题。Han et al(2016)提出了解释技术分析有效性的一般均衡模型,指出技术分法是投资者了解股票价格信息的一种途径。当市场上使用该方法的投资者越多时,技术面指标影响股价的程度会越深。因此,合理地使用技术面指标能够预测股票收益。

Park & Irwin(2007)将技术面指标主要分为了六类,包括标准型(如移动平均指标)、基于模型自助法型、事实检查型、基因编程法型、非线性型和图表模式型。他们研究发现,从1960到1987年间技术分析策略在股票市场获利十分有限,但却在外汇市场和期货市场中表现良好。而从1988到2004年间,技术分析在不同的金融市场中获取的经济利润显著。他们进一步指出,未来在研究技术面指标的收益预测效果时,应将风险和交易成本考虑进来,并应更多地从指标的可获利性角度出发。Faber(2007)构造了一个基于简单移动平均法则的技术面指标,通过最近一段时间股价的平均值来预测未来的股价。作者采用了移动平均指标的时点策略:当股价的移动平均值高于当前价格时卖出,反之则买入。通过投资组合分析发现,使用该指标能够显著提高多种资产风险调整后的回报。

近年来,不少研究从更广泛的角度探讨技术面指标对收益预测的作用。Moskowitz et al(2012)构造了时间序列动量策略,并发现采用该策略构建的多样化投资组合能够获得显著高于由标准的资产定价因子所构造的投资组合收益。他们发现月度层面的价格趋势广泛存在于股票市场、债券市场、期货市场以及商品市场中。Tang & Whitelaw(2011)使用时变股票市场夏普比率作为一种技术面指标,并发现条件夏普比率所显示出的时变性恰好与经济周期的阶段相符合。Han et al(2013)与Han et al(2017)的研究都指出,相较于传统的买入并持有的投资策略,利用波动率构建的投资组合在价格均线的技术面信号下会产生大量额外的超额收益。Neely et al(2014)通过将低频的宏观经济指标与技术面指标结合研究发现,结合后的投资组合能够获得可观的超额收益。Zhou & Zhu(2009)和Barberis et al(2015)试图利用现有的理论模型来解释为何趋势跟踪策略能预测股票收益。Han et al(2016)在提出了技术分析对于股票价格影响机制的一般均衡模型后,构建了一个稳健的趋势因子,并发现这个趋势因子能很好地解释横截面股票收益。Jiang et al(2017)检验了28000多种技术分析指标对中国股票市场的预测能力,在消除数据挖掘和过度拟合问题的影响后,他们发现仍然有大量的技术指标可以显著预测中国股票市场趋势。

(三)情绪指标

在金融市场中,情绪往往指对上市公司未来现金流或折现率过于乐观或悲观的有偏预期。情绪

指标刻画了人们对于市场、商业环境或者公司估值等因素的主观感受,这样的主观感受能够从多种渠道在市场中传导。比如当消费者情绪指数显示目前消费者悲观情绪较重时,市场上的公司会减少存货,因为公司担忧消费者不会进行消费从而增加了他们囤积存货的风险。

在早期涉及情绪指标的研究中, Lee et al(1991)提出封闭基金折价率是衡量投资者情绪的重要代理变量。同时,他们发现相同程度的情绪更能影响那些市值规模较小的公司或者主要被个人投资者所持有和交易的股票的未来回报率。Baker & Wurgler(2000)研究指出,新股股权发行和债券发行比例可以显著预测市场回报。这是由于新股发行占比包含了市场情绪,公司会在市场情绪高涨期发行更多股票以获得更高的回报。Ben-Rephael et al(2012)提出使用股票型基金累计净交易的标准误差衡量价格噪音和投资者情绪。Baker & Wurgler(2006)使用主成分分析法构建了一个综合的投资者情绪指数,并进一步研究发现,当投资者情绪低迷时,投机性强、套利困难的公司价值往往容易被低估,因此这些估值水平高、规模小、盈利水平低、成立时间短、股票波动大的公司会获得更高的股票收益。当投资者情绪高涨时,这类公司的价值却容易被高估,因此伴随着较低的未来股票收益。

Huang et al(2015)构造了一个改进的投资者情绪指数,发现投资者情绪可以显著预测股票市场收益。他们使用偏最小二乘法和换手率、封闭基金折价率、IPO 数量、IPO 首日回报率、新股发行占比和分红指标作为情绪代理变量构建情绪指数。他们通过剔除投资者情绪代理变量中公共的噪声部分,更精确地提取了变量中与股票未来收益相关的信息。研究显示,新的投资者情绪指数的预测能力无论在样本内还是样本外预测中都显著强于以往的指数。此外,投资者情绪指数是通过现金流渠道影响未来的股票市场收益:由于高情绪的投资者对公司未来现金流的预期过于乐观并脱离了基本面,最终造成了股票市场泡沫和泡沫破灭后的崩盘,并造成较低的未来投资收益。

Jiang et al(2018)指出上市公司经理人在公司运营中也带有一定的情绪,这种情绪会传导至金融市场中。他们利用文本数据构造了一个经理人情绪指数,并发现该指数能够反向预测未来总体的股票市场收益,即高涨的经理人情绪意味着更低未来股票收益,低迷的经理人情绪意味着未来更高的股票收益。经理人情绪指数甚至优于不同种类的投资者情绪指数,在月度层面能够更好地预测未来收益。另外,较高的经理人情绪往往意味着未来较高的投资增长以及较低的总盈利预期,这表明经理人对公司的过度投资以及对未来现金流的有偏预期是经理人情绪指数能够反向预测股票收益的来源。

从国际股票市场出发, Baker et al(2012)也分别对全球主要的股票市场构建了投资者情绪指数,并将各个指数组合成了一个全球的情绪指数。他们发现无论是国别的还是全球的投资者情绪指数都能反向预测股票收益。尤其是对于那些难以套利和估值的股票,当投资者情绪是高涨时,其未来收益会降低。同时,他们指出私人现金流是导致国际市场间投资者情绪互相传播的一个媒介。

(四)波动率指标

在传统的理性预测模型框架下,持有资产的风险和所获取的收益是相互联系的,如果一个人承担了更高的风险,那么其期望收益往往更高。同样,更低的风险预示着更低的收益。在金融市场中,风险被定义为资产回报的标准误差,高收益波动往往意味着高风险,也意味着高期望收益。因此,研究者们很自然地将市场波动用于预测未来的市场收益。French et al(1987)研究了股票回报和股票市场波动率之间的关系,研究发现,期望市场风险溢价(股票投资组合的期望收益率减去国债收益率)与未来股票收益间具有显著的正向关系。

Copeland & Copeland(1999)发现芝加哥期权交易所中市场波动率指标(VIX)的百分比变化可以显著地在月度层面预测市场回报。Bali & Hovakimian(2009)研究发现,来自风险中性分布的隐含波动率和来自真实分布的已实现波动率能够预测横截面股票期望收益的变化。Bollerslev et al(2009)的研究显示,隐含或实现的波动率可以显著地从时间序列角度解释股票市场总体收益的变动情况,即高波动率预示着未来高的股票回报,而低波动率预示着未来低的股票回报。Atilgan et al(2015)研究了由波动率构造的多空对冲组合收益与股票市场总体期望收益间的跨期关系。Kelly &

Jiang(2014)提出了可以从横截面股票收益中估计出市场时变尾部风险的方法,并研究发现尾部风险每提高1%则预示着下一年市场的超额收益提高4.5%,这说明尾部风险对于股票市场总体收益具有显著的预测能力。Chen et al(2017)使用7个国际主要的金融市场数据构造了国际波动率风险指数,并研究发现,国际波动率风险与中国股市隔夜日度收益负相关,但与第二日的中国股市收益正相关。较高的国际波动率风险指数意味着较高的中国股票市场波动率、较低的交易量和市场流动性。此外,美国市场的波动率风险在预测中国股票市场收益时尤为有效。

四、收益预测的计量方法

从资本市场的长期变化动态来看,任何预测模型都存在不确定性,模型参数也存在不稳定性,这导致单一的传统预测方法无法长期对收益进行准确预测。因此,使用恰当的计量方法的确能显著提高回归的预测能力。这些方法通过更准确把握收益数据的生成过程,从而减少潜在的模型不确定性,降低参数的不稳定性,最终提高了指标的预测准确度。这些方法包括加入经济机制约束法、潜在因子法、组合预测法、状态转换法和机器学习法等。

(一)加入经济机制约束法

第一种提高收益预测准确性的方法是在预测回归中加入经济机制约束。考虑一个简单的二元预测回归模型:

$$R_{t+1} = \alpha_i + \beta_i x_{i,t} + \epsilon_{i,t+1} \quad (1)$$

式(1)中, R_{t+1} 是取对数的股票超额收益, $x_{i,t}$ 代表在 t 期第 i 个可能预测股票收益的预测指标($i=1, \dots, K$)。通过预测回归得到的估计系数以及预期超额收益如下式所示:

$$\hat{R}_{i,t+1} = \hat{\alpha}_{i,t} + \hat{\beta}_{i,t} x_{i,t} \quad (2)$$

式(2)中, $\hat{\alpha}_{i,t}$ 和 $\hat{\beta}_{i,t}$ 分别由式(1)中预测回归方程通过最小二乘法计算而得出, $\hat{R}_{i,t+1}$ 代表预测指标 $x_{i,t}$ 所预测的超额收益。Campbell & Thompson(2008)研究指出,由于预测回归的样本空间有限,使得股票收益依然包含大量不可预测的部分,因此得到的估计系数 $\hat{\beta}_{i,t}$ 并不准确。这也使得回归预测的估计收益 $\hat{R}_{i,t+1}$ 与真实收益 R_{t+1} 之间存在较大差异,即预测精度较低。为了解决这一问题,他们提出可以在估计系数 $\hat{\beta}_{i,t}$ 和估计收益 $\hat{R}_{i,t+1}$ 中加入经济机制约束:当估计系数 $\hat{\beta}_{i,t}$ 的符号与对应的经济机制模型中所显示的影响收益的方向相反时,将该估计系数设置为零;当预测的超额收益 $\hat{R}_{i,t+1}$ 小于零时,由于投资者不会为了风险而承担负的收益,故也将此时的预测收益设置为零。加入经济机制约束能降低预测模型参数的不稳定性,提高预测准确性。Campbell & Thompson(2008)的研究进一步显示,和无约束的单变量预测回归相比,对一些不符合经济机制的预测变量加入约束后,回归的预测结果显著优于采用历史平均值的预测结果。Rapach & Zhou(2013)对14个经济变量施加约束得到了与Campbell & Thompson(2008)类似的结论。他们进一步研究发现,加入经济机制约束主要通过降低指标在经济扩张期的预测误差,从而提高了整体的预测准确度。

此外,还可以通过局部求合法或者贝叶斯法在预测模型中加入经济约束。Ferreira & Santa-Clara(2011)将对数化后的股票收益分解为市盈率的对数增长率、企业盈余的对数增长率以及对数化后的股利价格比与当期股价变化之和,并进行预测回归,研究发现用局部求和法所估计的预测结果显著优于历史平均值的预测结果。贝叶斯法主要通过将经济机制约束作为先验认识加入预测模型中。Pastor & Stambaugh(2012)使用了贝叶斯法将理性投资者的最优资产配置等信息加入先验认识中,并在此基础上预测股票收益。他们研究发现,对于投资者而言,由于存在未来收益的不确定性以及风险估计偏误,股票市场的长期收益波动远高于短期波动。因此,即使是并不相信股票收益可预测的投资者,也会因为收益预测准确性的变化而影响他们的资产配置方案。Pettenuzzo et al(2014)将条件夏普比率和非零的股票超额收益假设加入经济机制约束中,并以此调整模型参数的尾部分布特征。他们发现新方法适用于包含大量收益预测模型和指标的联合预测,并可以显著降低参

数的不确定性以及因选择表现不佳的预测模型所带来的风险。

(二)潜在因子法

另一种提高收益预测准确性的方法是潜在因子法,它通过把握大量股票收益预测指标中协同变化(co-movements)的部分来提高预测准确性。Ludvigson & Ng(2007), Kelly & Pruitt(2013)和 Neely et al(2014)都采用了潜在因子法,他们假设预测指标可以被一个潜在因子所表示,具体形式如式(3)所示:

$$x_{i,t} = \lambda'_{i,t} f_t + e_{i,t}, i = 1, \dots, K \quad (3)$$

上式中, $x_{i,t}$ 是预测指标 i 在第 t 期的值, f_t 是可以表示预测指标 $x_{i,t}$ 的一个 q 维的潜在因子, $\lambda_{i,t}$ 是因子系数, $e_{i,t}$ 是零均值干扰项。在式(3)中,预测指标间的协同变化主要由相对数量较小的因子 ($q \ll K$) 所控制。Stock & Watson(2004)使用主成分分析法估计潜在因子系数,然后根据该系数所表示的预测指标来估计股票收益。按照这一思路, Ludvigson & Ng(2007)采用股票回报的季度数据,从 209 个宏观和 172 个金融变量里提取出收益预测的潜在因子。他们发现该指数具有显著的样本内预测能力,并且优于历史平均法的预测结果。Neely et al(2014)选取了文献中常见的 14 个经济变量和 14 个基于移动平均、动量和交易量构造的技术面指标,构造了一个潜在因子来预测美国月度股票超额收益。他们发现潜在因子法的预测结果显著优于历史平均法或者基于 28 个经济变量和技术面指标的单变量预测回归的结果。

从预测的角度来看,潜在因子法的一个缺点是估计出的潜在因子被解释成各个预测指标间的协同变化,而忽略了预测指标与收益间的直接关系。Bai & Ng(2008)对潜在因子法进行了改进,并基于不同潜在因子对于收益预测的准确程度调整了因子系数的大小,使得收益预测更加准确。Kelly & Pruitt(2015)在此基础上发展并提出了一个三阶段回归过滤法来估计与所预测收益最接近的潜在因子系数,并指出三阶段回归过滤法优于主成分分析法,可以不断通过实际预测结果修正对于潜在因子的估计。Kelly & Pruitt(2013)使用该方法从一系列股票估值指标中提取潜在因子,并得到了很好的样本外收益预测结果。与此类似, Huang et al(2015)使用二阶段回归的偏最小二乘法从一系列情绪指标中提取投资者情绪指数,并得到了很好的样本外收益预测结果。Light et al(2017)在此基础上提出了基于横截面股票收益预测的偏最小二乘法,他们从 26 个公司特征变量中提取了一个或多个潜在因子,并发现该方法能够较好地预测不同股票间的收益差别。

(三)组合预测法

与只关注收益预测指标间协同变化的潜在因子法不同,组合预测法直接将不同变量、模型的估计进行有效整合。Timmermann(2006)指出,正如分散化的资产配置能够提高投资组合稳健收益一样,组合预测法也可以通过合理分配不同模型和变量的预测估计占比,从而提高整体收益预测的准确性。从直观上来讲,组合预测法结合了各个模型所包含的不同层面信息,诸如宏观经济形势、公司运营状况、投资者情绪等。另外,由于单变量模型的预测能力会随着时间改变,所以长时间依赖一个模型使投资者只把握了其在特定时期对预测有帮助的信息,但是在其他时期这样的信息却可能只是误导性的预测噪音。在组合预测法中,如果各个单变量模型的相关性较弱,其组合后的预测结果波动则会较小。这样即使不同的预测模型各自包含较高的模型不确定性和参数不稳定性,组合后的结果也要比单变量预测模型更准确,预测风险更低。

组合预测法通常是将由不同模型预测的收益进行加权平均得到组合预测收益:

$$\hat{R}_{t+1} = \sum_{i=1}^K w_{i,t} \hat{R}_{i,t+1} \quad (4)$$

式(4)中, $w_{i,t}$ 代表在 t 期中对第 i 个估计收益赋予的预测权重,所有预测权重之和为 1。最简单的赋权方法是对每个估计收益赋予相同的权重 $1/K$,类似投资组合分析中的等值加权方法,其收益预测表现往往也不错。Rapach et al(2010)指出,虽然等权重赋值法简单易行,但是有时权重向某些预测变量“倾斜”是有帮助的。他们根据不同预测指标的样本外预测均方误差计算了一组带有折价

因子的组合预测系数,使得那些在近期能够更准确预测收益的部分在预测中所占权重更大。他们的研究显示,调整系数后的组合预测结果与等值加权系数的预测结果一样好,并显著优于使用历史平均法进行预测。

另一种组合预测的方法是将所有预测变量同时加入类似式(1)的预测方程中进行多元回归(kitchen sink)。Goyal & Welch (2008)和 Rapach & Zhou(2013)将多个指标同时加入多元回归预测方程中进行研究发现,这一多元回归模型的样本外预测效果非常差。导致该问题的原因是在单一预测方程中加入过多参数从而使得估计量的方差较大,并引起了数据的过度拟合。解决该问题的方法可以通过引入传统的模型选择标准来减少方程中的估计变量,但效果依然有限。Cremers (2002)提出可以使用贝叶斯法进行改进。他们在每个预测指标系数前加入先验概率分布,通过预测回归不断调整模型中指标对收益预测的后验概率,并使得那些低后验概率的指标系数缩减至零。从本质上来讲,贝叶斯方法也是通过将那些导致较大预测误差的指标权重降低从而提高总体的预测精度。Zhang et al(2017)从贝叶斯方法出发,提出了基于自适应推进算法和平均窗口法的加权平均最小二乘法(weighted-average least squares)进行组合预测,研究发现新方法在降低了模型参数不确定性的同时也增加了预测精度与投资者效用,并优于传统的组合预测法。

(四)状态转换法

由于收益数据的生成过程受参数的不稳定性影响,如不同的经济周期预测模型的参数往往发生变化,因此根据不同的市场状态进行收益预测的状态转换方法受到了许多研究者的关注(Bai & Perron, 2003; Rapach & Wohar, 2006; Elliot & Muller, 2006)。状态转换法一般通过马尔可夫状态转移预测回归模型对未来收益进行估计。考虑如下预测回归:

$$R_{t+1} = \alpha_{S_{t+1}} + \beta_{S_{t+1}} x_t + \sigma_{S_{t+1}} u_{t+1} \quad (5)$$

其中, S_{t+1} 代表 $t+1$ 期的市场状态,且服从一阶马尔科夫状态转换过程。式(5)中其他的估计系数都是根据不同市场状态计算而得。由于市场状态往往不能直接被观察,因此需要根据马尔科夫状态转换方法计算不同时期所属的最大可能的市场状态。Guidolin & Timmermann(2007)通过极大似然法构造了一个包含美国股票和债券收益的多变量马尔科夫状态转换模型,他们将市场标记为“慢牛”、“牛”、“衰退”以及“复苏”四种状态,并根据不同的市场状态进行收益预测。研究结果显示,使用状态转换法可以显著提升投资者的效用和配置后资产的总收益。

Henkel et al(2011)构建了一个包括了市盈率、短期名义利率、期限利差和违约利差等不同预测指标的、具有两种市场状态(扩张和衰退期)的马尔可夫状态转换模型,并且通过贝叶斯方法估计出的两种市场状态与美国国家经济研究局对于经济扩张期和衰退期的划分十分接近。同时研究结果也显示,使用该方法能够很好地在经济衰退期预测股票收益。在进行非线性的股票收益预测时,研究者常使用状态转换法。这些非线性预测模型包括门限模型、神经网络模型等。Guidolin et al (2009)使用状态转换法分析了不同的市场状态中,哪一种关于股票收益预测的非线性模型可能更为有效。此外,另一种时变参数模型允许预测参数在各个时期不断变化,即每一期都代表一个新的市场状态。Pettenuzzo & Timmermann(2011)使用贝叶斯法估计时变参数预测模型的参数,并观察其随时间变化的特征。他们发现参数变动并不是由一个连续的随机变量变动而引起,而是由少量的突变造成,且模型参数变动的机制与某一时期市场的最优投资组合的突变有关。在收益预测的过程中,如果忽视了这种最优投资组合的改变,将使得长期的市场投资者蒙受巨大的效用以及利润损失。

(五)机器学习法

随着计算机科技的飞速发展,机器学习法在许多领域得到了广泛应用。从语音识别、人脸识别到智能翻译、智能管家,从人机围棋大战到人工智能基金的成立,机器学习在数据分析和优化方面甚至超越了人类。在收益预测中,机器学习法往往能从复杂的数据结构中识别出没有被人们发现或被广泛接受的规律,并能很好地避免数据的过度拟合,拥有较强的样本外预测能力。

简单来讲,机器学习法主要是通过对于已知样本的学习来对未来进行预测。比如,给定一段时

间内的股票收益与其他相关的金融数据作为训练集,计算机通过学习训练集内的收益与滞后一期其他变量的关系从而形成某种计算规则(函数)。当给计算机提供现期的金融市场数据后,它能根据学习得到的计算规则(函数)来对下一期的股票收益进行预测。Mullainathan & Spiess(2017)从计量方法的角度梳理了近年来主要的机器学习文献。他们指出传统的预测方法着重从参数估计的角度研究解释变量和被解释变量的关系,而机器学习则直接用被解释变量来构造解释变量。即使用机器学习法也可以得到类似连接解释变量与被解释变量的估计系数,但该系数也很少是固定不变的。因此,对于收益预测者而言,在使用机器学习法时尽可能全面地找到与收益相关的数据尤为重要。Gu et al(2018)的研究中就使用了 920 个与公司特征、行业分类、宏观经济状况相关的变量,并且通过将样本分为训练集、调试集和预测集不断地进行滚动预测,最终获得了较好的预测结果。

在金融和会计领域,机器学习包含监督学习法、神经网络法、决策树法、朴素贝叶斯法、支持向量机等。其中,监督学习法(supervised learning)主要是找到训练集中能够有效预测收益的变量,并进行有效的线性估计,最终得到相关的预测函数。在这一过程中,需要通过计算机剔除大量的无关变量,常用的剔除方法是最小绝对值收敛和选择算法(least absolute shrinkage and selection operator, LASSO)。该方法在不同变量与收益的估计参数前加入了惩罚因子,当无关的收益预测变量使得总体的预测方差与系数绝对值之和较大时,其估计系数会被赋零(或趋于零)。Green et al(2016)使用了 LASSO 法从上百个金融变量中构造了不同维度的能够预测收益的信号,研究发现,诸如规模、账面市值比、惯性因子等传统的金融变量在预测收益时往往忽略了重要信息;而盈利、换手率、行业调整的惯性因子等能够更好地预测横截面股票收益。在此基础上,Freyberger et al(2018)使用自适应的分组 LASSO(adaptive group LASSO)法从 62 个公司特征中最终挑选出能够显著预测未来股票收益的 11 个指标。

总体来看,机器学习法在收益预测方面的研究目前还处于起步阶段,但近期已逐渐涌现出一系列与之相关的研究。Moritz & Zimmermann(2016)使用机器学习法改进了目前资产定价领域内的投资组合分析法,提出了树形结构的条件投资组合排序法。他们发现在美国股票市场中公司的短期超额收益能够显著预测未来股票价格(去除了交易成本的影响),并且优于目前文献中主流的回归分析法。Hoberg & Phillips(2016)使用机器学习法逐年比较了不同的上市公司财务报告中的文本相似性以及产品特征。他们指出通过机器学习所获得的公司行业分类标准与传统上通过了解公司注册信息的行业分类不同,前者能更好地与实际的产业冲击、商业竞争相对应。这也为更准确地预测行业投资组合收益提供了新的方法。Rapach et al(2018)与 Han et al(2018)分别利用机器学习算法从行业历史收益以及公司特征中提取信息,并有效预测股票未来收益。当然,正如 Gu et al(2018)的研究所指出,由于金融市场往往充斥着各类噪音,所以并不是越复杂的机器学习算法越有效,研究者需要根据实际情况不断地选择最优的预测方法。

此外,对于未来的研究而言,除了找到能够显著预测股票收益的算法,研究者们更需要理解背后的市场传导机制,厘清机器学习法在不同市场时期赋予金融市场相关变量不同预测权重的原因。通过人类向机器再“学习”的过程,更深入地理解金融市场收益预测的真正来源。

五、收益预测准确性的度量

在实际投资中,往往有不少人声称自己能够准确预测未来收益。例如,有预测者被证明能预测出所有的熊市,但事实上是在绝大多数时候其对市场行情都非常悲观。因此,如何度量收益预测的准确性问题显得尤为重要。除了样本内拟合值外,目前在文献中常出现的衡量指标有均方预测误差,样本外拟合值,调整后的均方预测误差,基于效用增量和可获利性的指标。下面逐一进行介绍。

均方预测误差(mean squared forecasting error, MSFE)是较早估计收益预测准确性的指标。假设样本期为 T ,样本内的初始观测期长度为 t_1 ,样本外预测期长度为 $t_2 = T - t_1$,即最开始使用 t_1 期的观测值对 t_1 期后的收益 R_{t_1+t} 进行预测($t = 1, 2, \dots, t_2$)。对于每一个预测指标 x_i ,其在 t_2 期内均

方预测误差如(6)式所示。

$$MSFE_i = \frac{1}{t_2} \cdot \sum_{t=1}^{t_2} (R_{t_1+t} - \hat{R}_{i,t_1+t})^2 \quad (6)$$

其中, \hat{R}_{i,t_1+t} 表示预测指标 x_i 在 t_1+t 期的预测值, R_{t_1+t} 表示 t_1+t 期的真实收益值。 $MSFE_i$ 测度了预测指标 x_i 在所有的样本外预测期 t_2 内的预测误差平方和的平均值。当 $MSFE$ 的值越小时,说明所预测的收益准确度越高。Meese & Rogoff(1983)使用 $MSFE$ 衡量了多种国际汇率的预测准确性,研究发现传统的汇率结构模型在预测上并不准确,其一年内的预测精度甚至无法超过随机游走模型。从预测可获利性的角度出发,Leitch & Tanner(1991)研究了专业预测公司和时间序列模型对于短期国债利率的预测精度。研究发现,尽管在 $MSFE$ 度量下,专业公司的预测精度不比时间序列模型的高,但是其在实际操作中却可以获得更高的收益,这说明 $MSFE$ 存在被改进的空间。

样本外拟合值(R_{OS}^2)是对传统的 $MSFE$ 统计量的一个改进,Campbell & Thompson(2008)使用了以历史平均值作为基准的预测模型。模型假定不变的期望回报,即当期回报的预测值是过去所有实际回报的平均值,并以此计算了 $MSFE_0$,如式(7)(8)所示。

$$\bar{R}_{t+1} = \frac{1}{t} \cdot \sum_{t=1}^t R_t \quad (7)$$

$$MSFE_0 = \frac{1}{t_2} \cdot \sum_{t=1}^{t_2} (R_{t_1+t} - \bar{R}_{t_1+t})^2 \quad (8)$$

式(7)中,第 $t+1$ 期收益预测是 \bar{R}_{t+1} ,它由过去 t 期的历史收益平均值计算而出。式(8)中, $MSFE_0$ 衡量了 t_1 期后的真实收益与对应的历史平均收益的均方预测误差。 $MSFE_0$ 值越小,代表历史平均收益的预测准确度越高。Campbell & Thompson(2008)计算了样本外(out of sample)拟合值 R_{OS}^2 ,如式(9)所示。

$$R_{OS}^2 = 1 - MSFE_i / MSFE_0 \quad (9)$$

当 $R_{OS}^2 > 0$ 时,说明由 $MSFE_i$ 测度的指标 x_i 对收益的预测准确性要高于历史平均收益预测的准确性。当 $R_{OS}^2 < 0$ 时,说明 x_i 对收益的预测精度低于历史平均收益预测。在文献中,当样本外拟合值 R_{OS}^2 能够显著超过 2%,则说明预测指标的预测准确度较高。

调整后的均方预测误差(adjusted $MSFE$)是 Clark & West(2007)根据 Diebold & Mariano(1995)和 West(1996)的研究提出的。Diebold & Mariano(1995)和 West(1996)提出的 DMW 统计量能够很好地比较非嵌套模型间的预测能力是否相同。非嵌套模型(non-nested model)是指两个互补包容、模型参数间无法互相表示的模型。而一般的预测性回归模型和历史平均收益预测模型属于嵌套模型(nested model)。Clark & West(2007)修正了 DMW 统计量,使得新的调整后均方预测误差能够更准确地比较嵌套模型间的预测能力,同时服从渐进正态分布,并具有良好的小样本性质(即在有限样本预测中能够更准确进行比较)。此后许多研究都采用了该统计量衡量不同指标的样本外预测能力(Rapach et al, 2010; Dangel & Halling, 2012; Neely et al, 2014)。

当研究者对一组有限数据反复地使用不同模型进行样本内收益预测以期找出最佳估计时,往往会造成数据过度挖掘的问题。此时,对于这一组特定数据而言,尽管某个预测模型被证明是合适的,也很可能只是巧合。如果使用该模型对未来的收益再次进行预测,其结果常不尽如人意。因此,研究者在进行参数估计时应避免数据的过度挖掘,特别是对于包含大量收益预测指标的预测模型。White(2000)提出了基于“事实检查”的自助法(bootstrap)以解决该问题。他构造了一个衡量不同预测模型的平均损失差的统计量,并检验一组相互竞争的预测方法所得预测值是否显著地优于某一个基准的预测值。这个统计量可以通过非参数平稳自助法依据原始时间序列的损失差所计算得出。

Inoue & Kilian(2005)推导出 Diebold & Mariano(1995)和 West(1996)所提出的 DMW 统计量的渐进分布,并适用于衡量收益预测嵌套模型中的数据过度挖掘问题。Rapach & Wohar(2006)使用自助法检验了 1927—1999 年间的 9 个常用的经济变量是否能显著预测 1964—1999 年间的美国股票市场大盘指数(标准普尔指数)。研究结果显示,经由自助法所得到的统计量显著,即这 9 个经济变量所进行的样本外预测结果是稳健的,并不存在数据过度挖掘的问题。Clark & McCracken(2012)针对多元比较嵌套模型提出了自变量固定自助法,并发现该方法适用于包含条件方差的预测模型,且在有限样本的模拟预测中表现良好。总之,自助法可以从一定程度上减少收益预测中数据过度挖掘的问题,研究者需要通过这些方法更科学地选择不同的预测指标和模型。

除了使用预测误差来衡量收益预测指标的有效程度外,不少学者也从预测指标带给投资者的效用增量和可获利性大小出发。考虑一个简单的均值—方差投资者效用模型:

$$u_i = \mu_i - \frac{1}{2}\gamma\sigma_i^2 \quad (10)$$

式(10)中,投资者从第 i 个股票资产中获得的效用为 u_i , u_i 由该资产所带来的平均收益 μ_i 、投资者对风险的厌恶程度 γ 以及该股票收益的方差 σ_i^2 所决定。当该股票的平均收益越高,收益的方差(波动率)越小,投资者对风险的厌恶程度越低(一般假设不变)时,其所获得的效用越高。最优化资产配置后,投资者在第 t 期对股票 i 的投资规模如式(11)所示。

$$a_{i,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{R}_{i,t+1}}{\hat{\sigma}_{t+1}^2}\right) \quad (11)$$

式(11)中, $\hat{R}_{i,t+1}$ 代表投资者对于 $t+1$ 期的股票预测收益, $\hat{\sigma}_{t+1}^2$ 代表对于 $t+1$ 期的收益预测波动。式(11)说明,当投资者对于股票 i 在下一期的预期收益越高且预期波动越小时,其投资占比越高。正如 Leitch & Tanner(1991)所指出,使用基于经典的均值—方差投资者效用模型,可以更直接地从理性投资者的角度衡量预测指标所带给他们的效用和利润增量。在文献中,往往将股票回报预测指标作为投资者从期望效用最大化问题得出的资产配置决策的输入变量。投资者通过这一预测指标所实现的平均效用与通过历史平均方法所实现的平均效用之差,代表了使用这一预测指标得到股票溢价或效用的提高。这一效用提高也被称为确定性等价回报,即投资者为取得该指标预测信息所愿花费的投资组合管理费用。

Marquering & Verbeek(2004)、Cooper & Priestly(2008)、Ferreira & Santa-Clara(2011)等都从均值—方差投资者的角度衡量了预测指标对于投资者效用的提高程度,并发现基于经济变量的预测明显优于历史平均法预测。当采用效用度量时,需要注意所计算的均值方差投资者往往假设具有常相对风险厌恶系数。然而,股票收益可预测性的程度是由经济周期变化的时变风险厌恶产生的,因此在常相对风险厌恶系数假设下计算得到的效用收益仅仅得到了“一小部分”非代表性的不能影响市场价格的投资者。实际上,一个非代表性投资者还可以利用代表性投资者的时变风险系数创造可预测回报,从而获得更高的收益。

不少研究者也从投资者财富增长的角度衡量了预测指标准确性。Hong et al(2007)研究了行业收益增长对于股票收益的预测能力,并发现对于那些具有指数效用函数的投资者,考虑滞后的行业回报信息能获得更高的投资收益与效用增长。Pettenuzzo & Timmermann(2011)和 Cenesizoglu & Timmermann(2012)使用财富指数衡量了已经实现最优配置的投资者的效用提高程度,他们发现基于经济变量的预测指标可以使投资者获得显著的投资收益。Cenesizoglu & Timmermann(2012)还发现 MSFE 预测和投资者效用提高程度之间存在弱相关关系。

总之,随着对收益预测研究的不断深入,目前的研究者在比较某个预测方法或指标的收益预测准确性时,更倾向于从不同角度进行对比。因此,无论是从预测方程误差出发衡量的均方预测误差、样本外拟合值,还是从均值—方差投资者出发衡量的效用增量,都是科学衡量收益预测准确性所必要的指标。

六、结束语

本文梳理了资本市场(特别是股票市场)收益可预测性的研究进展,重点介绍了该领域的最新方法、多种指标以及研究热点。对于收益预测的可行性,研究者们已经基本认同资产收益的可预测性,且不一定违背有效市场假说。但科学地预测收益需要从样本内和样本外等多方面,使用不同的指标来进行。

从收益预测的主要指标来看,目前文献中的指标可以分为基本面指标、技术面指标、情绪指标以及波动率指标这四类。基本面指标和技术面指标是比较传统的预测指标,它们分别从宏观经济形势、公司运营状况以及股票价格变化趋势等角度对收益进行预测。情绪指标和波动率指标是近期文献中的研究热点,它们分别从行为和风险的角度诠释了收益预测的来源。研究者既可以从单个指标出发,探讨驱动该指标预测收益的经济机制;也可以从多种指标联合预测出发,研究如何有效整合不同种类的收益预测信息。

从提高收益预测准确性的主要方法来看,目前研究主要包括加入经济机制约束法、潜在因子法、组合预测法、状态转换法、机器学习法等。其中,机器学习法是近年来的研究热点。尽管该方法在业界已经得到了广泛应用,但在金融学研究领域,相关的文献还处于起步阶段。无论是用哪种方法提高收益预测的准确性,研究者更应关注方法背后的经济逻辑与机制。此外,由于在长期中,任何预测模型都存在不确定性,模型参数也存在不稳定性,因此研究者需要根据不同的市场时期选用合适的模型与方法才能有效减少这两个问题所带来的预测误差。

常见的收益预测准确性衡量指标则包括均方预测误差、样本外拟合值、调整后的均方预测误差,基于效用增量和可获利性的指标。其中,均方预测误差,样本外拟合值,调整后的均方预测误差是从预测回归方程的误差大小等角度出发,而基于效用增量和可获利性的指标则是从投资者获利大小的角度出发,未来的研究应至少包含这两个角度。

总之,随着不同种类的数据越来越丰富,特别是随着大数据分析方法越来越全面,学者们对于资本市场收益可预测的研究也在不断深入。选取合适的预测方法与指标,避免数据的过度挖掘,从多角度衡量预测的准确性等方面成为收益预测的关键。另外,通过明确收益预测的来源,更全面地理解金融市场的定价和信息传导机制,将有助于推动金融学理论研究的发展。

参考文献:

- 姜富伟 等,2011:《中国股票市场可预测性的实证研究》,《金融研究》第9期。
- Ang, A., G. Bekaert & M. Wei(2007), "Do macro variables, asset markets, or surveys forecast inflation better?", *Journal of Monetary Economics* 54(4):1163—1212.
- Atilgan, Y. et al(2015), "Implied volatility spreads and expected market returns", *Journal of Business & Economic Statistics* 33(1): 87—101.
- Bai, J. & S. Ng(2008), "Forecasting economic time series using targeted predictors", *Journal of Econometrics* 146(2): 304—317.
- Bai, J. & P. Perron(2003), "Computation and analysis of multiple structural change models", *Journal of Applied Econometrics* 18(1):1—22.
- Baker, M. & J. Wurgler(2000), "The equity share in new issues and aggregate stock returns", *Journal of Finance* 55(5):2219—2257.
- Baker, M. & J. Wurgler(2006), "Investor sentiment and the cross-section of stock returns", *Journal of Finance* 61(4):1645—1680.
- Baker, M., J. Wurgler & Y. Yuan(2012), "Global, local, and contagious investor sentiment", *Journal of Financial Economics* 104(2):272—287.
- Bali, T. G. & A. Hovakimian(2009), "Volatility spreads and expected stock returns", *Management Science* 55(11): 1797—1812.
- Barberis, N. et al(2015), "X-CAPM: An extrapolative capital asset pricing model", *Journal of Financial Economics* — 144 —

- 115(1):1—24.
- Ben-Rephael, A. et al(2012), “Measuring investor sentiment with mutual fund flows”, *Journal of Financial Economics* 104(2):363—382.
- Bollerslev, T., G. Tauchen & H. Zhou(2009), “Expected stock returns and variance risk premia”, *Review of Financial Studies* 22(11): 4463—4492.
- Campbell, J. Y. & R. J. Shiller(1988), “The dividend-price ratio and expectations of future dividends and discount factors”, *Review of Financial Studies* 1(3):195—228.
- Campbell, J. Y. (2000), “Asset pricing at the millennium”, *Journal of Finance* 55(4):1515—1567.
- Campbell, J. Y. & S. B. Thompson(2008), “Predicting excess stock returns out of sample: Can anything beat the historical average?”, *Review of Financial Studies* 21(4):1509—1531.
- Cenesizoglu, T. & A. Timmermann(2012), “Do return prediction models add economic value?”, *Journal of Banking & Finance* 36(11):2974—2987.
- Chen, J., F. Jiang & G. Tong(2017), “Economic policy uncertainty in China and stock market expected returns”, *Accounting & Finance* 57(5):1265—1286.
- Chen, J. et al(2017), “International volatility risk and Chinese stock return predictability”, *Journal of International Money and Finance* 70(2):183—203.
- Clark, T. E. & M. W. McCracken(2012), “Reality checks and comparisons of nested predictive models”, *Journal of Business & Economic Statistics* 30(1):53—66.
- Clark, T. E. & K. D. West(2007), “Approximately normal tests for equal predictive accuracy in nested models”, *Journal of Econometrics* 138(1):291—311.
- Clements, M. P. & D. F. Hendry(2006), “Forecasting with breaks”, in: G. Elliott, C. W. J. Granger & A. Timmermann(eds), *Handbook of Economic Forecasting*, Elsevier.
- Cochrane, J. H. (2008), “The dog that did not bark: A defense of return predictability”, *Review of Financial Studies* 21(4):1533—1575.
- Cooper, I. & R. Priestley(2008), “Time-varying risk premiums and the output gap”, *Review of Financial Studies* 22(7):2801—2833.
- Copeland, M. M. & T. E. Copeland(1999), “Market timing: Style and size rotation using the VIX”, *Financial Analysts Journal* 55(2):73—81.
- Cowles, A. (1933), “Can stock market forecasters forecast?”, *Econometrica* 1(3):309—324.
- Cremers, K. M. (2002), “Stock return predictability: A Bayesian model selection perspective”, *Review of Financial Studies* 15(4):1223—1249.
- Dangl, T. & M. Halling(2012), “Predictive regressions with time-varying coefficients”, *Journal of Financial Economics* 106(1):157—181.
- Diebold, F. X. & R. S. Mariano(1995), “Comparing predictive accuracy”, *Journal of Business & Economic Statistics* 13(3):253—263.
- Elliott, G. & U. K. Müller(2006), “Efficient tests for general persistent time variation in regression coefficients”, *Review of Economic Studies* 73(4):907—940.
- Faber, M. T. (2007), “A quantitative approach to tactical asset allocation”, *Journal of Wealth Management* 9(4):69—79.
- Fama, E. F. (1970), “Efficient capital markets: A review of theory and empirical work”, *Journal of Finance* 25(2):383—417.
- Fama, E. F. & M. E. Blume(1966), “Filter rules and stock-market trading”, *Journal of Business* 39(1):226—241.
- Fama, E. F. & G. W. Schwert(1977), “Asset returns and inflation”, *Journal of Financial Economics* 5(2):115—146.
- Fama, E. F. & K. R. French(1988), “Dividend yields and expected stock returns”, *Journal of Financial Economics* 22(1):3—25.
- Fama, E. F. & K. R. French(1989), “Business conditions and expected returns on stocks and bonds”, *Journal of Financial Economics* 25(1):23—49.
- Ferreira, M. A. & P. Santa-Clara(2011), “Forecasting stock market returns: The sum of the parts is more than the whole”, *Journal of Financial Economics* 100(3):514—537.

- French, K. R. et al(1987), "Expected stock returns and volatility", *Journal of Financial Economics* 19(1):3—29.
- Freyberger, J. et al(2017), "Dissecting characteristics nonparametrically", NBER Working Paper, No. w23227.
- Goh, J. et al(2013), "Can US economic variables predict the Chinese stock market?", *Pacific-Basin Finance Journal* 22(4):69—87.
- Goyal, A. & I. Welch(2008), "A comprehensive look at the empirical performance of equity premium prediction", *Review of Financial Studies* 21(4):1455—1508.
- Green, J., J. Hand & X. Zhang(2016), "The characteristics that provide independent information about average U.S. monthly stock returns", SSRN Working Paper, No. 2262374.
- Gu, S., B. T. Kelly & D. Xiu(2018), "Empirical asset pricing via machine learning", SSRN Working Paper, No. 159577.
- Guidolin, M. et al(2009), "Non-linear predictability in stock and bond returns: When and where is it exploitable?", *International Journal of Forecasting* 25(2):373—399.
- Guidolin, M. & A. Timmermann(2007), "Asset allocation under multivariate regime switching", *Journal of Economic Dynamics and Control* 31(11):3503—3544.
- Han, Y., A. He, D. E. Rapach & G. Zhou(2018), "How many firm characteristics drive US stock returns?", SSRN Working Paper, No. 3185335.
- Han, Y., D. Huang & G. Zhou(2017), "Anomalies enhanced: The use of higher frequency information", SSRN Working Paper, No. 2624650.
- Han, Y., K. Yang & G. Zhou(2013), "A new anomaly: The cross-sectional profitability of technical analysis", *Journal of Financial and Quantitative Analysis* 48(5):1433—1461.
- Han, Y., G. Zhou & Y. Zhou(2016), "A trend factor: Any economic gains from using information over investment horizons?", *Journal of Financial Economics* 122:352—375.
- Henkel, S. J. et al(2011), "Time-varying short-horizon predictability", *Journal of Financial Economics* 99(3):560—580.
- Hoberg, G. & G. Phillips(2016), "Text-based network industries and endogenous product differentiation", *Journal of Political Economy* 124(5):1423—1465.
- Hong, H. et al(2007), "Do industries lead stock markets?", *Journal of Financial Economics* 83(2): 367—396.
- Huang, D. et al(2015), "Investor sentiment aligned: A powerful predictor of stock returns", *Review of Financial Studies* 28(3):791—837.
- Inoue, A. & L. Kilian(2005), "In-sample or out-of-sample tests of predictability: Which one should we use?", *Econometric Reviews* 23(4):371—402.
- Jacobsen, B. et al(2016), "Stock market predictability and industrial metal returns", SSRN Working Paper, No. 1660864.
- Jensen, M. C. & G. A. Benington(1970), "Random walks and technical theories: Some additional evidence", *Journal of Finance* 25(2):469—482.
- Jiang, F. et al(2018), "Manager sentiment and stock returns", *Journal of Finance Economics*, forthcoming.
- Jiang, F., X. Qi & G. Tang(2018), "Q-theory, mispricing, and profitability premium: Evidence from China", *Journal of Banking and Finance* 87(2):135—149.
- Jiang, F., G. Tong & G. Song(2017), "Technical analysis profitability without data snooping bias: Evidence from Chinese stock market", *International Review of Finance*, forthcoming.
- Kelly, B. & H. Jiang(2014), "Tail risk and asset prices", *Review of Financial Studies* 27(10):2841—2871.
- Kelly, B. & S. Pruitt(2013), "Market expectations in the cross-section of present values", *Journal of Finance* 68(5):1721—1756.
- Kelly, B. & S. Pruitt(2015), "The three-pass regression filter: A new approach to forecasting using many predictors", *Journal of Econometrics* 186(2):294—316.
- Kilian, L. & C. Park(2009), "The impact of oil price shocks on the US stock market", *International Economic Review* 50(4):1267—1287.
- Lee, C., A. Shleifer & R. H. Thaler(1991), "Investor sentiment and the closed-end fund puzzle," *Journal of Finance* 46(1):75—109.

- Leitch, G. & J. E. Tanner(1991), "Economic forecast evaluation: Profits versus the conventional error measures", *American Economic Review* 81(3):580—590.
- Lettau, M. & S. Ludvigson(2001), "Consumption, aggregate wealth, and expected stock returns", *Journal of Finance* 56(3):815—849.
- Li, Y. et al(2013), "Predicting market returns using aggregate implied cost of capital", *Journal of Financial Economics* 110(2):419—436.
- Light, N. et al(2017), "Aggregation of information about the cross section of stock returns: A latent variable approach", *Review of Financial Studies* 30(4):1339—1381.
- Lo, A. W. & A. C. MacKinlay(1988), "Stock market prices do not follow random walks: Evidence from a simple specification test", *Review of Financial Studies* 1(1):41—66.
- Ludvigson, S. C. & S. Ng(2007), "The empirical risk-return relation: A factor analysis approach", *Journal of Financial Economics* 83(1):171—222.
- Malkiel, B. G. (1973), *A Random Walk Down Wall Street*, New York: Norton.
- Marquering, W. & M. Verbeek(2004), "The economic value of predicting stock index returns and volatility", *Journal of Financial and Quantitative Analysis* 39(2):407—429.
- Meese, R. A. & K. Rogoff(1983), "Empirical exchange rate models of the seventies: Do they fit out of sample?", *Journal of International Economics* 14(1—2):3—24.
- Moritz, B. & T. Zimmermann(2016), "Tree-based conditional portfolio sorts: The relation between past and future stock returns", SSRN Working Paper, No. 2740751.
- Moskowitz, T. J. et al(2012), "Time series momentum", *Journal of Financial Economics* 104(2): 228—250.
- Mullainathan S. & J. Spiess(2017), "Machine learning: An applied econometric approach", *Journal of Economic Perspectives* 31(2): 87—106.
- Neely, C. J. et al(2014), "Forecasting the equity risk premium: The role of technical indicators", *Management Science* 60(7):1772—1791.
- Park, C. H. & S. H. Irwin(2007), "What do we know about the profitability of technical analysis?", *Journal of Economic Surveys* 21(4):786—826.
- Pástor, L. & R. F. Stambaugh(2012), "Are stocks really less volatile in the long run?", *Journal of Finance* 67(2): 431—478.
- Pettenuzzo, D. & A. Timmermann (2011), "Predictability of stock returns and asset allocation under structural breaks", *Journal of Econometrics* 164(1):60—78.
- Pettenuzzo, D. et al(2014), "Forecasting stock returns under economic constraints", *Journal of Financial Economics* 114 (3):517—553.
- Rangvid, J. (2006), "Output and expected returns", *Journal of Financial Economics* 81(3):595—624.
- Rapach, D. E. , M. C. Ringgenberg & G. Zhou(2016), "Short interest and aggregate stock returns", *Journal of Financial Economics* 121(1):46—65.
- Rapach, D. E. , J. K. Strauss & G. Zhou(2010), "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy", *Review of Financial Studies* 23(2):821—862.
- Rapach, D. E. , J. K. Strauss, J. Tu & G. Zhou(2018), "Industry return predictability: A machine learning approach", SSRN Working Paper, No. 3120110.
- Rapach, D. E. & M. E. Wohar(2006), "Structural breaks and predictive regression models of aggregate US stock returns", *Journal of Financial Econometrics* 4(2):238—274.
- Rapach, D. E. & G. Zhou(2013), "Forecasting stock returns", in: G. Elliott & A. Timmermann(eds), *Handbook of Economic Forecasting*, Elsevier.
- Rozeff, M. S. (1984), "Dividend yields are equity risk premiums", *Journal of Portfolio Management* 11(1):68—75.
- Sadka, G. & R. Sadka(2009), "Predictability and the earnings-returns relation", *Journal of Financial Economics* 94 (1):87—106.
- Santos, T. & P. Veronesi(2005), "Labor income and predictable stock returns", *Review of Financial Studies* 19 (1):1—44.

- Stambaugh, R. F. (1999), "Predictive regressions", *Journal of Financial Economics* 54(3):375—421.
- Stock, J. H. & M. W. Watson(2004), "Combination forecasts of output growth in a seven-country data set", *Journal of Forecasting* 23(6):405—430.
- Tang, Y. & R. F. Whitelaw(2011), "Time-varying Sharpe ratios and market timing", *Quarterly Journal of Finance* 1(3):465—493.
- Timmermann, A. (2006), "Forecast combinations", in: G. Elliott, C. W. J. Granger & A. Timmermann (eds), *Handbook of Economic Forecasting*, Elsevier.
- West, K. D. (1996), "Asymptotic inference about predictive ability", *Econometrica* 64(5): 1067—1084.
- White, H. (2000), "A reality check for data snooping", *Econometrica* 68(5):1097—1126.
- Zhang, H. et al(2017), "Forecasting stock returns with model uncertainty and parameter instability", SSRN Working Paper, No. 3039844.
- Zhou, G. & Y. Zhou(2009), "Technical analysis: An asset allocation perspective on the use of moving averages", *Journal of Financial Economics* 92(3):519—544.

The Progress of Research on Return Predictability in the Capital Market

ZHANG Chunling¹ JIANG Fuwei² TANG Guohao³

(1. Shandong Technology and Business University, Jinan, China;

2. Central University of Finance and Economics, Beijing, China;

3. Hunan University, Changsha, China)

Abstract : Predicting asset returns with accuracy in the capital market is important for both asset management and financial studies. From the perspective of asset management, predicting asset returns is necessary for successful asset allocation. From the perspective of financial research, understanding the mechanism of return predictability helps researchers to develop more realistic and more accurate asset pricing theories and models. How to choose appropriate forecasting methods as well as predictors and investigate the sources of return predictability are hot topics in recent studies. This paper documents the research progress of return predictability in the capital market, especially in the stock market, starting with tests of return predictability and debates over the efficient market hypothesis. We introduce different forecasting predictors such as fundamental, technical, sentimental, and volatility-related predictors. We also review various methods to predict returns, including economically motivated model restrictions, latent factors, forecast combinations, regime shifts, and machine learning. After comparing these methods and predictors, this paper discusses how to evaluate forecast accuracy and alleviate the over-fitting problem.

Keywords : Capital Market; Return Predictability; Predictors; Asset Pricing

(责任编辑:刘洪愧)

(校对:刘新波)