迈向 Evology：美国股票共同基金的市场生态学代理模型 II

Aymeric Vi´e
Mathematical 研究所，牛津大学新经济思想研究所，牛津大学
vie@maths.ox.ac.uk

J. Doyne Farmer
Mathematical 研究所，牛津大学新经济思想研究所，牛津大学
圣菲研究所

arXiv:2302.01216v1 [cs.MA] 2 Feb 2023

摘要

基于智能体的模型（ABM）适用于模拟金融市场这类异质性互动系统。本文介绍 Evology 的最新进展：一种针对美国股票市场的异质性、经实证校准的市场生态学基于智能体的模型[1]。价格由市场参与者之间的互动及其多样化的投资行为，以及他们对基本面的反应内生地产生。这种方法能够在考量某一交易策略与其他市场参与者及市场状况之间相互作用的同时，对交易策略进行测试。这些初步结果表明，将基于智能体的模型与机器学习算法结合起来，可用于测试和优化投资策略。

1 引言

动机 金融市场是复杂适应系统，其参与者种类繁多。市场生态学理论 Farmer（2002）；LeBaron（2002）；Musciotto 等人（2018）；Lo（2019）；Levin 和 Lo（2021）；Scholl 等人（2021）借鉴生态学和生物学的概念来研究金融市场。交易策略类似于生物物种：它们利用市场低效率，并为生存或盈利而相互竞争。Scholl 等人（2021）强调了常见交易策略之间的互动特征，以及在风格化交易风格中，其收益率对密度的强烈依赖性。

相关工作 本研究延续了金融基于智能体的模型以及具有异质信念的市场选择这一广阔领域 Blume & Easley（1992；2006）。例如，近期已有若干基于智能体的模型被用于做市优化 Spooner 等人（2018）、理解闪崩 Paulin 等人（2019），以及为交易培训提供复杂的金融架构 Byrd 等人（2019）。我们尝试发展市场生态学 Farmer（2002）；Scholl 等人（2021）这一互补方法，重点关注不同类型代理和策略之间的生态互动。在这一视角下，基于智能体的模型尤为适用，因为其本质上强调智能体异质性及其相互作用。

贡献 本文报告 Evology 项目取得的进展。该项目是一个大规模的市场生态基于智能体建模项目，旨在解决若干紧迫问题：利用基于主体的模型预测市场变量；在金融基于智能体的模型中测试交易策略，以补充回测；以及研究市场参与者（例如散户投资者或指数基金）对波动性、价格和参与者规模的影响。建模工作已有所改进，纳入了更多市场参与者，并加入多只具有历史基本面的股票，基于主体的模型中的智能体利用这些基本面生成内生价格。该基于主体的模型如今支持测试用户指定的交易策略，并考虑传统回测可能忽略的异质性和交互效应。这些结果也进一步表明，基于智能体模型与机器学习可以更紧密地协作，在更真实的人工股票市场中测试和发现交易策略。Evology 的用户可以指定手工编码函数，而这一流程还可以实现自动化，从而在该仿真环境中利用机器学习搜索交易规则。随着模型现实性的提升，尤其是附录中所述研究方向的推进，Evology 中基于主体的模型交易策略测试的实用性和相关性也将增强。

[1]当前构建版本已在 GitHub 上以开源形式提供。

2 模型与校准

介绍 与 Evology 的先前版本及其他市场生态模型相比，Vie 等人（2022a；b）；Scholl 等人（2021），这一改进后的基于智能体的模型涵盖了种类更加丰富的市场参与者和金融资产。最初仅关注主动型共同基金的参与者范围，如今已扩展至国内/国际基金、指数基金、交易所交易基金和散户投资者。这些市场参与者交易多只股票的股份，而这些股票的基本面——例如收益、现金流、账面价值和股息——均取自历史数据。

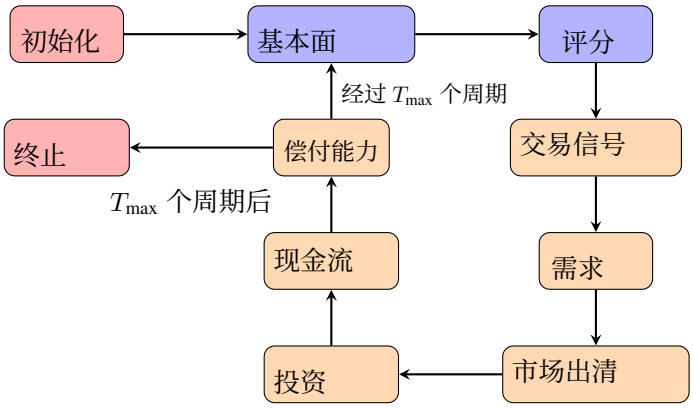

图 1：模型组成部分的可视化摘要。

校准 在模型初始化阶段，我们创建市场参与者（表 1）和股票。在此阶段，模型中的市场参与者采用聚合形式：例如，单个代理代表所有活跃价值型共同基金的集合。股票对应于实际股票，其基本面数据来自 Compustat IQ Market Intelligence；值得注意的是，模型从这些数据中提取股票的初始价格和股份数量。实体经济的活动是外生发生的，而基于主体的模型则以内生方式生成参与者之间的价格和财富动态。

模型步骤 在这个市场中，每天都会发生以下活动。公司会公布新的基本面信息，并将其提供给市场参与者；通常，每家公司每季度公布一次，我们引入一定延迟，使公告在一个63天的季度内均匀发布。参与者可根据这些基本面信息计算股票评分，以反映股票在不同标准下的相对吸引力。价值评分和成长评分目前采用晨星风格的分类尺度。例如，价值评分包括市盈率、市净率、市销率、现金流比率和股息率，而成长评分则涵盖盈利增长、收入增长等。市场参与者根据这些评分以及市值／权重等市场信息形成交易信号，这些信号作为投资组合权重，并决定参与者希望投资于各只股票的财富规模。他们根据未知股票价格制定需求函数，该函数取决于参与者的财富、交易信号、杠杆和当前股票持有情况。对于每只股票，汇总参与者的需求并求出由此产生的总需求函数的根，即可得到使供给与需求相匹配的市场出清价格。按照实际成交价格执行需求订单后，参与者可能会根据其表现获得资本投资，形式为现金流入／流出。Ha 与 Ko（2019）分析了 N-SAR 表格，发现基金的超额收益与投资流之间存在线性正相关关系，这表明外部投资者会追逐收益，也印证了 Chevalier 与 Ellison（1997）的早期研究结果。接下来，现金／股票会支付利息和股息现金流。在开始下一交易日之前，我们会检查参与者的偿付能力。附录介绍了该基于主体的模型的更多建模细节、校准结果以及需要改进的关键领域。

表 1：市场参与者及其行为。

参与者	行为规则
价值（主动型共同基金）	对排名靠前的 k 股票给予相同权重（按价值评分）
混合（主动型共同基金）	对价值与成长类顶尖股票等权配置
成长型（主动型共同基金）	按成长评分对顶尖股票等权配置
指数共同基金	复制整个股票市场权重
ETFs	归并为指数共同基金
国际共同基金	按相同比例归并为主动共同基金
散户投资者（家庭）	噪声交易，权重遵循奥恩斯坦-乌伦贝克过程
“策略基金”	用户指定

3 结果

3.1 基于代理的模型市场动态概览

Evology 展示了不同参与者按照不同交易规则互动，并对公司发布的基本面公告作出反应，从而生成逼真价格的能力（图2）。股票市场的活动也会影响各类市场参与者的财富及其财富占比。尽管这些结果令人鼓舞，但也凸显了许多有待改进之处。模拟运行具有随机性，目前价格受到散户投资者抽取的特定奥恩斯坦-乌伦贝克过程路径的显著影响。改进 Evology 时，应确保生成的价格与历史价格相符，以验证其预测能力。增加股票数量应能降低平均价格：目前价格过高（达到数千），这很可能是因为股票市场中的全部财富都集中在仅有的21只股票上。模拟运行的早期阶段也呈现出较高的波动性，因为一些股票的评分时而升至前列、时而跌出前列，导致投资于这些股票的财富骤然变化。

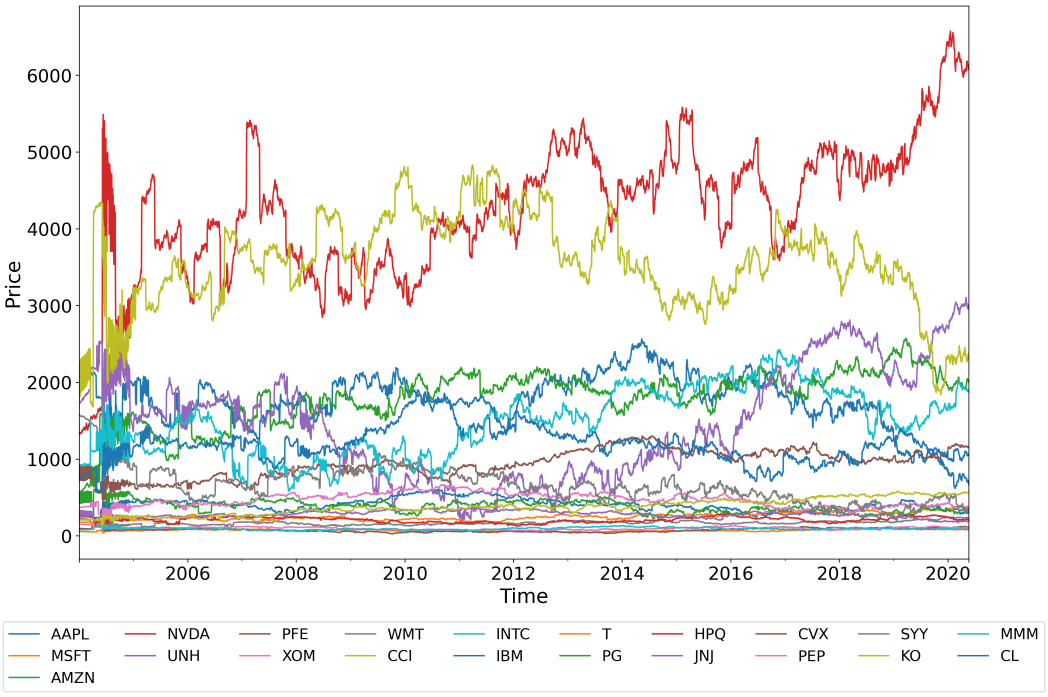

图 2：内生生成的价格序列。

3.2 在 Evology 中测试交易策略

除上述数据中提到的市场参与者外，基于主体的模型如今还包含一个规模小得多的参与者——“策略基金”，它采用用户指定的策略在市场中进行交易。作为回测的补充，这使得人们能够在基于主体的模型中测试交易策略，并评估其回报、波动性及其他特征，同时考虑市场影响、与其他参与者的互动以及拥挤效应 Farmer（2002）。使用此类基于主体的模型测试策略，可以克服回测中的一项假设，即策略不会影响用于评估它的数据 Vie 等（2022a）。图3展示了与图2相同实验中，策略基金采用一种简单策略（将全部财富投资于 AAPL 股票）时，其净资产值的演化。在此情形下，财富变化轨迹并非重点；更重要的是，基于主体的模型能够测试交易策略，同时让市场其他部分按照各自的交易规则运行。这一目标在先前研究中已有提出 Vie 等（2022a），并使基于主体的模型能够用作机器学习驱动的交易策略搜索中的评估环节。提升基于主体的模型的真实性，将有助于提高交易策略测试在现实世界中的相关性，并增加人工智能/机器学习算法的研究兴趣与挑战。

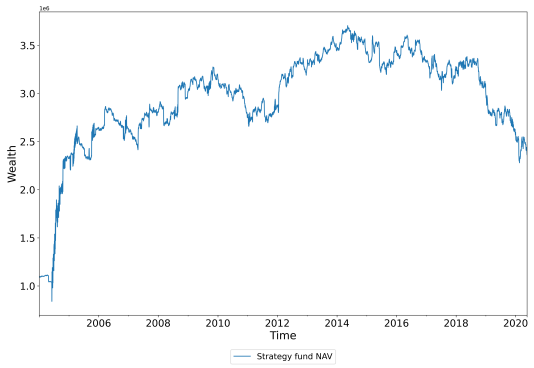

图3：用户指定的策略基金净资产值，采用“100% 投资于 AAPL”策略

致谢

本出版物基于以下项目的支持性工作：EPSRC 随机系统数学博士培训中心：分析、建模与模拟（EP/S023925/1），以及 Baillie Gifford 的支持。

参考文献

Lawrence Blume 和 David Easley。演化与市场行为。经济理论杂志 (1):9–40 1992。

Lawrence Blume 和 David Easley。既然你这么聪明，为什么还不富有？完全市场与不完全市场中的信念选择。*Econometrica*，74(4):929–966，2006。

David Byrd、Maria Hybinette 和 TuckerHybinette Balch。ABIDES：面向人工智能研究的高保真市场仿真。*arXiv* 预印本 *arXiv:1904.12066*，2019。

Judith Chevalierand Glenn Ellison。共同基金对激励的风险承担反应。政治经济学杂志，105(6):1167–1200，1997。

Rama Cont. 资产收益的实证特征：典型事实与统计问题。定量金融，1(2):223，2001。

Rama Cont. 金融市场中的波动率聚集：实证事实与基于智能体的模型。载于经济学中的长记忆，第289–309页。施普林格，2007。

J Doyne Farmer. 市场力量、生态学与演化。产业与公司变革，11(5): 895–953, 2002.

Yeonjeong Ha 和 Kwangsoo Ko. 基金流量与业绩关系中的设定错误。《金融中介》，38:69–81,2019.

投资公司协会。ICI，2022。

Blake LeBaron。构建圣菲人工股票市场。*Physica A*，第 1–20 页，2002。

Simon A Levin 和 Andrew W Lo。关于金融市场演化模型专题的介绍。《美国国家科学院院刊》，118(26)，2021。

Andrew W Lo。适应性市场假说。普林斯顿大学出版社，2019。

Federico Musciotto、Luca Marotta、Jyrki Piilo 和 Rosario N Mantegna。金融市场中投资者的长期生态。*PalgraveCommunications*，4(1)：1–12，2018。

James Paulin、Anisoara Calinescu 和 Michael Wooldridge。复杂金融系统的基于智能体的建模。《*IEEE*智能系统》，33(2):74–82，2018。

James Paulin、Anisoara Calinescu 和 Michael Wooldridge。理解闪电崩盘传染与系统性风险：微观—宏观基于智能体的方法。《经济动态与控制杂志》，100:200–229，2019。

Sebastian Poledna、Stefan Thurner、J Doyne Farmer 和 John Geanakoplos。巴塞尔协议II及其他信用风险政策下杠杆引发的系统性风险。《银行与金融杂志》，42: 199–212，2014。

Maarten P Scholl、Anisoara Calinescu 和 J Doyne Farmer。市场生态学如何解释市场etm故障。《美国国家科学院院刊》，118(26)，2021。

Thomas Spooner、John Fearnley、Rahul Savani 和 Andreas Koukorinis。通过强化学习做市。*arXiv*预印本 *arXiv:1804.04216*，2018。

Aymeric Vie、Maarten Scholl、Alissa M Kleinnijenhuis 和 Doyne J Farmer。《迈向 Evology：美国股票共同基金的市场生态学基于智能体的模型》。*arXiv* 预印本 *arXiv:2210.11344*，2022a。

Aymeric Vie、Maarten Peter Scholl 和 Doyne James Farmer。Evology：用于交易策略搜索的经实证校准的市场生态学基于智能体的模型。载于*ICML 2022* 基于智能体建模人工智能研讨会，2022b。

A 附录

A.1 建模细节

A.1.1 资产需求

基金的每日交易信号是资产超额需求函数 的输入，Scholl 等人（2021）。对于每只股票，超额需求函数将基金对该资产的需求表示为未知价格 $p(t)$ 的函数。基金财富 W 是智能体现金、股票份额现值和负债之和。我们的需求函数包含最大杠杆 λ 和策略激进度 β。我们的需求函数表示一位投资者，其资产头寸为 $S(t)$、预算为$\lambda W(t)$，并将其预算的 $\phi(t)$ 用于该资产，Poledna 等（2014）；Scholl 等人（2021）。

$$D(t, p(t)) = \beta\phi(t)\frac{\lambda W(t)}{p(t)} - S(t) \tag{1}$$

A.1.2 市场出清

市场出清过程找出使基金需求总和等于固定资产供给 Q 的价格，即供需匹配，Poledna 等（2014）。这等同于 Scholl 等人（2021）所述的市场出清程序，即求解总超额需求函数的根。每只股票的市场出清条件为 $\sum_i D_i(t,\ p(t)) = Q$。尽管许多金融智能体模型使用限价订单簿（LOBs），我们的研究重点是数十年至数百年的长时间尺度，而这些模型关注的是日内动态。我们并不排除今后在本模型中使用限价订单簿的可能性，但我们认为，在所考察的时间范围内，市场出清是一种有效的近似。

A.1.3 股票评分

表 2：晨星成长评分。YOY 表示同比增长。

标准	**权重**
长期预期盈利增长（%，截至目前）	50%
历史盈利增长（%，同比）	12.5%
销售增长（%，同比）	12.5%
现金流增长（%，同比）	12.5%
账面价值增长（%，同比）	12.5%

表3：晨星价值评分

标准	**权重**
股价与预期收益比	50%
市净率	12.5%
市销率	12.5%
市价现金流比	12.5%
股息率（%）	12.5%

A.2 模型验证——资产价格的实证风格化事实（Cont（2001））

生成所谓的金融“风格化事实”是验证金融基于智能体的模型的常见要求。我们的模型再现了 Cont（2001）提出的资产价格主要风格化事实。我们的对数价格呈现间歇性。对数价格收益在平凡频率之外未显示出显著的线性自相关。与正态分布相比，收益率呈现厚尾分布和超额峰度。我们还可以再现杠杆效应——价格收益与波动性之间的负相关——以及成交量与波动率的正相关性，并且绝对收益的自相关衰减缓慢。

自相关缺失资产收益的自相关应不显著，极短时间尺度除外，因为微观结构会产生一定影响。

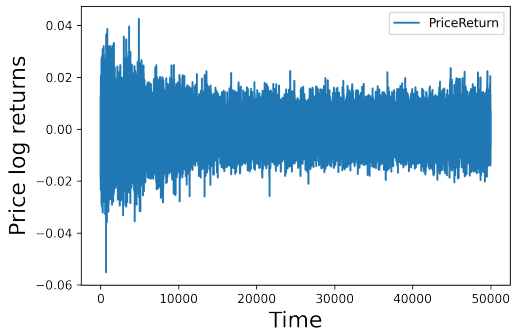

图 4：资产价格的日对数收益率。对于价格 $p(t)$，日时间尺度上的对数收益率为 $r(t) = \ln p(t) - \ln p(t-1)$。

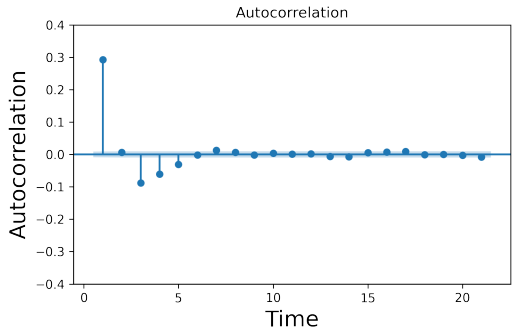

图 5：资产价格对数收益率在最多 21 个周期内的自相关函数。经过 5 个周期的短时间尺度后，自相关与零值无显著差异，这证实了自相关缺失。

厚尾收益率的无条件分布应呈现幂律分布或帕累托尾部，且方差有限，因此不属于正态分布。

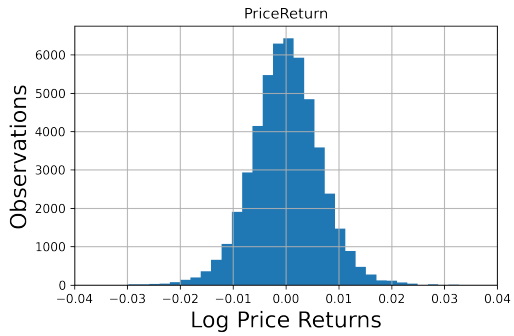

图 6：日对数价格收益率的直方图显示出厚尾。该序列的费雪（或超额）峰度 (κ) 值为 1。3，高于正态分布的费雪峰度，而正态分布的费雪峰度为
is equal to 0. $\kappa = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3.$

收益/损失不对称性 人们应当观察到股票价格出现大幅回撤，却没有同样大幅的上涨。

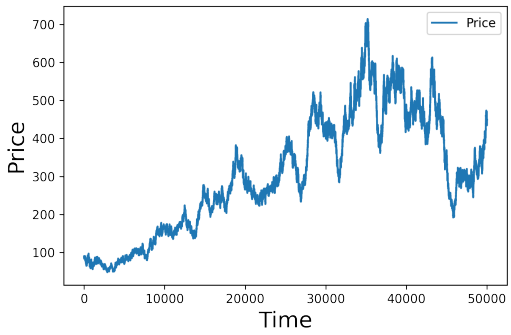

图 7：模拟运行中的资产价格序列。图形分析显示，价格下跌比上涨更快：从回撤中恢复通常需要比回撤持续时间更长的周期。

聚合高斯性 随着计算收益率所用时间尺度的增加，其分布应当越来越接近正态分布。

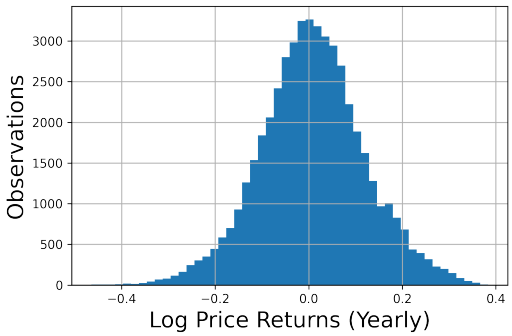

图 8：年度对数价格收益的分布比图 6 中的日度对数价格收益分布更接近正态分布。从定量角度看，月度收益的超额峰度较低，为 0.85；年度收益的超额峰度则更低，为 0.32。因此，我们证实了聚合高斯性。

间歇性 收益率应表现出较高的变异性，这可从波动率估计量时间序列中不规则的突发波动看出。

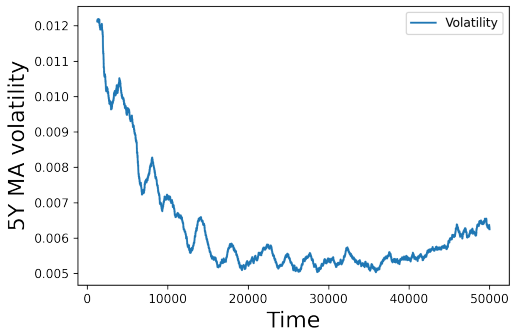

图 9：随时间变化的 5年移动平均波动率显示出间歇性。我们认为，模拟运行初期波动率相对较高，是由基于智能体的模型初始化所导致的暂态。不过，图 4 中对数收益率的不规则突增并不明显，这或许是因为图中涵盖的周期较多。

波动率聚集与绝对收益的自相关缓慢衰减不同的波动率度量应在数日内呈现正自相关，这表明高波动事件往往会随时间聚集。我们可以从绝对收益自相关函数的缓慢衰减中观察到这种波动率聚集。

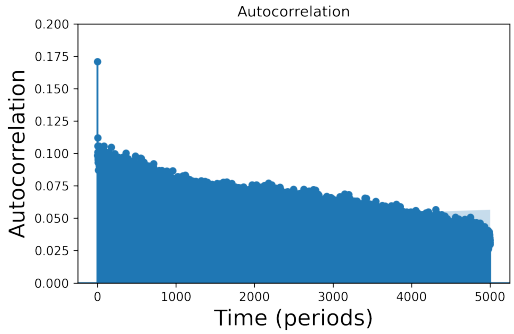

图 10：绝对对数价格收益率的自相关函数显示出长程依赖，且随时间缓慢衰减。自相关在多个时间滞后下仍显著为正，表明存在波动率聚集（Cont（2007））。

杠杆效应资产波动率度量应与资产收益率负相关。在我们的模拟运行中，这一皮尔逊相关系数为负，且等于 -0.005。

成交量与波动率相关性交易量应与波动率度量相关。在整个模拟运行过程中，这一皮尔逊相关系数为正，等于 0.48。此外，交易量与价格收益呈正相关（0.001）。

A.3 改进 Evology

- 纳入此前缺失的市场参与者，尤其是养老基金、现金流和对冲基金。风险转移活动还需要加入中介类市场参与者（做市商、期货/交易所交易基金/期权套利者、批发商）。
- 更真实地模拟利率，将固定利率框架改为动态利率。可以引入外生的历史序列，改善利率现金流，并为内生价格加入估值效应，也就是说，价格的变动不仅由基本面的变化解释，也由贴现率的变化解释。

• 改进校准对基于主体的模型至关重要（Paulin 等人（2018）），且应遵循资产／共同基金风格化事实（Cont（2001；2007）；Institute（2022））。具体而言，在我们的案例中，应利用机构参与者的监管申报文件（尤其是 13F 表格），并利用其他来源（例如美联储家庭资产负债表）来校准不同时期的参与者规模。未来，Evology 的运行应能够涵盖特定周期（例如“01-01-1997”至“31-12-2022”），同时仍能实现准确校准。
• 为各类市场参与者制定数据驱动的行为规则，以便更好地捕捉不同投资风格的决策过程，至少是在总体层面上。鉴于散户投资者在美国股票市场持有量中的重要性，更准确地评估其行为及交易活动的决定因素尤为重要。

• 继续拓展股票范围。从 Scholl 等人（2021）和 Vie 等人（2022a；b）中的 1 只风格化股票，到目前版本的 Evology 已纳入 21 只具有历史基本面数据的股票。随着模型逐步扩展至数百只甚至数千只股票，使其更贴近现实，探索市场参与者对小盘、中盘和大盘股的区分将会很有意义。还需要对数据进行一些调整，以纳入在模拟期间开始交易的股票，尤其是 TSLA 和 FB/META。

• 通过检查价格收益、参与者收益、换手率和活动统计数据（如流动性份额），确保参与者交易规则的真实性。
• 改进各类参与者初始股票头寸的初始化，并寻找有关这些参与者股票持有情况的新数据。更准确的初始头寸将缓解模拟开始时的波动性，因为参与者会重新配置投资组合，以符合其交易规则。