Shifting Power: Leveraging LLMs to Simulate Human Aversion in ABMs of Bilateral Financial Exchanges, A bond market study

Extended Abstract

Alicia Vidler
UNSW
Sydney, Australia

Toby Walsh
UNSW
Sydney, Australia

ABSTRACT

Bilateral markets, such as those for government bonds, involve decentralized and opaque transactions between market makers (MMs) and clients, posing significant challenges for traditional modeling approaches. To address these complexities, we introduce **TRIBE** an agent-based model (ABM) augmented with a large language model (LLM) to simulate human-like decision-making in trading environments. TRIBE leverages publicly available data and stylized facts to capture realistic trading dynamics, integrating human biases like risk aversion and ambiguity sensitivity into the decision-making processes of agents. Our research yields three key contributions: first, we demonstrate that integrating LLMs into ABMs, to enhance client agency, is feasible and enriches the simulation of agent behaviors in complex markets; second, we find that even slight trade aversion encoded within the LLM leads to a complete cessation of trading activity, highlighting the sensitivity of market dynamics to agents' risk profiles; third, we show that incorporating human-like variability shifts power dynamics towards clients and can disproportionately affect the entire system, often resulting in systemic agent collapse across simulations. These findings underscore the emergent properties that arise when introducing stochastic, human-like decision processes, revealing new system behaviors that enhance the realism and complexity of artificial trading societies.

KEYWORDS

Multi-agent systems; Large Language models; Agents; Financial markets

ACM Reference Format:

Alicia Vidler and Toby Walsh. 2025. Shifting Power: Leveraging LLMs to Simulate Human Aversion in ABMs of Bilateral Financial Exchanges, A bond market study: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

ABMs are versatile applications for modeling complex and dynamic systems, particularly suited for bilateral markets like government bond markets. These markets, characterized by direct transactions between two parties without centralized exchanges, present modeling challenges due to their decentralized nature, complex interactions between heterogeneous agents, and lack of transparency. We introduce the *TRIBE* model, a generative-ABM focused on **T**rading **R**elationships, **I**nteractions, and **B**ilateral **E**xchange of assets. *TRIBE* incorporates client agency with dynamically assigned asset distributions and probabilistic trading availability, extending this approach by integrating a LLM for more human-like decision-making and negotiation behavior.

TRIBE focuses on the Australian government bond market, a decentralized market emphasizing liquidity over pricing. This structure provides an ideal environment for exploring how human-like behaviors, such as trade aversion and ambiguity sensitivity, influence market dynamics and stability.

2 RELATED WORK

ABMs simulate complex financial market behaviors but often use homogeneous agents and publicly transparent, high visibility, environments - less applicable in opaque over-the-counter (OTC) markets [7, 20]. [12] extends ABM theory by introducing network agency, showing how external features influence agent behaviors, though data limitations persist in OTC markets [18].

Recent developments in LLMs like GPT-4 improve agent decision-making [6, 11], despite challenges in reasoning and financial applicability [1, 2, 22, 23]. Prior research has explored LLMs for simulating complex human behaviors [19], and we enhance reliability by focusing on the state-of-the-art GPT4o-mini-2024-07-18 and developing a risk management framework [21].

Calibrating ABMs involves varied methodologies [4]. We use inductive approaches with data from [16, 20] and build on existing frameworks [9]. The Australian bond market serves as our research foundation, leveraging studies on liquidity and bilateral exchanges [13, 17].

While traditional ABMs employ deterministic or basic stochastic models [7], integrating LLMs introduces human-like variability and enhances simulation capabilities [15]. Our analysis considers behavioral changes like trade aversion, influenced by trading psychology and ambiguity [5, 8, 10].

3 MODEL DESCRIPTION

TRIBE is a bespoke ABM uses to simulate the trading activities of heterogeneous MMs dealing in the stylized Australian government bond market. We utilize a series of artificial trading simulations (200 epochs for each experiment) with parameter calibration from literature and government data ([16], [20], [13], [3], [14]). TRIBE incorporates advanced features in a series of experiments where TRIBE(LLM) (Experiment 3) integrates the use of a LLM, specifically GPT4o-mini-2024-07-18, to enhance model decision-making. We focus on two key metrics: average agent lifespan across simulations and percentage of environment assets that are transacted.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA.

3.1 Agent Design

TRIBE models two key agent types:

- **Market Makers (MMs)**: Adaptive agents with individual utility functions that trade bonds and cash with clients and other MMs.
- **Clients**: Grid-based agents initialized with bonds and cash from a log-normal distribution, reflecting real-world asset distributions. Clients have probabilistic trading availability & dynamic preferences influenced by LLM-driven decisions.

3.2 LLM Integration

A defining feature of **TRIBE** is the use of an LLM to simulate client decision-making (Experiment 3). At each time step, the LLM determines whether a client is willing to trade *right now* with a market maker, introducing variability and capturing behaviors like risk aversion and ambiguity sensitivity. **Prompt Design**: Client agents provide to an LLM a unique prompt (with bond and cash holdings etc), asking if they wish to trade at that time step. **Decision Outcomes**: The LLM's response (trade or not) is binary. If a client chooses to trade, the direction (buy or sell) is determined probabilistically.

3.3 Model Dynamics

TRIBE operates over discrete time steps, simulating interactions between MMs and clients. Key components include: **Asset Distribution**: Clients' bonds and cash are initialized using publicly available data, following a log-normal distribution. **Client Availability**: Clients are only available for trading at certain time steps, modeled probabilistically. **Decision-Making Process**: For each available client, the LLM decides whether to trade. If yes, a Bernoulli distribution determines the trade direction. **Market Maker Strategies**: MMs adapt to trading opportunities, optimizing their portfolios while maintaining market liquidity.

3.4 Emergent Properties and Client Realism

Integrating LLM-driven decisions introduces stochastic behaviors that mirror human unpredictability, leading to emergent properties in market dynamics:

- **Probabilistic Availability**: Reflecting real-world market patterns, not all clients are available at each time step.
- **Dynamic Preferences**: Clients' trade decisions fluctuate based on LLM-driven reasoning and current holdings.
- **Systemic Impact**: The stochastic nature of LLM-driven decisions can shift power towards clients, destabilizing the system and leading to market collapse.

4 EXPERIMENTS & RESULTS

(1) Experiment 1 (Benchmark ABM): A traditional ABM without LLMs tested with 20% average client availability. **Results:** Stable markets, high trading volumes, long agent lifespans.
(2) Experiment 2 (LLM with Trade Aversion): LLM prompts include aversion-related language. **Results:** Complete cessation of trading, rapid systemic collapse, average simulation lifespan reduced to 27 time steps.
(3) Experiment 3 (LLM with Timeliness Focus): LLM prompts focused on timeliness rather than aversion. **Results:** Increased short-term variability in trading decisions, 22.8% average availability with very high variance, shifting power to clients, frequent market instability, significant reduction in agent lifespans.

Table 1: Experimental Results Summary

Experiment	**Avg. Agent Lifespan (max 1,500)**	**Client Asset Trading (%)**	**Systemic Stability**
1. ABM	1136 steps	90%	High
2. LLM Aversion	27 steps	0%	Collapse
3. ABM + LLM	365 steps	7%	Unstable

5 LLM-DRIVEN VARIABILITY AND MARKET IMPACT

Our simulations reveal that LLM-generated yes/no trade decisions average 57% over time, equating to a long run average of 22.8% client availability, directly comparable to Experiment 1. However, we see that using an LLM, short-term (10 sample) fluctuations in average yes/no availability vary from 0% to 100%. This large-frequency variability disrupts market stability, as clusters of negative responses hinder MMs' ability to maintain liquidity, often leading to premature market collapse.

6 DISCUSSION AND CONCLUSION

Our research introduces **TRIBE**, a novel ABM integrating an LLM to simulate human-like decision-making in bilateral markets. We demonstrate how subtle behavioral variations, expressed through LLMs, can fundamentally reshape market dynamics, potentially leading to systemic instability and market cessation. The emergent properties reveal how slight variability, as modeled through an LLM, shifts power dynamics between participants, towards clients, with profound implications for market design and regulation.

These findings underscore the potential of LLMs to enhance the realism of ABMs, providing a powerful tool for exploring market dynamics, testing regulatory interventions, and advancing the field of computational finance. Future work will focus on refining LLM prompts, exploring additional market types, and extending the model to multi-agent negotiation scenarios.

ACKNOWLEDGMENTS

This work is funded in part by an ARC Laureate grant FL200100204 and NSF-CSIRO grant to Toby Walsh. A complete version of TRIBE is available online at arXiv.org

REFERENCES

[1] Sahar Abdelnabi and Mario Fritz. 2024. Cooperation, Competition, and Maliciousness: LLM-Stakeholders Interactive Negotiation.
[2] Janice Ahn, Rishu Verma, Renze Lou, Di Liu, Rui Zhang, and Wenpeng Yin. 2024. Large Language Models for Mathematical Reasoning: Progresses and Challenges. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*. Association for Computational Linguistics, St. Julian's, Malta, 225–237.

[3] Australian Prudential Regulation Authority. 2003. Annual Fund-level Superannuation Statistics. https://www.apra.gov.au/annual-fund-level-superannuation-statistics. Accessed: 2024-08-05.

[4] Priscilla Avegliano and Jaime Simão Sichman. 2019. Using surrogate models to calibrate agent-based model parameters under data scarcity. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1781–1783.

[5] Peter Bossaerts, Paolo Ghirardato, Serena Guarnaschelli, and William R. Zame. 2010. Ambiguity in Asset Markets: Theory and Experiment. *The Review of Financial Studies* 23, 4 (01 2010), 1325–1359. https://doi.org/10.1093/rfs/hhp106 arXiv:https://academic.oup.com/rfs/article-pdf/23/4/1325/24430084/hhp106.pdf

[6] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, John A. Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuan-Fang Li, Scott M. Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. https://api.semanticscholar.org/CorpusID:257663729

[7] Rama Cont, Mihai Cucuringu, Vacslav Glukhov, and Felix Prenzel. 2023. Analysis and modeling of client order flow in limit order markets. *Quantitative Finance* 23, 2 (2023), 187–205. https://doi.org/10.1080/14697688.2022.2150282

[8] Larry Epstein and Martin Schneider. 2010. Ambiguity and Asset Markets. https://EconPapers.repec.org/RePEc:nbr:nberwo:16181

[9] Steven Mann Frank J. Fabozzi. 2005. *The Handbook of Fixed Income Securities, Seventh Edition*. McGraw-Hill, New York.

[10] Cosmin L Ilut and Martin Schneider. 2022. *Modeling Uncertainty as Ambiguity: a Review*. Working Paper 29915. National Bureau of Economic Research. https://doi.org/10.3386/w29915

[11] Haipeng Luo, Qingfeng Sun, Can Xu, Pu Zhao, Jianguang Lou, Chongyang Tao, Xiubo Geng, Qingwei Lin, Shifeng Chen, and Dongmei Zhang. 2023. WizardMath: Empowering Mathematical Reasoning for Large Language Models via Reinforced Evol-Instruct. arXiv:2308.09583 http://arxiv.org/abs/2308.09583

[12] Zakaria Mehrab, Logan Stundal, Samarth Swarup, Srinivasan Venaktramanan, Bryan Lewis, Henning Mortveit, Christopher Barrett, Abhishek Pandey, Chad Wells, Alison Galvani, Burton Singer, David Leblang, Rita Colwell, and Madhav Marathe. 2024. Network Agency: An Agent-based Model of Forced Migration from Ukraine. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems* (Auckland, New Zealand) *(AAMAS '24)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1372–1380.

[13] Reserve Bank of Australia. 2021. Government-bond-markets-in-advanced-economies-during-the-pandemic. , 52 - 60 pages. www.rba.gov.au/publications/bulletin/2021/sep/government-bond-markets-in-advanced-economies-during-the-pandemic.html

[14] Bank of England. 2015. GEMM Guidebook: A guide to the roles of the DMO and Primary Dealers in the UK Government bond market. https://dmo.gov.uk/media/ownmub2i/guidebook20150401.pdf

[15] Joon Park, Joseph O'Brien, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. arXiv:2304.03442 [cs.HC] https://arxiv.org/abs/2304.03442

[16] Gabor Pinter. 2023. An anatomy of the 2022 gilt market crisis. https://doi.org/10.2139/ssrn.4488189

[17] Gabor Pinter, Chaojun Wang, and Junyuan Zou. 2024. Size Discount and Size Penalty: Trading Costs in Bond Markets. *Review of Financial Studies* 37, 7 (2024), 2156–2190. https://doi.org/10.1093/rfs/hhae007

[18] Nelson Vadori, Leo Ardon, Sumitra Ganesh, Thomas Spooner, Selim Amrouni, Jared Vann, Mengda Xu, Zeyu Zheng, Tucker Balch, and Manuela Veloso. 2022. Towards Multi-Agent Reinforcement Learning driven Over-The-Counter Market Simulations. arXiv:2210.07184 [cs.MA]

[19] Alicia Vidler and Toby Walsh. 2024. TraderTalk: An LLM Behavioural ABM applied to Simulating Human Bilateral Trading Interactions . In *2024 IEEE International Conference on Agents (ICA)*. IEEE Computer Society, Los Alamitos, CA, USA, 164–167. https://doi.org/10.1109/ICA63002.2024.00042

[20] Alicia Vidler and Toby Walsh. 2024. Decoding OTC Government Bond Market Liquidity: An ABM Model for Market Dynamics. arXiv:2501.16331 [q-fin.TR] https://arxiv.org/abs/2501.16331

[21] Alicia Vidler and Toby Walsh. 2025. Evaluating Binary Decision Biases in Large Language Models: Implications for Fair Agent-Based Financial Simulations. arXiv:2501.16356 [cs.LG] https://arxiv.org/abs/2501.16356

[22] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388

[23] Yadong Zhang, Shaoguang Mao, Tao Ge, Xun Wang, Adrian Wynter, Yan Xia, and Wenshan Wu. 2024. LLM as a Mastermind: A Survey of Strategic Reasoning with Large Language Models. arXiv:2404.01230 [cs.CL] https://arxiv.org/abs/2404.01230