Enhancing macroeconomic Agent-Based Models (ABMs) with Inverse Reinforcement Learning

Data-anchored household objectives for crisis-policy evaluation

Elena Lickel

Department of Computer Science, University of Oxford

UNIVERSITY OF OXFORD

15 January 2026

What are macroeconomic Agent-Based Models (ABMs)?

Bottom-up simulation with heterogeneous agents

- An **Agent-Based Model (ABM)** represents the economy as interacting agents: **households, firms, banks, government**.
- Agents face constraints (income, debt, liquidity, credit limits) and follow decision rules.
- Macro outcomes emerge endogenously: **defaults, unemployment, inequality, contagion**.

State of the art: Multi-Agent Reinforcement Learning

- ABMs increasingly use **reinforcement learning (RL) / multi-agent RL (MARL)** to replace fixed heuristics.
- Benefit: households for example can **adapt** when constraints tighten and policy rules change (crisis regimes).
- This supports crisis analysis:
  - policy counterfactuals (transfers, moratoria, credit support)
  - nonlinear cascades (defaults $\rightarrow$ bank stress $\rightarrow$ credit crunch)
  - distributional outcomes (effects differ by liquidity and debt exposure)

State of the art: Multi-Agent Reinforcement Learning

Why learning agents helps in crises

- ABMs increasingly use **reinforcement learning (RL) / multi-agent RL (MARL)** to replace fixed heuristics.
- Benefit: households for example can **adapt** when constraints tighten and policy rules change (crisis regimes).
- This supports crisis analysis:
  - policy counterfactuals (transfers, moratoria, credit support),
  - nonlinear cascades (defaults $\rightarrow$ bank stress $\rightarrow$ credit crunch),
  - distributional outcomes (effects differ by liquidity and debt exposure).

Limitation

Learning behaviour still requires specifying **what households optimise**.

Bottleneck: the reward/objective is hand-designed

MARL improves adaptation, not the underlying assumptions

- In practice, household rewards are often **hand-designed** (utility proxies, smoothing, default penalties).
- Under crisis regimes, multiple objectives are plausible; the reward is **underdetermined**.
- Small reward changes can shift behaviour: **buffer use**, **deleveraging vs default**, **risk-taking**.

Policy rankings can change under equally plausible reward choices.

Why the reward matters for Ethics

The reward encodes normative trade-offs

Concrete crisis-policy question

In a shock, how should policy trade-off **default prevention**, **consumption stability (essentials)**, and **inequality**?

- The reward is a **value-laden behavioural assumption**: it defines what counts as “good outcomes”.
- If it is implicit, conclusions are hard to **justify**, **audit**, or **contest**.
- Ethical stakes: reward choices can change **who benefits** from interventions (e.g., moratoria vs transfers).

Ethics goal $\Rightarrow$ Make behavioural assumptions **explicit and auditable**, then stress-test policy conclusions.

Proposed solution: Inverse Reinforcement Learning (IRL) to infer household objectives

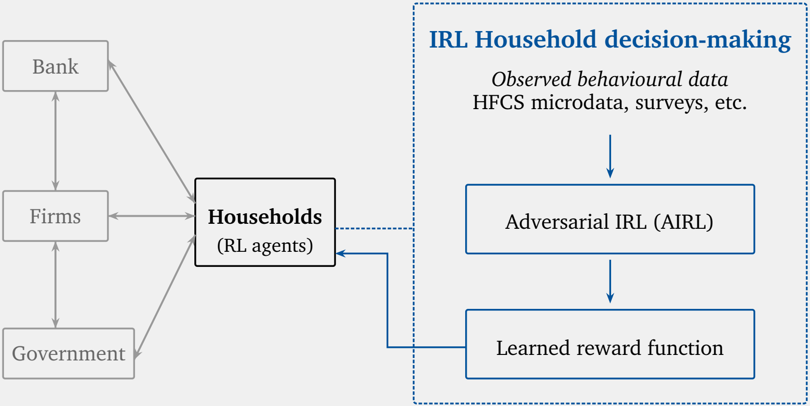
From reward engineering to data-anchored objectives

- In RL: objective $\rightarrow$ behaviour.
- **In IRL: behaviour $\rightarrow$ objective.**
- Use anonymised microdata to infer an objective that explains observed choices under constraints.
- Output: an **inspectable objective** used as the household reward inside the ABM.

Adversarial IRL as proposed implementation

AIRL is designed to recover objectives that transfer better across regime changes (policy/shock changes).

Approach in one figure



Keep the RL architecture; replace hand-designed household rewards with AIRL-inferred objectives.

Data is central (and so are the ethical constraints)

Anonymised microdata + governance + proportional use

- Data: anonymised household microdata (e.g., **HFCS**).
- Signals (examples): income, assets/liquidity buffers, liabilities/debt service, housing proxies, employment status.

Ethical safeguards

- Approved access via formal application and secure conditions;
- **data minimisation** (only what is needed);
- **purpose limitation** (objective inference + subgroup policy evaluation);
- **no re-identification/linkage**;
- reporting only in aggregates/subgroups.

Conclusion

Data-anchored objectives for transparent crisis-policy simulation

- ABMs are useful for crises because they capture heterogeneity, constraints, and cascades.
- MARL improves adaptation, but the **reward/objective** remains the normative bottleneck.
- IRL/AIRL makes that objective **explicit, inspectable, and data-anchored**.
- This supports **auditable assumptions** and **robustness-tested** policy rankings under strong data governance.

Thank you for your attention!

Elena Lickel
elena.lickel@cs.ox.ac.uk

References I

[1] J. Fu, K. Luo, and S. Levine, “Learning robust rewards with adversarial inverse reinforcement learning,” *arXiv preprint arXiv:1710.11248*, 2018.

[2] S. Brusatin, M. Tedeschi, M. Gallegati, and D. Delli Gatti, “Simulating the economic impact of rationality through reinforcement learning and agent-based modelling,” in *Proceedings of the ACM International Conference on AI in Finance (ICAIF)*, 2024.

[3] T. R. Cook and N. M. Palmer, “Reinforcement learning in macroeconomics,” *Oxford Research Encyclopedia of Economics and Finance*, 2025.

[4] European Central Bank, “Household finance and consumption survey (hfcs).” https://www.ecb.europa.eu/stats/ecb_surveys/hfcs/html/index.en.html, 2025.

[5] L. Yu, J. Song, and S. Ermon, “Multi-agent adversarial inverse reinforcement learning,” in *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.