# Learning Solutions in Large Economic Networks using Deep Multi-Agent Reinforcement Learning

Extended Abstract

Michael Curry
University of Zurich
Zurich, Switzerland
curry@cs.umd.edu

Alexander Trott*
Salesforce Research
United States

Soham Phade
Salesforce Research
United States

Yu Bai
Salesforce Research
United States

Stephan Zheng*
Salesforce Research
United States

## ABSTRACT

Real-world economies can be modeled as a network with many *heterogeneous* and *strategic* agents. In this setting, it is very challenging to find optimal mechanisms, e.g., taxes, 1) when taking strategic best responses into account and 2) even when using restrictive assumptions, e.g., that supply always meets demand. Deep multi-agent reinforcement learning (MARL) is a natural framework to *learn* mechanisms and model strategic best responses, but independent MARL often collapses to trivial solutions (e.g., where nobody works) as joint exploration severely distorts rewards and constraints. Here, we show how to use structured learning curricula and GPU-accelerated simulations to find non-trivial solutions in networks with many heterogeneous agents. We validate our approach in models with 100 worker-consumers, 10 firms, and a social planner who taxes and redistributes. We use empirical best-response analyses across agent types to show that it is difficult for agents to benefit by deviating from the learned solutions. In particular, we find income and corporate taxes that achieve 15% higher social welfare compared to baselines.

## KEYWORDS

multi-agent RL; economics; tax policy

**ACM Reference Format:**

Michael Curry, Alexander Trott, Soham Phade, Yu Bai, and Stephan Zheng. 2023. Learning Solutions in Large Economic Networks using Deep Multi-Agent Reinforcement Learning : Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

## 1 INTRODUCTION

In many (dynamic) general equilibrium (DGE) models of economic systems, consumers and firms engage in production and trade, and a social planner (the government) sets policies in order to achieve a (set of) desired social outcome(s). Since such economic systems can be seen as general-sum sequential imperfect-information games with many heterogeneous agents [25], it can be challenging for the social planner to find good public policies in the face of strategic behavior by the economic agents.

*Our goal.* Here, we aim to find a social planner (tax) policy that empirically achieves higher social welfare than fixed baselines, and where the other agents play an (approximate) best response. In particular, *we focus on the methodological challenge of using multi-agent RL for finding empirical solutions in complex economic systems.*

In particular, we analyze models with a large number of heterogeneous agents. To make this feasible, we 1) run behavioral models and simulation on the GPU using WarpDrive [23], and 2) generalize the *curriculum learning approach* from Zheng et al. [46].

This approach overcomes significant limitations of existing theoretical and computational methods [21, 36]. Moreover, to make models tractable, many strong simplifying assumptions have to be made, e.g., that there are a small number of representative agents or goods [20, 35]. As such, our work is a step towards real-world economic modeling which requires analyzing a wide spectrum of possible outcomes and solutions in diverse simulations [13].

## 2 RELATED WORK

DGE models describe the relationships and behavior of aggregate economic variables, such as productivity, consumption, savings, etc. Mathematically, they are akin to a system of temporal (partial) differential equations. Microfoundations research bases such models on individual agents: consumers, firms, and governments [24, 35]. Solving the stochastic game defined by a DGE is very difficult in general. Such models thus may represent the various interactions and context of agents well, but have to make many unrealistic assumptions to become tractable.

Another approach to economic modeling is agent-based modeling [5], which studies emergent phenomena in simulations of populations of interacting agents. Often, though such agent-based models use relatively simple (sub-optimal) behavioral rules. Thus, although the environmental simulation might have more realistic features, and there is no need to make representative agent assumptions, the agents themselves may not behave realistically.

Ideally, one could preserve some advantages of agent-based modeling when studying DGE models, while also allowing agents to adapt strategically and rationally. Multi-agent RL is a natural conceptual fit for this goal, because it allows for agents to learn to

*Work done while at Salesforce.

*Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.* 

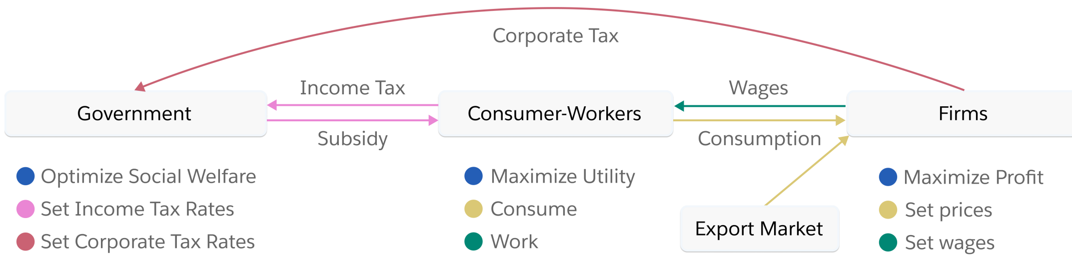


**Figure 1: RBC model with consumers, firms, and governments. Arrows represent money flow. Consumer-workers earn wages through work and consume goods from firms. They also strategically choose *which* firm to work for and *which basket of goods* to buy, but this is not explicitly visualized. Firms produce goods, pay wages, and set a price for their goods. They also invest a fixed fraction of profits to increase capital. The government taxes labor income and firm profits, and redistribute the tax revenue to the consumer-workers. Firms can also sell goods to an external export market, which acts as a price-taker that is willing to consume goods at any price.**

interact in potentially very complicated environments, while optimizing their behavior. This has been observed by economists also [13]. There is some previous work in this direction [7, 15, 34]. For example, the AI Economist used two-level RL to design optimal taxes to improve social welfare in spatiotemporal simulations, where both agents and governments use RL policies, and to find interpretable public health policies in pandemic simulations [40, 46]. For an extensive review of related work across ML, game theory, and economic modeling, see the Appendix of the full paper.

## 3 RESULTS

*Model.* We focus on a *real-business cycle* model, an instance of a DGE model. It involves 3 agent types: consumer-workers, firms that set prices and wages and use labor to produce different types of goods, and a government social planner that taxes and redistributes; all use learned RL policies (see Figure 1). Unlike approaches such as that of Hill et al. [15], we do not enforce the assumption that markets clear successfully. Firms can accumulate stocks of produced goods, and goods may be over-demanded. A more detailed and formal description of this model is in the full paper.

*Experiments.* We study variations of our RBC model with 100 consumers and 10 firms. The key empirical challenge is that joint learning using independent multi-agent RL is highly unstable in our setting. A key idea of our approach is *using structured multi-agent curricula as a meta-algorithm to stabilize joint learning*, extending the approach used by Zheng et al. [47]. These curricula consist of staged training, annealing of allowed actions, and annealing of penalty coefficients. They help to prevent the agents from learning trivial, uninteresting behavior, e.g., a situation where no production or consumption takes place.

We compare policies trained using policy gradient [43] or PPO [22, 32], with both simulation and RL agents on a GPU [23]. The learned agent behaviors are economically plausible – see Figure 2 for a characteristic run from the environment. There may be many possible DGE solutions that our approach may converge to; with neural network training, it is also hard to guarantee more than local convergence. We test the quality of our learned solutions by allowing each agent type to train separately, with other policies frozen. Using this empirical best-response analysis, we find that

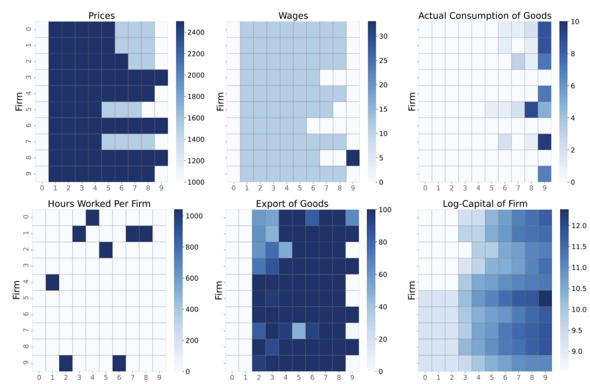


**Figure 2: Sample roll-out; open RBC model. We observe that firms have different strategies: some set prices high and rely on exporting goods (e.g. firm 3); others set prices lower and also sell to consumers (for example, firm 0). Consumers respond sensibly, only consuming when prices are low and mainly working when wages are not 0. Firms have different levels of starting capital and production technology.**

agent types are not able to improve their reward much, suggesting that at least a local, approximate equilibrium has been reached. For further experimental results and details of the training procedure and best-response analysis, see the full paper.

## 4 DISCUSSION AND FUTURE WORK

Economic models often assume the existence of a small number of representative agents whose behavior is simple and analytically tractable. Meanwhile, agent-based modeling tools do incorporate more complexity and heterogeneity, but often fail to model optimal behavior by the agents. In this work, we adapt multi-agent RL to enable economic analysis in models with more complexity and at larger scales than were previously possible. Future work might involve taking advantage of the flexibility of our framework to expand the realism of the economic models considered, further analyzing the behavior of non-equilibrium solutions, or trying to establish convergence guarantees for special cases.

## ACKNOWLEDGMENTS

This paper is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 805542).

## REFERENCES


[1] G.C. Archibald, E.S. Phelps, A.A. Alchian, and C.C. Holt. 1970. *Microeconomic Foundations of Employment and Inflation Theory.* Norton.

[2] Manoj Atolia, Santanu Chatterjee, and Stephen J Turnovsky. 2010. How misleading is linearization? Evaluating the dynamics of the neoclassical growth model. *Journal of Economic Dynamics and Control* 34, 9 (2010), 1550–1571.

[3] Yu Bai, Chi Jin, Huan Wang, and Caiming Xiong. 2021. Sample-Efficient Learning of Stackelberg Equilibria in General-Sum Games. *arXiv preprint arXiv:2102.11494* (2021).

[4] David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. 2018. The Mechanics of N-Player Differentiable Games. *arXiv preprint arXiv:1802.05642* (Feb. 2018).

[5] Eric Bonabeau. 2002. Agent-Based Modeling: Methods and Techniques for Simulating Human Systems. *Proceedings of the National Academy of Sciences* 99, suppl 3 (May 2002), 7280–7287. https://doi.org/10.1073/pnas.082080899

[6] Lena Mareen Boneva, R. Anton Braun, and Yuichiro Waki. 2016. Some unpleasant properties of loglinearized solutions when the nominal rate is zero. *Journal of Monetary Economics* 84 (2016), 216–232. https://doi.org/10.1016/j.jmoneco.2016.10.012

[7] Shu-Heng Chen, Bin-Tzong, Chie, Ying-Fang Kao, Ragupathy, and Kartik Venkatachalam. 2017. Agent-Based Modeling of a Non-Tâtonnement Process for the Scarf Economy The Role of Learning.

[8] Wilbur John Coleman. 1990. Solving the stochastic growth model by policy-function iteration. *Journal of Business & Economic Statistics* 8, 1 (1990), 27–29.

[9] Panayiotis Danassis, Aris Filos-Ratsikas, and Boi Faltings. 2021. Achieving Diverse Objectives with AI-driven Prices in Deep Reinforcement Learning Multi-agent Markets. *arXiv preprint arXiv:2106.06060* (2021).

[10] Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. 2021. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems* 34 (2021).

[11] Oliver De Groot, Ceyhun Bora Durdu, and Enrique G Mendoza. 2019. *Approximately Right?: Global v. Local Methods for Open-Economy Models with Incomplete Markets.* Technical Report. National Bureau of Economic Research.

[12] Jakob N. Foerster, Richard Y. Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. 2017. Learning With Opponent-Learning Awareness. *arXiv:1709.04326 [Cs]* (Sept. 2017). http://arxiv.org/abs/1709.04326 arXiv: 1709.04326.

[13] Andrew G Haldane and Arthur E Turrell. 2019. Drawing on different disciplines: macroeconomic agent-based models. *Journal of Evolutionary Economics* 29, 1 (2019), 39–66.

[14] Burkhard Heer and Alfred Maussner. 2009. *Dynamic general equilibrium modeling: computational methods and applications.* Springer Science & Business Media.

[15] Edward Hill, Marco Bardoscia, and Arthur Turrell. 2021. Solving Heterogeneous General Equilibrium Economic Models with Deep Reinforcement Learning. *arXiv preprint arXiv:2103.16977* (2021).

[16] Tom D Holden. 2017. Existence and uniqueness of solutions to dynamic models with occasionally binding constraints. *The Review of Economics and Statistics* (2017), 1–45.

[17] Kenneth L Judd. 1992. Projection methods for solving aggregate growth models. *Journal of Economic theory* 58, 2 (1992), 410–452.

[18] Kenneth L Judd. 1997. Computational economics and economic theory: Substitutes or complements? *Journal of Economic Dynamics and Control* 21, 6 (1997), 907–942.

[19] Nitin Kamra, Umang Gupta, Fei Fang, Yan Liu, and Milind Tambe. 2018. Policy learning for continuous space security games using neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence.*

[20] Greg Kaplan, Benjamin Moll, and Giovanni L Violante. 2018. Monetary policy according to HANK. *American Economic Review* 108, 3 (2018), 697–743.

[21] Alan P Kirman. 1992. Whom or what does the representative individual represent? *Journal of economic perspectives* 6, 2 (1992), 117–136.

[22] Ilya Kostrikov. 2018. PyTorch Implementations of Reinforcement Learning Algorithms. https://github.com/ikostrikov/pytorch-a2c-ppo-acktr-gail.

[23] Tian Lan, Sunil Srinivasa, Huan Wang, Caiming Xiong, Silvio Savarese, and Stephan Zheng. 2021. WarpDrive: Extremely Fast End-to-End Deep Multi-Agent Reinforcement Learning on a GPU. arXiv:2108.13976 [cs.LG]

[24] Robert E Lucas and Thomas Sargent. 1981. After keynesian macroeconomics. *Rational expectations and econometric practice* 1 (1981), 295–319.

[25] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory.* Oxford University Press, Oxford, New York.

[26] Enrique G Mendoza. 1991. Real business cycles in a small open economy. *The American Economic Review* (1991), 797–818.

[27] Enrique G Mendoza and Sergio Villalvazo. 2020. FiPIt: A simple, fast global method for solving models with two endogenous states & occasionally binding constraints. *Review of Economic Dynamics* 37 (2020), 81–102.

[28] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. 2007. *Algorithmic Game Theory.* Cambridge University Press. https://doi.org/10.1017/CBO9780511800481

[29] OpenAI. 2018. OpenAI Five. https://blog.openai.com/openai-five/.

[30] Jean Pierre Danthine and John B. Donaldson. 1993. Methodological and empirical issues in real business cycle theory. *European Economic Review* 37, 1 (1993), 1–35. https://doi.org/10.1016/0014-2921(93)90068-L

[31] Dylan Radovic, Lucas Kruitwagen, Christian Schroeder de Witt, Ben Caldecott, Shane Tomlinson, and Mark Workman. 2021. Revealing Robust Oil and Gas Company Macro-Strategies Using Deep Multi-Agent Reinforcement Learning. (Sept. 2021).

[32] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).

[33] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. 2017. Mastering the Game of Go Without Human Knowledge. *Nature* 550, 7676 (2017), 354.

[34] Ekaterina Sinitskaya and Leigh Tesfatsion. 2015. Macroeconomies as constructively rational games. *Journal of Economic Dynamics and Control* 61 (2015), 152–182.

[35] Frank Smets and Rafael Wouters. 2007. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American economic review* 97, 3 (2007), 586–606.

[36] Joseph E Stiglitz. 2018. Where modern macroeconomics went wrong. *Oxford Review of Economic Policy* 34, 1-2 (2018), 70–106.

[37] Nancy L. Stokey, Robert E. Lucas, and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics.* Harvard University Press. http://www.jstor.org/stable/j.ctvjnrt76

[38] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement Learning: An Introduction.* MIT Press.

[39] Kristal K Trejo, Julio B Clempner, and Alexander S Poznyak. 2016. Adapting strategies to dynamic environments in controllable stackelberg security games. In *2016 IEEE 55th Conference on Decision and Control (CDC).* IEEE, 5484–5489.

[40] Alexander Trott, Sunil Srinivasa, Douwe van der Wal, Sebastien Haneuse, and Stephan Zheng. 2021. Building a Foundation for Data-Driven, Interpretable, and Robust Policy Design using the AI Economist. *arXiv preprint arXiv:2108.02904* (2021).

[41] Oriol Vinyals, Igor Babuschkin, Junyoung Chung, Michael Mathieu, Max Jaderberg, Wojciech M. Czarnecki, Andrew Dudzik, Aja Huang, Petko Georgiev, Richard Powell, Timo Ewalds, Dan Horgan, Manuel Kroiss, Ivo Danihelka, John Agapiou, Junhyuk Oh, Valentin Dalibard, David Choi, Laurent Sifre, Yury Sulsky, Sasha Vezhnevets, James Molloy, Trevor Cai, David Budden, Tom Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Toby Pohlen, Yuhuai Wu, Dani Yogatama, Julia Cohen, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Chris Apps, Koray Kavukcuoglu, Demis Hassabis, and David Silver. 2019. AlphaStar: Mastering the Real-Time Strategy Game StarCraft II. https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/.

[42] Yufei Wang, Zheyuan Ryan Shi, Lantao Yu, Yi Wu, Rohit Singh, Lucas Joppa, and Fei Fang. 2019. Deep reinforcement learning for green security games with real-time information. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 1401–1408.

[43] Ronald J. Williams. 1992. Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Mach. Learn.* 8, 3-4 (May 1992), 229–256. https://doi.org/10.1007/BF00992696

[44] Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, and Patrick Lucey. 2018. Generative multi-agent behavioral cloning. *arXiv* (2018).

[45] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control* (2021), 321–384.

[46] Stephan Zheng, Alexander Trott, Sunil Srinivasa, Nikhil Naik, Melvin Gruesbeck, David C Parkes, and Richard Socher. 2020. The ai economist: Improving equality and productivity with ai-driven tax policies. *arXiv preprint arXiv:2004.13332* (2020).

[47] Stephan Zheng, Alexander Trott, Sunil Srinivasa, David C. Parkes, and Richard Socher. 2022. The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science Advances* 8, 18 (2022), eabk2607. https://doi.org/10.1126/sciadv.abk2607 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.abk2607

[48] Stephan Zheng, Yisong Yue, and Jennifer Hobbs. 2016. Generating long-term trajectories using deep hierarchical networks. *Advances in Neural Information Processing Systems* 29 (2016).