TaxAgent: Large Language Model-Empowered Adaptive Taxation Optimizer

Anonymous ACL submission

Abstract

Economic inequality is a severe global issue. It intensifies disparities in education and healthcare, impairing social stability. Traditional systems such as the US federal income tax reduce inequality but they lack adaptability. Frameworks like the Saez optimal taxation introduce dynamic adjustments, but they rely on rigid economic assumptions and taxpayer homogeneity. This study introduces the **TaxAgent**, an integration of large language models (LLMs) with agent-based modeling (ABM) to design adaptive tax policy that accounts for taxpayer heterogeneity and moves beyond arbitrary assumptions. In our macroeconomic simulation, heterogeneous H-Agents (households) simulate real-world taxpayers and the TaxAgent (government) iteratively optimizes tax rates. Benchmarked against Saez optimal taxation, US federal income tax, and free-market system, TaxAgent achieves superior equality-productivity trade-offs and maintains a healthy economy with low unemployment and stable inflation. Two behavioral experiments further suggest that H-Agents better simulate real human decision-making compared to rule-based models. This research provides a novel taxation solution and a scalable framework for fiscal policy evaluation, demonstrating the potential of LLMs in addressing social challenges.

1 Introduction

Economic inequality is a critical global issue with profound social, political, and economic impacts. Researches highlight its detrimental effects on education, healthcare, political stability, and economic growth (Flug et al., 1998; Ferreira et al., 2022; Glaeser et al., 2003). Tackling inequality is essential for building a fair and prosperous society.

Progressive taxation emerged to address inequality. By imposing higher tax rates on higher incomes, systems like the US federal income tax have shown potential to reduce poverty and improve health, education, and economic opportunities (Hoynes and Patel, 2015; Nichols and Rothstein, 2015). However, these systems remain static, constrained by legislative inertia and unable to adapt dynamically to shifting economic conditions.

The evolution of dynamic optimal taxation frameworks, from Mirrlees' foundational work on incentive compatibility (Mirrlees, 1971) to Saez's rule-based social welfare optimization (Saez, 2001), has sought to reconcile equality and efficiency. Yet unresolved challenges persist:

- **Arbitrary Assumptions in Economic Models:** Critical assumptions—such as the Slutsky matrix's sufficiency in capturing consumer behavior (Sørensen, 2007) and utility aggregation methods (Kleven et al., 2009)—lack empirical grounding and face challenges for overlooking behavioral complexities (Weisbach, 2023; Afriat, 1980).
- **Behavioral Homogeneity in Taxpayer Responses:** Traditional economic models rely on rigid functional forms for taxpayer behavior (Diamond and Mirrlees, 1971; Saez, 2001), neglecting heterogeneity and bounded rationality in real-world decision-making. Although the latest customized neural networks can capture various responses, they rely on a significant amount of training data and expert knowledge, making them difficult to calibrate and deploy (Tesfatsion and Judd, 2006; Trott et al., 2021; Zheng et al., 2022; Mi et al., 2023).

To address these limitations, we propose integrating agent-based modeling (ABM) with large language models (LLMs). This synergy enables adaptive tax optimization by replacing restrictive assumptions with simulations of heterogeneous human-like behaviors. LLMs enhance policymakers' ability to interpret complex socioeconomic trends while avoiding contentious welfare calculations (Li et al., 2024).

Our framework comprises three components. TaxAgent: An LLM-powered government agent that dynamically adjusts tax policies using real-time socioeconomic data, circumventing reliance on arbitrary welfare metrics. H-Agents Group: Heterogeneous households represented by LLMs, simulating diverse behaviors absent in traditional models. Macroeconomic Environment: A simulator capturing market dynamics (wages, prices, production) to evaluate policy impacts.

By iteratively optimizing tax rates through agent interactions, the framework resolves the equality-efficiency dilemma more effectively than traditional systems. Benchmarking against Saez taxation, the US federal taxation, and free-market scenario reveals the TaxAgent's superior balance of equality and productivity: 12.8% better than the second-best US federal taxation. Mechanistic analysis demonstrates that the TaxAgent achieves policy consistency and prioritizes equality without sacrificing flexibility, mitigating productivity decline through dynamic adjustments. Evaluation of economic side-effects shows that the TaxAgent does not compromise macroeconomic stability; instead, it fosters low unemployment rate and price stability. Finally, behavioral experiments substantiated our claim that H-Agents outperform rule-based ones in simulating real-world taxpayers.

This study contributes:

- **A Scalable Policy Evaluation Framework:** Combines adaptive policymakers, behavioral heterogeneity, and market dynamics in a simulation environment, enabling scalable experimentation for real-world fiscal policy design.
- **LLM-Driven Tax Innovation:** Eliminates rigid assumptions by using LLMs to model both policymaker reasoning and taxpayer behavior, enabling dynamic and data-informed taxation.
- **Comprehensive Empirical Validation:** This study highlights our framework's effectiveness in addressing taxation trade-offs, showing the LLM's potential in fiscal policy design. By benchmarking against established models and analyzing economic side-effects, we confirm that TaxAgent maintains macroeconomic stability while optimizing the equality-productivity trade-off. Additional experiments show H-Agents outperform rule-based models in replicating real taxpayer behavior.

2 Related Work

Traditional Tax Systems Progressive taxation implements higher rates for higher incomes, reducing economic inequality (Hoynes and Patel, 2015; Nichols and Rothstein, 2015) but lacks adaptability to dynamic economic conditions (Foo, 2019; Patjoshi, 2015).

Optimal taxation theory maximizes social welfare while considering economic constraints (Diamond and Saez, 2011). Pioneered by Mirrlees (Mirrlees, 1971) and Diamond and Mirrlees (Diamond and Mirrlees, 1971), this approach optimizes aggregate utility. Saez (Saez, 2001) advanced the field by incorporating earnings elasticity and income distribution into tax rate calculations. Diamond and Saez (Diamond and Saez, 2011) further developed this framework into a closed-loop system balancing social welfare and income inequality.

Recent economic research emphasizes taxpayer behavioral responses. Studies on top tax rates' elasticity (Piketty et al., 2014) reveal its impacts on labor supply and tax avoidance. Research on unemployment effects (Kroft et al., 2020) demonstrates how wage responses shape tax structures, leading to innovations like the Earned Income Tax Credit.

AI in Economic Policy Research AI optimizes economic policy beyond traditional equilibrium-based models. The AI Economist (Zheng et al., 2020) demonstrates the potential of reinforcement learning in tax policy optimization, and the integration of causal inference improves the impact assessment (Athey, 2018).

Agent-based models (ABMs) can simulate complex economic phenomena (Axtell and Farmer, 2022), enabling business cycle, policy intervention, and inflation studies (Delli Gatti et al., 2018). Improved computation power and data quality have enhanced the empirical validity of these studies (Zheng et al., 2020).

Large Language Models (LLMs) introduce advanced reasoning capabilities to economic research, enabling market behavior simulation and policy evaluation (Zhao et al., 2024; Nie et al., 2024).

Current optimal taxation models rely on arbitrary assumptions and underestimate behavioral heterogeneity (Weisbach, 2023; Zheng et al., 2020). With the advancements in ABM and LLMs, our work addresses these limitations by employing LLMs both as tax planners and taxpayers, eliminating explicit economic assumptions while capturing diverse behavioral responses (Li et al., 2024).

3 Taxation Evaluation Framework

Our framework integrates three components—the TaxAgent (government), H-Agents Group (households), and the macroeconomic simulation environment. This framework models household-government interactions within an evolving economy. The system operates as follows:

- **Household Decision-Making:** H-Agents, representing households, observe economic dynamics such as taxation and market conditions from the macroeconomic environment and incorporate past experiences. They decide on work and consumption propensities based on these inputs.
- **Macroeconomic Environment Dynamics:** Heterogeneous household decisions are processed, updating metrics such as production, wages, and prices, reflecting supply-demand dynamics and financial market influences.
- **Government Decision-Making via the TaxAgent:** The TaxAgent, representing the government, analyzes updated economic metrics and household behavior using an LLM. It proposes tax rates optimized for social goals.
- **Iterative Feedback:** New tax rates are implemented and updated metrics are fed back to H-Agents for the next round decision making. This cycle continues for a set number of iterations.

The following subsections detail the roles and mechanisms of the three components.

3.1 H-Agent: Agent as a Household

One H-Agent represents one distinct household, using an LLM to model decision-making in response to economic conditions. By interacting with the macroeconomic environment and the TaxAgent, H-Agents Group collectively influence macroeconomic outcomes of tax policies.

Each H-Agent operates through two modules. Decision-Making: H-Agents determine production and consumption propensities based on economic inputs, including household-specific background information, taxation, labor markets, price levels, and financial markets. Self-Reflection: A reflection module enhances decision-making by maintaining a memory pool of past economic data and decisions (Li et al., 2024). Quarterly, H-Agents refine future behavior by reviewing this information.

3.1.1 Decision-Making

H-Agents decide on working and consumption propensities (p_i^w, p_i^c) based on current economic observations and past reflections, as shown in Figure 1, process ①:

$$(p_i^w, p_i^c) = H_i(Pmt_i, \theta_i^R) \tag{1}$$

H_i is the decision function for the i-th household, Pmt_i includes observable economic data and background information. θ_i^R represents reflection-based parameters from previous decisions.

3.1.2 Self-Reflection

At the end of each quarter, represented as loop ② in Figure 1, H-Agent reviews its decisions and economic history to update its reflection parameters:

$$\theta_i^R \leftarrow H_i(Memo_i) \tag{2}$$

where $Memo_i$ represents the i-th household's prior prompts and decision history.

3.2 TaxAgent: Agent as a Government

The TaxAgent is the central authority of the macroeconomic simulation, leveraging an LLM to dynamically adjust tax rates, balancing two societal goals: productivity and equality.

The TaxAgent iteratively performs two steps. Tax rate adjustment: Using a heuristic prompt that combines household data, global performance metrics, and decision-making flexibility, the TaxAgent integrates traditional tax data with the LLM reasoning. It analyzes economic trends and the productivity-equality trade-off to generate tax rates aimed at optimizing societal objectives. Iterative Feedback: Generated tax rates influence household behavior and economic conditions within the simulation. Updated metrics are fed back to the TaxAgent, enabling continuous refinement of its strategy through an iterative feedback loop.

3.2.1 Tax Rate Adjustment

The tax rate adjustment process corresponds to loop ⑧ in Figure 1 and is represented as follows:

$$TX = Gov(Pmt, \theta_G, \theta_H) \tag{3}$$

where TX represents the tax rate; Gov is the tax rate determination function; Pmt includes household data, global performance metrics, and decision-making flexibility; θ_G represents the LLM's trained parameters on the government's optimal taxation strategy and θ_H represents the

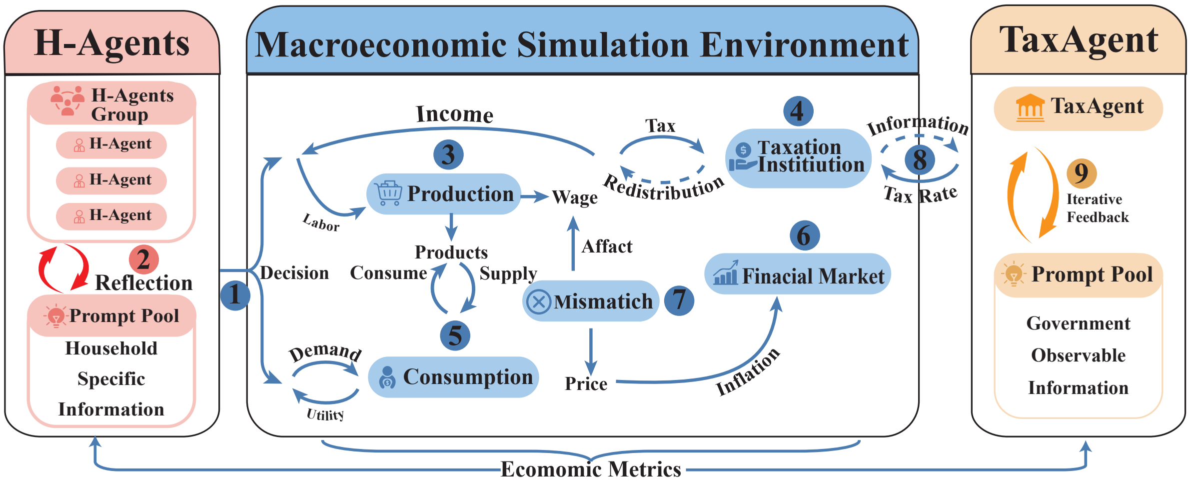

Figure 1: The illustration of the taxation evaluation framework.

trained parameters on household reactions to adjusted tax rates.

3.2.2 Iterative Feedback

Once proposed, the tax rates are implemented in the simulation environment, influencing household behavior and market dynamics. The resulting economic metrics are then fed back to the TaxAgent, forming an iterative feedback loop as shown in Figure 1, loop 9:

$$\theta_G, \theta_H \leftarrow Gov(Pmt_{upd}, \theta_G, \theta_H) \tag{4}$$

Where Pmt_{upd} includes latest household data and global performance metrics.

3.3 Macroeconomic Simulation Environment

The macroeconomic simulation environment models key aspects of a real-world economy. It includes four modules: production, taxation, consumption, and the financial market, which interact dynamically.

Economic metrics in each module are updated based on decisions of the TaxAgent and H-Agents Group. These metrics, in turn, inform agent decisions, creating a cyclical feedback loop that begins with production and wage distribution, followed by taxation, income allocation, and adjustments to wages and prices based on production-consumption relationship.

3.3.1 Production Module

Production is the starting point of economic activity, as shown in Figure 1, Module 3. The production, S, is determined by the total individual labor supply, s_j:

$$S = \sum_{j=1}^{N} s_j \tag{5}$$

The inventory G is updated after production as follows:

$$G \leftarrow G + S \tag{6}$$

Households receive wages upon completion of production.

3.3.2 Taxation Module

Taxation, central to this study, follows wage distribution. The taxation is bracketed: income in each bracket is taxed at a specific rate. The brackets are set according to the 2018 version of US federal income tax. In addition, redistribution is even and implicit, emulating real-world scenarios:

$$z_i = z_i^{pre} - T(z_i) + z^r \tag{7}$$

Where z_i is individual income, $z_i^{pre}$is pre-tax income, $T(z_i)$ is tax levied, and z^r is the redistribution.

3.3.3 Consumption Module

After taxation, households allocate post-tax income between consumption and savings. Total demand, D, is the sum of individual demands, d_j, and inventory G is updated dynamically:

$$D = \sum_{j=1}^{N} d_j \tag{8}$$

$$G \leftarrow G - d_j \tag{9}$$

3.3.4 Financial Module

The financial market incorporates the interest rate, a critical metric influencing household savings. Individual saving, s_i, increase annually by the prevailing interest rate r:

$$s_i \leftarrow s_i \times (1 + r). \tag{10}$$

Interest rate is adjusted based on the unemployment rate and inflation rate (Dawid and Gatti, 2018).

3.3.5 Global Interdependency

Interactions among production, taxation, consumption, and the financial market drive wage and price changes. When supply exceeds demand, prices drop, restraining profits and wages. Conversely, when demand exceeds supply, the opposite occurs.

3.4 Advancing Beyond Rule-Based Optimal Tax Systems

The TaxAgent addresses the limitations of traditional optimal taxation.

- Beyond Assumption Optimization: By combining data with an understanding of collective human welfare, the LLM-based TaxAgent shifts from rigid assumptions to adaptive decision-making, reflecting diverse societal perspectives as a "superposition of social consciousness."
- Modeling Heterogeneous Behavior: Learning from the heterogeneous and irrational behaviors of H-Agents in the macroeconomic environment, the TaxAgent adapts to various responses and moves beyond the rational-agent assumption, producing superior outcomes.

4 Experiments

This section assesses the ability of the TaxAgent to achieve balanced social outcomes compared to traditional tax systems, guided by the following research questions:

- **RQ1** Does the TaxAgent achieve an improvement over traditional tax systems?
- **RQ2** What is the tax rate determination mechanism of TaxAgent?
- **RQ3** What are the macroeconomic side-effects of deploying the TaxAgent?
- **RQ4** Do H-Agents simulate real-person behaviors better than rule-based agents?

4.1 Baselines

Traditional machine learning approaches' underlying assumptions often fail to reflect real-world complexities (Zhao et al., 2021; Ezeife et al., 2021).

Therefore, we propose an LLM-based approach that better captures real-world economic behavior (Li et al., 2024) and evaluate it against three representative tax systems. Saez Optimal Taxation: A theoretical benchmark based on elasticity estimates and income distribution. US Federal Income Tax: A bracketed progressive system, representing the real-world approach. Free-market: A Zero-tax scenario, a theoretical Pareto optimal, serving as a baseline to measure redistributive effects.

These baselines include theoretical optimal, real-word implementation, and theoretical Pareto optimal. Our work introduces the first explainable LLM-based tax system that combines theoretical rigor with contextual awareness.

4.2 Implementation Details

The simulation comprises $N = 200$ households over $P = 120$ months. Four tax systems are tested: free-market, US federal income taxation, Saez optimal taxation, and the TaxAgent. Productivity is fixed at 1, with results remaining consistent across parameter variations. The simulation utilizes the glm-4-plus model.

4.3 Performance Evaluation Metrics

Tax system performances are evaluated using equality and productivity per capita. Equality is calculated as the complement of the normalized Gini-index, while productivity is defined as the current average wealth of H-Agents. The social outcome is assessed as the product of equality and productivity (Zheng et al., 2020).

4.4 Experiment Results

4.4.1 The TaxAgent in Generating Social Outcomes (RQ1)

Performance Comparison Across Tax Systems The TaxAgent is evaluated against Saez optimal taxation, US federal income tax, and the free-market over 120 months by tracking the Equality-Productivity Index (EPI). Figure 2 illustrates the temporal evolution of the performance of each system:

- Short-term (0–40 months): TaxAgent performs comparably to Saez and US federal systems, with all systems showing high volatility.

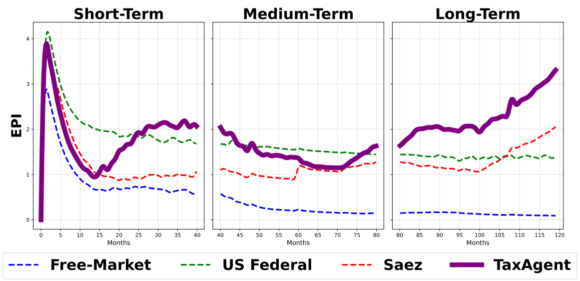

Figure 2: The EPI performance of all tax systems over 120 months. The TaxAgent (purple) performs significantly better in the long-term.

- Medium-term (40–80 months): TaxAgent experiences a slight decline before month 70, maintaining performance similar to Saez and US federal systems.
- Long-term (80–120 months): TaxAgent and Saez taxation show significant improvement, with TaxAgent consistently outperforming all other systems.

Statistical Validation of the Performance of TaxAgent Fixed-effect regression analysis quantifies performance differences: Table 1 shows each system's performance relative to the free-market baseline. **TaxAgent achieves the best overall EPI performance, outperforming the second-best US federal taxation by 12.8% overall and 70.6% in the long term.**

Table 2 analyzes temporal trends relative to the free-market. The TaxAgent shows significant positive trends in all periods except the medium-term, demonstrating superior progress. The Saez taxation shows moderate improvement; the US federal taxation and free-market systems remain stagnant.

4.4.2 Mechanisms Behind TaxAgent Performance (RQ2)

Tax Rate Determination Mechanism We analyze factors influencing the tax rate decisions of TaxAgent using time series regression:

$$T \sim \sum_{l=1}^{6} T_l + \sum_{l=1}^{6} P_l +, \sum_{l=1}^{6} E_l + \epsilon_i \qquad (11)$$

where T is the mean tax rate, E and P represent equality and productivity indices, and l indicates the lag period. Table 3 reveals two key insights:

- **Policy Consistency and Adaptability:** Strong positive effects from lag three and four tax rates, with no significant influence from recent periods, indicate balanced policy consistency and flexibility.

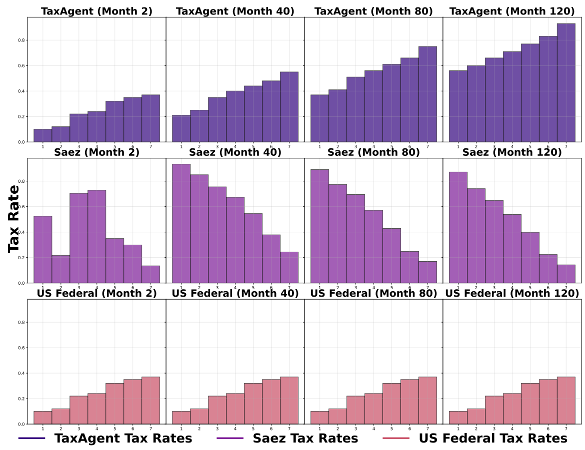

Figure 3: Sample tax rates for seven income brackets of the TaxAgent (top), the Saez taxation (middle), and US federal income tax (bottom). Regressiveness of Saez taxation and rigidness of US federal income tax limit their performances.

- **Emphasis on Equality:** The magnitude of the negative correlation between past equality levels and current tax rates is larger than the correlation between productivity and tax rates. This phenomenon demonstrates TaxAgent prioritizes addressing inequality over productivity considerations.

Comparative Analysis We also examined tax rate trajectories and their impacts on equality and productivity performance across systems, as shown in Figures 3 and 4:

- Saez Optimal Taxation: Despite theoretical optimality, its regressiveness impairs long-term performance due to the resistance to higher rates for lower-income groups.
- US Federal Income Tax: Fixed rates maintain stable equality (0.6–0.65) but lack dynamic adjustment capability, limiting optimization potential.
- Free-Market: Shows minimal productivity and equality, contradicting traditional assumptions of free-market Pareto optimality and demonstrating the necessity of taxation.

4.4.3 Macroeconomic Side-Effects of the TaxAgent (RQ3)

Inflation and unemployment rates are fundamental indicators of economic health (Masca, 2017; Cuche-Curti et al., 2008). To evaluate whether the TaxAgent achieved better social outcomes while maintaining economic stability, we analyze inflation and unemployment patterns across different

Table 1: Regression Results on the Fixed Effects of Tax Systems

Variable	Overall	Short-Term	Medium-Term	Long-Term
TaxAgent	1.41^{***}	0.89^{***}	1.19^{***}	2.15^{**}
Saez	0.82^{***}	0.42^{***}	0.82^{***}	1.24^{**}
US Federal	1.25^{***}	1.17^{***}	1.31^{***}	1.26^{**}

Note: Significance levels: $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. (The same below)

Table 2: Regression Results on the Trend of Tax Systems

Variable	Overall	Short-Term	Medium-Term	Long-Term
TaxAgent	0.02^{***}	0.03^{***}	0.00^{**}	0.04^{***}
Saez	0.01^{***}	-0.01	0.01^{***}	0.02^{***}
US Federal	0.00	0.00	0.00	0.00

The coefficients of TaxAgent in Table 1 and Table 2 demonstrate the superior EPI performance of the TaxAgent.

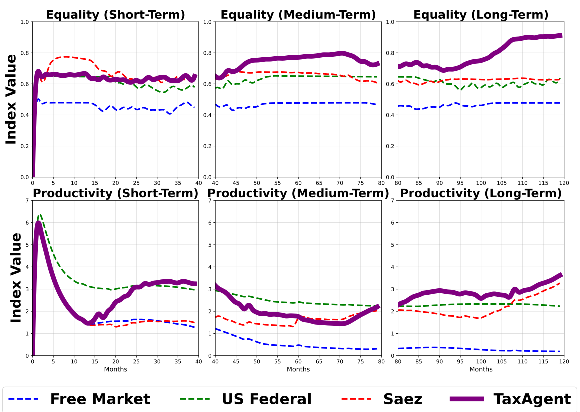

Figure 4: The equality (above) and productivity (below) performance of different tax systems. The TaxAgent demonstrates its prioritization on equality and its flexibility in making equality-productivity trade-offs.

tax systems. This analysis was crucial since high volatility in these metrics creates economic uncertainty that can discourage investment, negatively impact stock markets, and burden households (Dixit, 1992; Ampudia et al., 2020).

Figure 5 presents the inflation and unemployment rate distributions of each system. As shown in Figure 5, the TaxAgent demonstrates superior performance with three notable characteristics: first, it maintained a moderate inflation rate of approximately 8%; second, it achieved the lowest unemployment rate among all systems, consistently maintaining levels between 2% and 4%; third, both indicators show minimal fluctuations compared to other systems, suggesting a more stable economic

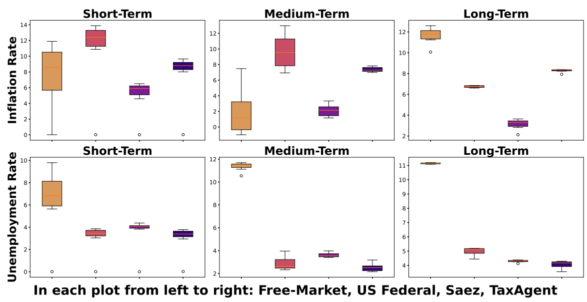

Figure 5: The distribution of inflation and unemployment rate of different tax systems. The TaxAgent (the rightmost one in each plot) achieved stable inflation control and low unemployment rate.

environment, providing predictability that allow society to adjust effectively.

4.4.4 Authentic Behaviors of the H-Agents (RQ4)

To validate the authenticity of H-Agents in modeling human behavior, we conducted two experiments, comparing H-Agents with rule-based models and responses from interviewees under identical initial conditions. We select a composite of LEN (Lengnick, 2013) and CATS (Gatti et al., 2011) as the rule-based model.

Free-Market Experiment This experiment examined decision-making in a zero-tax environment, testing the classical "rational man" assumption where individuals optimize work until marginal cost equals marginal reward (Barger, 1936). Behavioral economics suggests that perceived inequal-

Table 3: Regression Results on the Tax Rate Determination Factors

Variable	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6
Mean Tax Rate	0.00	-0.04	0.28***	0.42***	0.02	0.30*
Productivity	-0.00	-0.00	0.01	0.03**	0.02**	0.01*
Equality	-0.03	-0.02	-0.08*	-0.07**	-0.05**	-0.06***

The TaxAgent emphasizes equality and shows strong policy consistency and adaptability.

Table 4: Consumption Reduction Facing Unemployment (Percent of Overall Wealth)

Variable	Median	25th%	75th%
H-Agent	5.03	2.89	5.19
Real-person	8	2	10
Rule-based	2.25	2.14	2.36

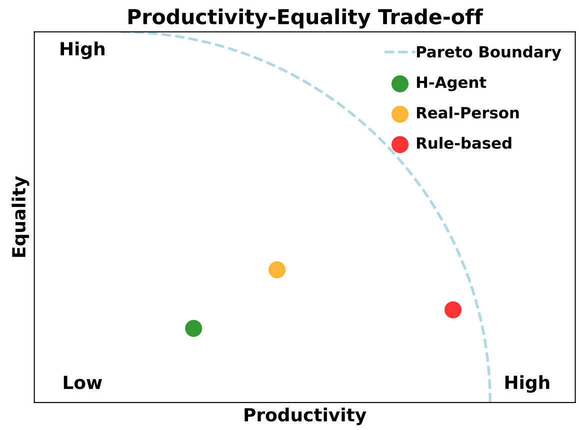

Figure 6: The equality-productivity trade-off made by H-Agents, real persons, and rule-based simulations. H-Agents(green) behaved more similar to humans(yellow) compared to rule-based simulations(red).

ity reduces work incentives (Cohen-Charash and Mueller, 2007; Sainz et al., 2023). As shown in Figure 6, H-Agents demonstrated lower equality and productivity levels that aligned more closely with real-person responses than rule-based models, confirming the impact of inequality perceptions on work behavior.

Risk Preference Analysis This experiment analyzed risk preferences by assessing consumption reductions during unemployment. Human decision-making under uncertainty involves risk aversion, probability weighting of rare events, and loss aversion (Jia et al., 2024). Table 4 shows that real individuals exhibited the highest risk aversion and H-Agents displayed intermediate levels that better approximated human behavior compared to rule-based models, further validating their authenticity in simulating real-person decisions.

5 Conclusion

This study introduces the TaxAgent, an innovative LLM-based tax planner, and evaluates its performance within an authentic simulation framework. Benchmarking against Saez optimal taxation, US federal income tax, and free-market systems demonstrates its superior performance: it achieves a 12.8% higher EPI than the second best US federal taxation. The TaxAgent emphasizes equality and shows strong policy consistency and adaptability. Analysis on the economic environment shows the TaxAgent maintained a healthy economy with low unemployment and stable inflation. Behavioral experiments validate that H-Agents models irrational and diverse taxpayer behaviors better than rule-based agents.

By eliminating reliance on rigid economic assumptions, TaxAgent pioneered in LLM-augmented fiscal policy design. This work contributes to a more equitable and prosperous society through LLM-driven tax strategies.

6 Limitations

Potential Biased Behavior of the TaxAgent Because of the training process of the LLMs, its values may not fully reflect the society. Although this study empirically validated its superior performance over traditional tax systems, adjusting training biases may generate even better social outcomes.

More Complicated Simulation Environment Although our simulation environment incorporated many important economic factors, it is not as complex as the real world. However, a more complex simulation environment is a double-edged sword: it may provide more accurate predictions or cause over-fitting and take up excessive computational power.

The Potential of Reinforcement Learning (RL) With adequate training data, RL may act as a good TaxAgent in tax policy design. However, there are two major challenges. First, a reward function is needed. Manually defining an explicit reward function means that we regress to the classical approaches that require arbitrary assumptions and value judgments. Second, training steps needed by RL will significantly increase the cost of deploying LLM-based agents to simulate taxpayers. Therefore, this may be a promising topic for future researchers with more adequate resource.

References

Sydney N. Afriat. 1980. Demand functions and the slutsky matrix.

Miguel Ampudia, Daniel Kapp, Kristian Kristiansen, and Cornelius Nicolay. 2020. Euro area equity markets and shifting expectations for an economic recovery. *Economic Bulletin Boxes*, 5.

Susan Athey. 2018. *The Impact of Machine Learning on Economics*, pages 507–547. University of Chicago Press.

R. Axtell and J. Farmer. 2022. Agent based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*.

Harold M. Barger. 1936. The general theory of employment, interest and money. *Nature*, 137:761–762.

Yochi Cohen-Charash and Jennifer S Mueller. 2007. Does perceived unfairness exacerbate or mitigate interpersonal counterproductive work behaviors related to envy? *Journal of applied psychology*, 92(3):666.

Nicolas A Cuche-Curti, Harris Dellas, and Jean-Marc Natal. 2008. Inflation targeting in a small open economy. *International finance*, 11(1):1–18.

Herbert Dawid and Domenico Delli Gatti. 2018. Agent-based macroeconomics. *Handbook of computational economics*, 4:63–156.

Domenico Delli Gatti, Giorgio Fagiolo, Mauro Gallegati, Matteo Richiardi, and Alberto Russo. 2018. *Contents*, page vii–v. Cambridge University Press.

Peter Diamond and James Mirrlees. 1971. Optimal taxation and public production: I–production efficiency. *American Economic Review*, 61(1):8–27.

Peter Diamond and Emmanuel Saez. 2011. The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives*, 25(4):165–90.

Avinash Dixit. 1992. Investment and hysteresis. *Journal of economic perspectives*, 6(1):107–132.

Enuma Ezeife, Eseoghene Kokogho, Princess Eloho Odio, and Mary Oyenike Adeyanju. 2021. The future of tax technology in the united states: A conceptual framework for ai-driven tax transformation. *International Journal of Multidisciplinary Research and Growth Evaluation*.

Ines A Ferreira, Rachel M Gisselquist, and Finn Tarp. 2022. On the impact of inequality on growth, human development, and governance. *International Studies Review*, 24(1):viab058.

Karnit Flug, Antonio Spilimbergo, and Erik Wachtenheim. 1998. Investment in education: do economic volatility and credit constraints matter? *Journal of Development Economics*, 55(2):465–481.

Check Woo Foo. 2019. Process and critical approaches to solving the systemic climate change governance problem. *Politics & Energy eJournal*.

Domenico Delli Gatti, Saul Desiderio, Edoardo Gaffeo, Pasquale Cirillo, and Mauro Gallegati. 2011. *Macroeconomics from the Bottom-up*, volume 1. Springer Science & Business Media.

Edward Glaeser, Jose Scheinkman, and Andrei Shleifer. 2003. The injustice of inequality. *Journal of Monetary Economics*, 50(1):199–222.

Hilary W Hoynes and Ankur J Patel. 2015. Effective policy for reducing inequality? the earned income tax credit and the distribution of income. Working Paper 21340, National Bureau of Economic Research.

Jingru Jia, Zehua Yuan, Junhao Pan, Paul E McNamara, and Deming Chen. 2024. Decision-making behavior evaluation framework for llms under uncertain context. *arXiv preprint arXiv:2406.05972*.

Henrik Jacobsen Kleven, Claus Thustrup Kreiner, and Emmanuel Saez. 2009. The optimal income taxation of couples. *Econometrica*, 77(2):537–560.

Kory Kroft, Kavan Kucko, Etienne Lehmann, and Johannes Schmieder. 2020. Optimal income taxation with unemployment and wage responses: A sufficient statistics approach. *American Economic Journal: Economic Policy*, 12(1):254–92.

Matthias Lengnick. 2013. Agent-based macroeconomics: A baseline model. *Journal of Economic Behavior & Organization*, 86:102–120.

Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. 2024. Econagent: Large language model-empowered agents for simulating macroeconomic activities. *Preprint*, arXiv:2310.10436.

Mahmut Masca. 2017. Economic performance evaluation of european union countries by topsis method. *North Economic Review*, 1(1):83–94.

Qirui Mi, Siyu Xia, Yan Song, Haifeng Zhang, Shenghao Zhu, and Jun Wang. 2023. Taxai: A dynamic economic simulator and benchmark for multi-agent reinforcement learning. *arXiv preprint arXiv:2309.16307*.

J. A. Mirrlees. 1971. An exploration in the theory of optimum income taxation12. *The Review of Economic Studies*, 38(2):175–208.

Austin Nichols and Jesse Rothstein. 2015. The earned income tax credit (eitc). Working Paper 21211, National Bureau of Economic Research.

Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, and Stefan Zohren. 2024. A survey of large language models for financial applications: Progress, prospects and challenges. *Preprint*, arXiv:2406.11903.

Rajesh Kumar Patjoshi. 2015. Design and development of advanced control strategies for power quality enhancement at distribution level.

Thomas Piketty, Emmanuel Saez, and Stefanie Stantcheva. 2014. Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, 6(1):230–71.

Emmanuel Saez. 2001. Using elasticities to derive optimal income tax rates. *The Review of Economic Studies*, 68(1):205–229.

Mario Sainz, Eva Moreno-Bella, and Laura C Torres-Vega. 2023. Perceived unequal and unfair workplaces trigger lower job satisfaction and lower workers' dignity via organizational dehumanization and workers' self-objectification. *European Journal of Social Psychology*, 53(5):921–938.

Peter Birch Sørensen. 2007. The theory of optimal taxation: what is the policy relevance? *International Tax and Public Finance*, 14:383–406.

Leigh Tesfatsion and Kenneth L Judd. 2006. *Handbook of computational economics: agent-based computational economics*. Elsevier.

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*.

David A. Weisbach. 2023. Constrained income redistribution and inequality: Legal rules compared to taxes and transfers. *SSRN Electronic Journal*.

Eric Zhao, Alexander R Trott, Caiming Xiong, and Stephan Zheng. 2021. Ermas: Becoming robust to reward function sim-to-real gaps in multi-agent simulations. *CoRR*.

Qinlin Zhao, Jindong Wang, Yixuan Zhang, Yiqiao Jin, Kaijie Zhu, Hao Chen, and Xing Xie. 2024. Competeai: Understanding the competition dynamics in large language model-based agents. *Preprint*, arXiv:2310.17512.

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. *Preprint*, arXiv:2004.13332.

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):eabk2607.

A Appendix

A.1 H-Agent and TaxAgent

H-Agent Prompt Example The following prompt is one example of the information provided to H-Agents in the decision-making process.

> You're Adam Mills, a 58-year-old individual living in San Antonio, Texas. A tax planner adjusts your tax rates periodically. Now it's 2001.03. Last month, you worked as a(an) Newspaper Delivery. If you continue working this month, your expected income will be $567.18, which decreased compared to last month due to deflation of the labor market. Besides, your consumption was $544.68. Part of your income last month was witheld as income tax. Last month, the tax brackets are: [0.00, 808.33, 3289.58, 7016.67, 13393.75, 17008.33, 42525.00] and their corresponding rates are: [0.10, 0.12, 0.22, 0.24, 0.32, 0.35, 0.37]. Income earned within each bracket is taxed only at that bracket's rate. This month, according to the tax planner, the brackets are not changed. But the planner updated corresponding rates: [10.00%, 12.00%, 22.00%, 24.00%, 32.00%, 35.00%, 37.00%]. Income earned within each bracket is taxed at that bracket's rate. Pay attention to the tax rates because they may be different from the previous ones and you need to make your decision based on the current rates Deflation has led to a price decrease in the consumption market, with the average price of essential goods now at $126.78. Your current savings account balance is $13072.25. Interest rates, as set by your bank, stand at 3.00%. Considering aspects like your living costs, future aspirations, broader economic trends, and the tax you need to pay, how is your willingness to work this month? How would you plan your expenditures on essential goods? Provide your decisions in a JSON format. The format should have two keys: 'work' (a value between 0 and 1 with intervals of 0.02, indicating the willingness or propensity to work) and 'consumption' (a value between 0 and 1 with intervals of 0.02, indicating the proportion of all your savings and income you intend to spend on essential goods).

The following prompt is one example of the prompt provided to H-Agents in the self-reflection process. For brevity, the previous prompts and decisions of the H-Agents is omitted.

> Given the previous quarter's economic environment, reflect on the labor, consumption, and financial markets, as well as their dynamics. What conclusions have you drawn? Your answer must be less than 200 words!

TaxAgent Prompt Example The TaxAgent is provided with comprehensive household and macroeconomic data for tax rate determination. In addition, the social performance of the TaxAgent past decisions are also provided, providing the information needed for iterative feedback. For brevity, part of the income and wealth information is omitted.

> You are a tax planner in charge of adjusting the tax rates of each income brackets. You will decide the tax rate in next period applied cumulatively to the income of agents in the seven [0.00, 808.33, 3289.58, 7016.67, 13393.75, 17008.33, 42525.00] income brackets. Last month, the incomes and wealth of individuals living in your society were $[529.42, 820.63, 1255.18, ..., 0.0, 80016.97, 0.0] and $[29273.73, 35603.8, 36976.86, ..., 180063.13, 286498.22, 294174.67]. The tax rates you set in the past months were [([[0.1, 0.12, 0.22, 0.24, 0.32, 0.35, 0.37], [0.1, 0.12, 0.22, 0.24, 0.32, 0.35, 0.37], [0.1, 0.12, 0.22, 0.24, 0.32, 0.35, 0.37], [0.1, 0.15, 0.25, 0.3, 0.35, 0.4, 0.45], [0.1, 0.15, 0.25, 0.3, 0.35, 0.4, 0.45], [0.1, 0.15, 0.25, 0.3, 0.35, 0.4, 0.45]])]. The average per-capita productivity in the last months were [(0.0), (11.26), (14.32), (15.13), (15.28), (14.54)]: the past months' equality performances were [(0.0), (0.66), (0.66), (0.66), (0.67), (0.67)](the higher, the more equal). Adjust the tax rates to build a society that you consider best for society. You have the total freedom to adjust the rates! Provide your decision in a JSON format. The decision should be a list with seven values (each value between 0 and 1 with intervals of 0.01).

A.2 Macroeconomic Simulation Framework Details

The production is determined by the total labor supplied by households. For simplicity, we assume the production of a single homogeneous commodity, with each household contributing 168 hours (21 eight-hour working days) of labor if employed in a given period. The total production, S, is defined as:

$$S = \sum_{j=1}^{N} l_j \times 168 \times A \tag{12}$$

Where l_j represents the labor supplied by household j, and A denotes productivity.

Taxation is modeled using a progressive, bracketed structure. The tax bracket limitation is represented by b_k The tax levied on a household with income z_i is given by:

$$\begin{aligned} T(z_i) = \sum_{k=1}^{B} \tau_k \left((b_{k+1} - b_k) \mathbf{1} \left[z_i > b_{k+1} \right] \right. \\ \left. + (z_i - b_k) \mathbf{1} \left[b_k < z_i \le b_{k+1} \right] \right), \end{aligned} \tag{13}$$

Redistribution in the simulation is even and latent. The actual post-tax income for a household is:

$$z_i = z_i^{pre} - T(z_i) + z^r = z_i - T(z_i) + \frac{1}{N} \sum_{j=1}^{N} T(z_j), \tag{14}$$

Demand for commodities is inversely proportional to price and directly proportional to wealth. Total societal demand is expressed as:

$$D = \sum_{j=1}^{N} d_j = \sum_{j=1}^{N} \frac{c_j}{P} = \sum_{j=1}^{N} \frac{p_j^c s_j}{P}, \tag{15}$$

where c_j stands for individual consumption intention; p_j^c is the working propensity and s_j is the accumulated wealth.

Due to inventory constraints, actual consumption, $\hat{d}_j$, is bounded by available supply:

$$\hat{d}_j = \min(d_j, G), \hat{c}_j = \hat{d}_j \times P \tag{16}$$

To ensure fairness, households consume in a randomized sequence, with the inventory updated after each transaction:

$$G \leftarrow G - \hat{d}_j. \tag{17}$$

Interest rate is defined by the Taylor Rule:

$$r = \max(r^n + \pi^t + \alpha^\pi (\pi - \pi^t) + \alpha^u (u^n - u), 0), \tag{18}$$

where, r^n represents the natural interest rate, π^t is current inflation, and u^n is the natural unemployment rate.

Demand-supply mismatch is quantified as:

$$\bar{\varphi} = \frac{D - G}{\max(D, G)}, \tag{19}$$

This imbalance triggers price and wage adjustments modeled as:

$$w_i \leftarrow w_i(1 + \varphi_i), \varphi_i \sim sign(\bar{\varphi}) U(0, \alpha_w |\bar{\varphi}|), \tag{20}$$

$$P \leftarrow P(1 + \varphi_P), \varphi_P \sim sign(\bar{\varphi}) U(0, \alpha_P |\bar{\varphi}|) \tag{21}$$

where, α_p and α_w represents the maximum adjusting rates of prices and wages, respectively.

Inflation is defined as:

$$\pi = \frac{\overline{p}_n - \overline{p}_{n-1}}{\overline{P}_{n-1}} \tag{22}$$

Unemployment is defined as:

$$u = \frac{\sum_{m=1}^{12} \sum_{j=1}^{N} (1 - l_j)}{12N} \tag{23}$$

Equality is defined as:

$$\text{eq}(x_c) = (1 - \text{gini}(x_c)) \times \frac{N-1}{N} \tag{24}$$

where $\text{gini}(x_c)$ is the standard Gini Index of the wealth of H-Agents.

Productivity is defined as:

$$\text{prod}(x_c) = \sum_{i=1}^{N} x_c^i \tag{25}$$

A.3 Baselines

US Federal Income Tax The United States federal income tax system operates under a progressive tax structure. Individuals and households are taxed at increasing rates as their taxable income rises. After ajusting for monthly income, the tax brackets and corresponding rates for the 2018 tax year are as follows:

- 10% Tax Rate: Applied to taxable income up to \$808.33 for single filers.
- 12% Tax Rate: Applied to taxable income from \$808.33 to \$3289.58 for single filers.
- 22% Tax Rate: Applied to taxable income from \$3289.58 to \$7016.67 for single filers.
- 24% Tax Rate: Applied to taxable income from \$7016.67 to \$13393.75 for single filers.

- 32% Tax Rate: Applied to taxable income from \$13393.75 to \$17008.33 for single filers.
- 35% Tax Rate: Applied to taxable income from \$17008.33 to \$42525.00 for single filers.
- 37% Tax Rate: Applied to taxable income above \$42525.00 for single filers.

The Saez Optimal Taxation The Saez tax framework is formalized as hereunder.

The utility of an individual depends positively on consumption c and negatively on labor effort z, and is given by:

$$u(c, z) = v(c) - h(z), \tag{26}$$

where $v(c)$ captures the utility from consumption, and $h(z)$ represents the disutility from labor effort. Individuals face a budget constraint:

$$c = z(1 - \tau) + R, \tag{27}$$

where z is earnings, τ is the marginal tax rate, and R is virtual income.

Behavioral responses to taxation are captured through three key elasticities:

- Uncompensated Elasticity (ϵ_u):

$$\epsilon_u = \frac{1 - \tau}{z} \cdot \frac{\partial z}{\partial(1 - \tau)}, \tag{28}$$

  which measures the sensitivity of earnings to changes in the net-of-tax rate $(1 - \tau)$;

- Income Effect (η):

$$\eta = \frac{1 - \tau}{z} \cdot \frac{\partial z}{\partial R}, \tag{29}$$

  which represents how changes in virtual income influence labor supply;

- Compensated Elasticity (ϵ_c):

$$\epsilon_c = \epsilon_u + \eta, \tag{30}$$

  which captures the pure substitution effect after accounting for income effects.

The government's objective is to maximize social welfare:

$$W = \int_z w(z) u(c, z)\, dz, \tag{31}$$

where $w(z)$ are welfare weights, decreasing with income to reflect redistributive goals.

For high-income earners, Saez derives a simple formula for the optimal marginal tax rate:

$$\tau^* = \frac{1}{1 + a \cdot \epsilon_u}, \tag{32}$$

where $a = \frac{\bar{z}}{\bar{z} - z^*}$ is the Pareto parameter, reflecting the thickness of the income distribution's top tail. This formula balances revenue gains from increased tax rates with losses due to reduced labor supply, ensuring progressivity without excessive distortion.

Extending to the full income distribution, Saez provides a general nonlinear tax schedule:

$$T'(z) = \frac{(1 - G(z)) + e \cdot z \cdot g(z)}{1 + e \cdot g(z)}, \tag{33}$$

where $G(z)$ is the cumulative income distribution, $g(z)$ is the income density, and e is the elasticity of taxable income.

Saez's framework emphasizes progressive taxation with higher marginal rates for top earners, justified by diminishing marginal utility of income and empirical evidence on elasticities.

A.4 Experiments

Ablation Study The results of using qwen-max-2024-09-19 and gpt-4o-2024-08-06 as the TaxAgent are shown in Figure 7. In general, the social outcome generated by the TaxAgent is superior in the long term. An exception is that its performance experiences a slight drop after the 100th month when the base LLM is Chatgpt, but the performance is not significantly lower than its competitors'. This indicates that the TaxAgent has low sensitivity to changes in its LLM base, enhancing its reliability.

Elaboration on Experiments in RQ4 In the Free-Market experiment we randomly selected 50 responses from the H-Agent under the free-market scenario from our main experiment and documented their economic condition and decisions on work and consumption. We provided the same prompts to real persons and asked them to make the same decisions as H-Agents. Finally, we used the formulas (Lengnick, 2013; Gatti et al., 2011) of the rule-based agents to calculate the decisions of rule-based agents.

We use the decisions made and initial conditions specified in the prompts to calculate the equality and productivity of the three kind of agents.

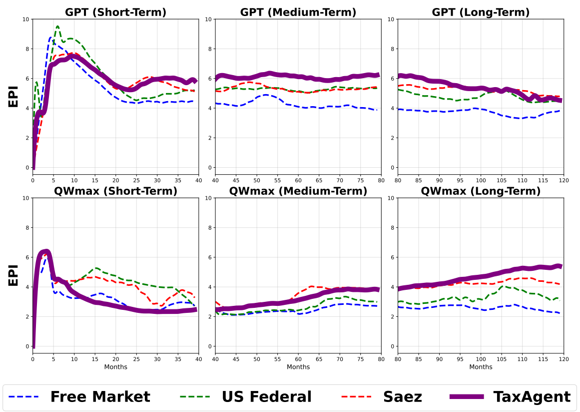

Figure 7: Ablation study of the robustness of the TaxAgent. The TaxAgent shows low sensitivity to changes in its base LLM.

In the risk preference experiment, we selected 63 unemployment incidences from our main experiment under the US federal taxation scenario and documented their economic condition and decisions on work and consumption. We provided the same prompts to real persons and asked them to decide on the consumption reduction percentage of their total wealth. Finally, based on the income loss, we used the formulas of the rule-based agents to calculate their decisions.

The participants of the two experiments were student volunteers from Huazhong University of Science and Technology. The data collected are totally anonymous. The usage of the data is fully explained and gained consensus from the participants.

It is important to note that the response of rule-based simulations to unemployment is derived from their reaction to income reduction, meaning it can be adjusted to any desired value by modifying model parameters. Nevertheless, the variance in consumption reduction among rule-based simulations remains significantly lower than that observed in the other two groups.