Designing Progressive Tax Schedules Under Relocation Risk: A Reinforcement Learning Approach for the UK

Abstract

Recent reports have found that the UK may be experiencing a wealth exodus as a result of tax hikes imposed by the current administration. On one hand, it is clear that higher tax rates are required to fund public services and address long-standing fiscal challenges; on the other, critics have argued that they undermine the intended economic benefits by prompting relocation among high-income workers. When evaluating policy impacts in this context, traditional economic models cannot capture the dynamic strategies of heterogeneous workers. The following work proposes a multi-agent reinforcement learning framework that can be leveraged to derive income tax brackets that optimize government revenue in an economy where wealthy individuals are able to relocate. Specifically, I include the following entities in this model: (a) agents with heterogenous preferences calibrated to real-world data, (b) a central planner who 'learns' optimal tax brackets in response to worker's behaviour, and (c) a representative firm whose actions are held constant. After deriving the relevant utility functions and parameters, I run a rudimentary version of the simulation with 100 workers to demonstrate that the proposed model converges at an optimal equilibrium where government revenue is maximized. The results offer promising insights into how new computational methods can complement traditional economic models.

Candidate Number: 1093276

Word Count: 2995

Contents

1 Introduction **2**

2 Literature review **2**

3 Methodology **3**

3.1 Framework 3
3.2 Worker Behaviour 4
3.3 Firm behaviour 5
3.4 Government behaviour 5
3.5 Defining parameters 6

4 Preliminary results and discussion **7**

5 Limitations and areas for future research **9**

6 Conclusion **10**

1 Introduction

In October 2024 Rachel Reeves unveiled a £40 Billion increase in the tax budget. This included maintaining a freeze on income tax thresholds introduced by the previous government and pushing the capital gains tax to 24% for high earners (Partington and Elgot, 2024). Her administration claimed that these are necessary measures to fund essential public services and address long-standing fiscal challenges but critics tend to cite a recent survey on capital flight, which found that 8% of high net worth individuals plan on leaving the UK within the next 12 months (Khan, 2024). Extensive literature has been dedicated to exploring tax systems that accommodate to this trade-off between raising taxes to fund public services that promote a more equitable society where individuals can live up to their potential, and lowering taxes to avoid discouraging investment, entrepreneurship, or talent retention (see Akcigit et al., 2016; Bhagwati and Srinivasan, 1983). But most existing studies either use rigid macroeconomic models or rely on historical data and therefore face the Lucas critique (i.e., that historical data cannot capture future behavioural responses) (Favero and Hendry, 1992).

In this context, multi-agent reinforcement learning can complement traditional methods by simulating interactive relationships between a government, corporations, and heterogenous agents. But despite its advantages, there is currently a shortage of studies that leverage the tool for tax policy evaluation. The following work proposes a multi-agent reinforcement learning algorithm that captures how workers, firms and a single government adaptively adjust their behaviour. Specifically, I build on previous work by Luo et al. (2025), by adjusting their model to (a) allow for relocation if the expected utility of doing so exceeds domestic utility, (b) hold the firm's behaviour constant and (c) calibrate relevant parameters to the UK's economy.

Within this proposed framework the central planner (i.e., government) is able to 'learn' tax brackets that maximise revenue. Upon deriving the model, I run a preliminary simulation with 100 agents to demonstrate how reinforcement learning can derive an optimal tax policy based on the expected behaviour of heterogenous workers in an open economy over a 50 year period. The derived tax schedule corresponds to a progressive system, where marginal tax rates increase in discrete steps as income rises. The optimal policy begins with a 5% marginal rate on low incomes and gradually climbs to a 25% marginal rate for the highest earners. The proposed study therefore provides both empirical insights into tax policy in the UK and a framework that policymakers can adapt to derive optimal tax schedule in complex economic environments.

2 Literature review

Over the past five decades, the study of optimal taxation has evolved significantly to account for the increasing globalization of labour markets. The groundwork was laid by Mirrlees (1971), who developed a model of optimal non-linear income taxation based on individual heterogeneity in productivity. Further as globalization intensified in the 1990s and 2000s, the literature also began to incorporate open-economy constraints. Scholars introduced capital mobility as a key constraint on redistribution to account for the risk of capital flight and the erosion of national tax bases due to tax competition. This is evident in the work of Lehmann et al. (2014), who modelled optimal nonlinear income tax schedules in the presence of competing governments and mobile top earners. Empirical research has also played a crucial role in validating these theoretical models. For example, Akcigit and Stantcheva (2021) examine the international mobility of inventors, finding that innovation-related income is highly sensitive to tax differentials, especially among top earners

This research underscore the importance of accounting for cross-border mobility when designing tax policy, especially in the knowledge economy. But traditional economic models are rigid in their underlying assumptions. As noted by Stiglitz (2019), they assume that all agents in an economy are equipped with decreasing marginal utilities and rational expectations in the face of perfectly clearing markets. Agent-based modelling has evolved as an alternative, offering a flexible framework that can accommodate heterogeneity, bounded rationality, and dynamic interaction between agents (Axtell and Farmer, 2025). This flexibility makes it particularly well-suited for simulating tax systems under social mobility.

But despite its evident advantages there is currently a shortage of large-scale multi-agent reinforcement learning applications in the field of tax policy. To my knowledge, there are only three studies that have leveraged the tool to optimize tax brackets (see Zheng et al., 2020; Mi et al., 2023; Luo et al., 2025). All of these use an existing economic model – either the Real Business Cycle, the Bewley-Aiyagari model or a dynamic stochastic general equilibrium model – to simulate how firms, workers and a central planner learn to behave optimally by adapting their behaviour to one another over several times steps. These works offer promising insights into how reinforcement learning can complement traditional economic methods, but they also fall short in that they (a) model a closed economy where agents merely optimize the trade-off between leisure and working hours, (b) they are not calibrated with reference to real-world data. I propose an adapted model that address these shortcomings to demonstrate how multi-agent reinforcement learning can be used to capture the complex relationship between micro-level decisions and macro-level phenomena.

3 Methodology

3.1 Framework

Following Luo et al. (2025) I propose using a Real-Business Cycle as a foundation for the agent-based model. For our purposes The Real-Business Cycle consists of three different entities: (a) workers equipped with heterogenous preferences, (b) a representative firm and (c) a single government who can adjust income tax brackets on a yearly basis. It should be noted that Luo et al. (2025) also account for the government's ability to adjust corporate tax brackets and firms' subsequent response in this context. Since our model merely focuses on income tax we hold the corporate tax rate constant at a baseline rate of 0%. Further, they include a the government that redistributes tax revenue evenly across all workers. This redistributive subsidy increases effective income, especially for lower earners. The approach is generally common, but weighted redistribution or a simulation of social services are arguably more representative (Suslov et al., 2016). To strike a middle ground between computational simplicity and real-world representation, I propose that a share is evenly redistributed to the population. The remainder is used on government administration, foreign aid, military or emergency reserves and debt servicing (i.e., it trickles out of the system). The representative firm in this simulated economy is recruiting workers and selling a product. Workers in turn are contributing their labour to production and subsequently spending their 'reward' on consumption.

3.2 Worker Behaviour

The model includes a population of N workers characterized by varying wages, skills and leisure preferences (the empirical calibration of these parameters will be discussed in section 3.5). For simplicity, the model excludes borrowing/saving decisions and it assumes that workers consume all post-tax income within each period. Each worker j is assigned an individual ability parameter e_j, calibrated to match the empirical distribution of earnings heterogeneity observed in the UK. All workers then choose an optimal labour supply $h_{j,t}$ in each period. Wages are determined endogenously based on aggregate labour supply and productivity. Thus, individual income can be denoted as:

$$z_{j,t} = e_j \cdot w_t \cdot h_{j,t}. \tag{1}$$

This formulation introduces heterogeneity in earnings based on skill; variations in this parameter across workers lead to heterogeneous incomes, allowing the model to capture differential exposure to progressive taxation and tax-induced relocation incentives. To replicate the empirical distribution of income in the UK, the values of e_j are drawn from a log-normal distribution calibrated to match the observed income distribution. Next, following Fabretti and Herzel (2017) I use the following Constant Relative Risk Aversion (CRRA) function to model how workers derive utility from consumption and disutility from labour:

$$U_{j,t} = \frac{C_{j,t}^{1-\eta} - 1}{1-\eta} - \phi_j \cdot h_{j,t}, \tag{2}$$

where $\eta > 0$ is the coefficient of relative risk aversion, and ϕ_j represents the individual's preference for leisure. Workers optimize $U_{j,t}$ by choosing the number of hours worked ($h_{j,t}$) in each period. Next we introduce the redistributed lump-sum denoted as:

$$x_{j,t} = \frac{\lambda T_t}{N}. \tag{3}$$

where $\lambda \in (0, 1)$ is the fixed redistribution parameter to capture government spending not returned to households, whilst T_t denotes the total tax revenue collected in period t. This value feeds into post-tax consumption $C_{j,t}$, given by:

$$C_{j,t} = (1 - \tau(z_{j,t})) \cdot z_{j,t} + x_{j,t}. \tag{4}$$

Workers with income above a threshold z_{reloc} are considered mobile and eligible to relocate abroad. Each eligible worker j assesses whether relocating would yield a higher net payoff than staying, accounting for a relocation cost of κ_j. Relocation therefore occurs if:

$$(1 - \tau(z_{j,t})) \cdot z_{j,t} + x_{j,t} < z^*_{\text{foreign}} - \kappa_j. \tag{5}$$

where $x_{j,t}$ is the lump-sum redistribution received, z^*_{foreign} is the expected income abroad (a fixed or stochastic value) and κ_j is the cost of relocating. Relocated agents are removed from the domestic tax base and no longer contribute to or benefit from redistribution.

3.3 Firm behaviour

Firms are modelled implicitly through exogenous wage assignments and labour demand; they do not actively participate in decision-making. To this end, the model uses a single representative firm that passively transforms aggregate labour into output. This allows for a focus on worker behaviour, taxation, and relocation dynamics while preserving consistency between labour supply and wages. The firm operates under a fixed Cobb-Douglas production function, where total output is determined by the aggregate hours worked:

$$Y_t = A \cdot L_t^{\psi} \tag{6}$$

where Y_t denotes total output in period t, $L_t = \sum_{j=1}^{N} h_{j,t}$ is the aggregate labour input across all workers, and $\psi \in (0, 1)$ is the labour share parameter. The constant A captures total factor productivity and the influence of a fixed capital stock. Wages in each period are derived from the marginal product of labour:

$$w_t = \frac{\partial Y_t}{\partial L_t} = A \cdot \psi \cdot L_t^{\psi-1}. \tag{7}$$

All workers are paid the same wage w_t based on aggregate labour supply. Note that although all workers have the same market wage, heterogeneity in income arises endogenously through differences in individual ability e_j and chosen labour supply $h_{j,t}$, as discussed previously.

3.4 Government behaviour

The government is simulated as a central planner that sets progressive income tax brackets and redistributes a share of collected revenue back to residents as uniform lump-sum transfers. The planner updates its tax policy using a Deep Q-Network (DQN) to optimize its long-term economic objectives. It observes a state vector at each period t:

$$s_t = [\mu_z, \sigma_z, N_{\text{relocated}}, R_t] \tag{8}$$

where μ_z and σ_z denote the mean and standard deviation of worker income, respectively, $N_{\text{relocated}}$ represents the number of workers who relocated in the last period and R_t denotes total tax revenue collected. Based on s_t, the government selects a tax action $a_t = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5\}$, which represents a progressive income tax policy defined by a five-tier bracket system:

$$\tau(z) = \begin{cases} \tau_1 & \text{if } z < b_1 \\ \tau_2 & \text{if } b_1 \leq z < b_2 \\ \tau_3 & \text{if } b_2 \leq z < b_3 \\ \tau_4 & \text{if } b_3 \leq z < b_4 \\ \tau_5 & \text{if } z \geq b_4 \end{cases}, \quad R_t = \sum_{j=1}^{N} \tau(z_{j,t}) \cdot z_{j,t} \tag{9}$$

where $a_t = \{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5\}$ are marginal tax rates, and $\{b_1...b_4\}$ are fixed bracket thresholds. This structure reflects the UK's actual tax system and allows for varying progressivity across income levels. R_t then denotes total tax revenue. The tax action a_t is defined by the Deep Q-Network, which approximates the optimal policy to maximize long-term rewards. The government's reward function is then modelled as a weighted combination of two objectives:

$$r_t = \alpha R_t - \beta N_{\text{relocated}} \tag{10}$$

where α encourages revenue generation and β penalizes relocation (tax base erosion). Although relocations already reduce total tax revenue endogenously, this additional term provides a clearer learning signal and helps the agent internalize the behavioural consequences of high marginal tax rates more directly. This design choice improves training efficiency and allows for flexible calibration of the government's sensitivity to talent flight. The Deep Q-Network updates the action-value function via temporal difference learning:

$$Q(s, a) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right] \tag{11}$$

where $Q(s, a)$ denotes the value of taking action a in state s, $\mathbb{E}[\cdot]$ is expected future reward and r_t constitutes the reward received at time t. Finally, the term γ^t serves as a discount factor that gives more weight to near-term rewards than distant ones. This approach to reinforcement learning is standard practice in the literature (Knox and Stone, 2015), given that there is greater certainty in the expected short-term reward than in the long-term. The equation essentially holds that the value of a tax policy a in state s is the expected sum of discounted future rewards the government will receive if it starts with this action and acts optimally thereafter. The adaptability of this approach separates it from traditional economic models, which typically rely on optimizing static or myopic objectives.

3.5 Defining parameters

In the above described framework workers choose the number of hours they work by maximising the utility function in Equation 2 and they choose to relocate if expected income abroad minus the relocation cost is greater than their domestic income plus the lump sum that they receive as a result of the redistributed tax. Simultaneously the government learns a tax schedule with reference to a Deep Q-Network that maximizes their expected revenue over the 50-year period. Finally, although tax revenue is the what we optimize for, Gini coefficients and GDP are also recorded to help us better understand the simulated economy.

All relevant parameters are listed in Table 1. These essentially capture the economic framework within which the simulation will be ran. Parameters such as worker efficiency (e_j), labour disutility (ϕ_j), initial income ($w_{j,0}$) and relocation cost (κ_j) are drawn from distributions that should be defined with reference to empirical data on socio-economic indicators for the UK. Specifically, I propose using data from the UK Household Longitudinal Study (UKHLS) to calibrate the parameters e_j (via hourly income conditional on hours worked), ϕ_j (via observed labour-leisure trade-offs), and κ_j (via regional mobility patterns and migration responses to income differentials). Additionally, the Family Resources Survey (FRS) can be used to define the initial income distribution $w_{j,0}$ based on household earnings percentiles. These datasets provide a rich empirical basis for generating plausible agent heterogeneity in line with real-world distributions.

Parameter	Role	Value / Range	Source
N	Agents in economy	34 million	Calibrated
e_j	Worker efficiency	$\sim$ Lognormal(μ_e, σ_e)	Calibrated
ϕ_j	Labor disutility	$\in [0.1,\ 0.5]$	Calibrated
κ_j	Relocation cost	$\in [0.7,\ 1.2]$	Calibrated
w_t	Wage	$\in [0,\ 1]$	Calibrated
z^*_{foreign}	Income abroad	1.0	Calibrated
λ	Redistribution efficiency	0.85	Theory
η	Risk aversion (CRRA)	2.0	Theory
ψ	Labour share in production	0.7	Calibrated
A	Productivity scale	1.0	Calibrated
b	Tax bracket thresholds	$\{0.2, 0.4, 0.6, 0.8\}$	Design
τ_{vals}	Marginal tax choices	$\{0.05, \ldots, 0.6\}$	Design
γ	Discount factor (DQN)	0.95	Design
α, β	Reward weights	1.0	Design

Table 1: Key parameters used in the agent-based policy framework.

Table 1 includes all parameters relevant to the simulation. Some of these parameters constitute a distribution (e.g., worker efficiency and labour disutility), which will equip agents with heterogenous preferences. Other parameters such as income abroad and relocation cost are the same for all agents. These are currently attributed with placeholder values that were used for the preliminary simulation. Tax policy, by contrast, is not fixed but learned dynamically by the model. A five-tier progressive tax system is implemented, where marginal tax rates $\{\tau_1, \ldots, \tau_5\}$ are selected from a discrete set of values ranging from 0.05 to 0.6. This range can be seen as an *action space* (i.e., a list of options from which the model can choose). The rates are updated through reinforcement learning via the DQN, which selects from $6^5 = 7{,}776$ possible tax rate combinations. Bracket thresholds are fixed ex ante, while the government agent learns the optimal set of marginal rates that maximize long-run revenue net of relocation costs.

4 Preliminary results and discussion

A simple rudimentary simulation was performed to test the 'feasibility' of the experiment. For computational efficiency the simulation was conducted with merely 100 agents and all parameters outlined in Table 1 that should be empirically calibrated with reference to survey data are randomly drawn from a continuous uniform distribution within the prescribed range. Agents are therefore heterogenous in their preferences and their behaviour, but they are not calibrated to the UK economy[1]. Figure 1 represents how the model converged to an optimal tax return.

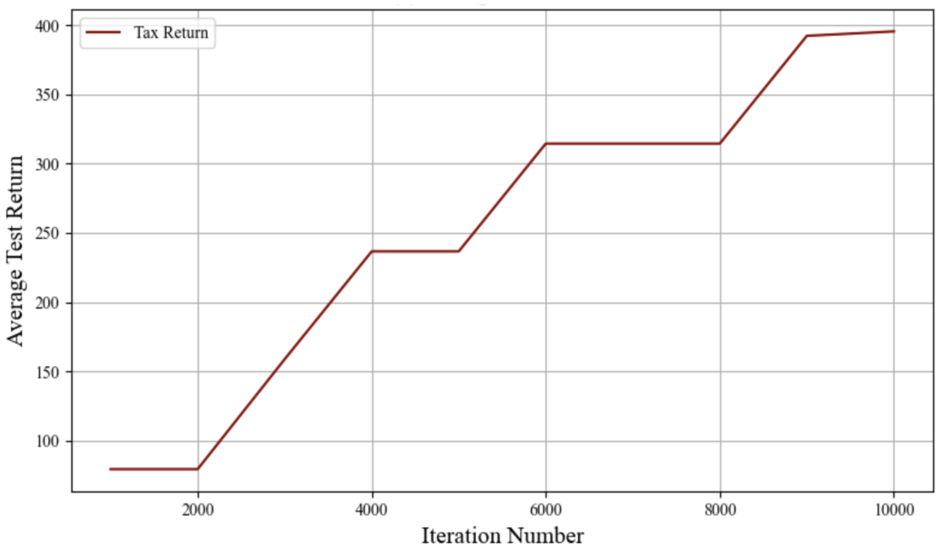

Figure 1: Convergence to optimal revenue

The plateaus reflect periods where the agent converges to locally optimal tax policies, while the subsequent jumps arise from stochastic exploration (i.e., not deterministic) and subsequent policy updates triggered by improved reward signals. Note that an iteration refers to a single update step in which the agent changes its policy using data from interactions with the environment. While iterations reflect learning progress, they are not equivalent to simulated time (e.g., years); instead, they measure how often the agent's decision-making is refined. More iterations enable deeper learning and help the agent converge toward a policy that is 'optimal' under the prescribed objective.

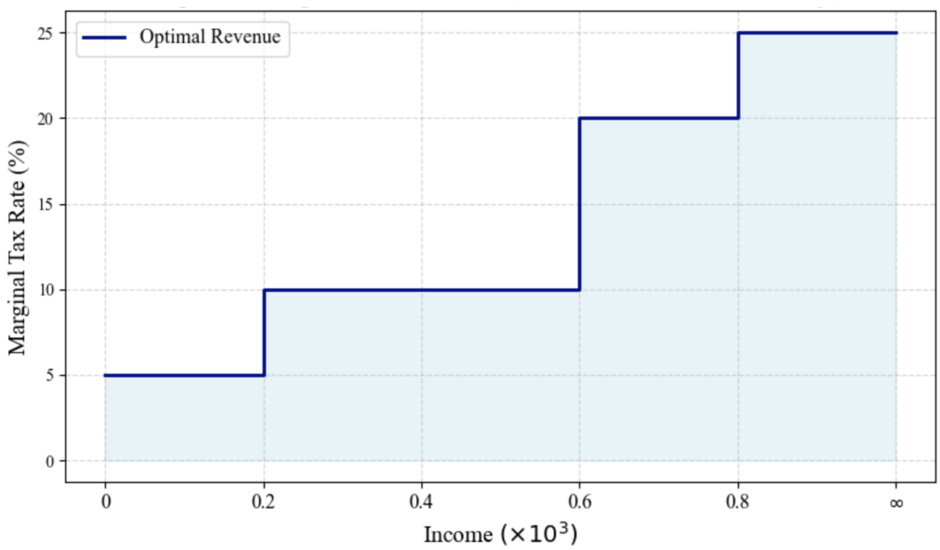

Figure 2: Optimal marginal tax schedule under preliminary simulation

[1]An anonymized GitHub repository with the full code can be found here.

Figure 2 illustrates the optimal marginal tax schedule learned by the DQN under a single policy objective. The highest bracket sees a 25% marginal tax rate, while lower earners are taxed at rates between 5–15%. This optimised strategy led to a GDP of 31.75 and a Gini coefficient of 0.17 with no agents relocating. Interestingly, the model essentially learned the most aggressive tax schedule under which no high-income earners would relocate. Note that although the model permits dynamic adjustments to the marginal tax schedule in each period, the trained planner converged to a static tax schedule. This outcome reflects the DQN's assessment that a consistent tax structure maximized long-run government reward, balancing revenue generation against relocation incentives. The stability of the schedule suggests that under the preliminary model frequent tax adjustments were suboptimal.

It is also worth discussing the computational feasibility of the proposed methodology. This preliminary simulation involved 10,000 iterations with merely 100 agents, taking 30 seconds to run. The full simulation would involve 34 million UK workers. Note that TaxAI, a similar agent-based model that optimizes tax brackets in a closed economy, ran a simulation for 10,000 agents on a NVIDIA GA100 (A100 GPU) that took 12 minutes (Mi et al., 2023. But since their script was written in Python, it could not be parallelized across cores or GPUs. Running it in C++ would allow for a 10-100× speedup (Nourisa et al., 2022). The expected runtime for the full simulation with 34 million workers proposed by this study would therefore be approximately 7–72 hours.

5 Limitations and areas for future research

There are three primary limitations to the proposed study. The first is that the model cannot capture the full range of opportunities, costs, and socio-economic considerations that inform the decisions made by people in the real world. For example, relocation is not possible in all professions and factors such as family ties, legal barriers, or credential recognition often constrain mobility. Although this 'cost' is accounted for as a general financial barrier it is not homogenous for all agents. Future studies could attribute 'sticky' preferences (e..g, by age, family size, geography, profession, industry) based on representative survey data, thereby extending the list of parameters in Table 1.

The second limitations is that although workers in the proposed model are allowed to leave the country in response to unfavourable tax policies, it does not allow for inward migration. The simulated government therefore cannot attract foreign talent to offset domestic relocation. A more holistic open-economy framework would allow for both inflows and outflows of labour, capturing the two-sided nature of tax-induced mobility observed in empirical studies (Kleven et al., 2020; Akcigit and Stantcheva, 2021). The third limitation is that the model must be calibrated with reference to current data to forecast an optimal tax schedule for the next 50 years. This constitutes a fundamental challenge to agent-based modelling, since the calibrated parameters may not remain stable over long horizons (Heathcote et al., 2009; Haldane and Turrell, 2019). As a result, forecasts may be sensitive to assumptions about the persistence of current behavioural and economic conditions.

6 Conclusion

In summary, this work has presented a framework that extends the Real Business Cycle to include a central planner capable of 'learning' an optimal policy with reference to the behavioural response it observes in the simulated economy. In a full implementation of the study, all 34 million workers in the UK economy can be equipped with an empirically representative distribution of their preferences. Agents within the simulation are continuously adapting their behaviour to maximize their respective utility function by choosing (a) the number of hours they work and (b) whether or not to relocate. The central planner in this simulated economy chooses an optimal tax schedule by 'learning' how specific combinations of tax brackets impact total revenue via the Deep Q-Network algorithm. A preliminary simulation was then conducted with 100 agents to demonstrate how the proposed framework effectively converges to an optimal equilibrium under the prescribed objective. These results offer promising insights into how reinforcement learning can extend conventional economic models to allow for more accurate contextualized forecasts by capturing the complex relationship between micro-level decisions and macro-level phenomena.

References

Akcigit, U., Baslandze, S., & Stantcheva, S. (2016). Taxation and the international mobility of inventors. *American economic review*, *106*(10), 2930–2981.

Akcigit, U., & Stantcheva, S. (2021). 6 taxation and innovation: What do we know? In *Innovation and public policy* (pp. 189–212). University of Chicago Press.

Axtell, R. L., & Farmer, J. D. (2025). Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*, *63*(1), 197–287.

Bhagwati, J. N., & Srinivasan, T. (1983). On the choice between capital and labour mobility. *Journal of International Economics*, *14*(3-4), 209–221.

Fabretti, A., & Herzel, S. (2017). An agent based model for a double auction with convex incentives. *Journal of Artificial Societies and Social Simulation*, *20*(1).

Favero, C., & Hendry, F. D. (1992). Testing the lucas critique: A review. *Econometric reviews*, *11*(3), 265–306.

Haldane, A. G., & Turrell, A. E. (2019). Drawing on different disciplines: Macroeconomic agent-based models. *Journal of Evolutionary Economics*, *29*, 39–66.

Heathcote, J., Storesletten, K., & Violante, G. L. (2009). Quantitative macroeconomics with heterogeneous households. *Annu. Rev. Econ.*, *1*(1), 319–354.

Khan, A. (2024). *More than half of hnws consider moving abroad.* FTAdviser.

Kleven, H., Landais, C., Muñoz, M., & Stantcheva, S. (2020). Taxation and migration: Evidence and policy implications. *Journal of Economic Perspectives*, *34*(2), 119–142.

Knox, W. B., & Stone, P. (2015). Framing reinforcement learning from human reward: Reward positivity, temporal discounting, episodicity, and performance. *Artificial Intelligence*, *225*, 24–50.

Lehmann, E., Simula, L., & Trannoy, A. (2014). Tax me if you can! optimal nonlinear income tax between competing governments. *The Quarterly Journal of Economics*, *129*(4), 1995–2030.

Luo, S., Liu, S., Cai, T., & Wu, C. (2025). Utilizing rbc system for taxation policy evaluation: An adaptive interaction framework based on deep reinforcement learning. *Expert Systems with Applications*, *268*, 126365.

Mi, Q., Xia, S., Song, Y., Zhang, H., Zhu, S., & Wang, J. (2023). Taxai: A dynamic economic simulator and benchmark for multi-agent reinforcement learning. *arXiv preprint arXiv:2309.16307*.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, *38*(2), 175–208.

Nourisa, J., Zeller-Plumhoff, B., & Willumeit-Römer, R. (2022). Cppyabm: An open-source agent-based modeling library to integrate c++ and python. *Software: Practice and Experience*, *52*(6), 1337–1351.

Partington, R., & Elgot, J. (2024). Budget 2024: Reeves reveals £40bn in tax rises as she promises to rebuild public services. *The Guardian*.

Stiglitz, J. E. (2019). *People, power, and profits: Progressive capitalism for an age of discontent.* Penguin UK.

Suslov, V., Novikova, T., & Tsyplakov, A. (2016). Simulation of the role of government in spatial agent-based model. *Ekonomika Regiona= Economy of Regions*, (3), 951–965.

Zheng, S., Trott, A., Srinivasa, S., Naik, N., Gruesbeck, M., Parkes, D. C., & Socher, R. (2020). The ai economist: Improving equality and productivity with ai-driven tax policies. *arXiv preprint arXiv:2004.13332*.