Scholars Journal of Economics, Business and Management OPEN ACCESS
Abbreviated Key Title: Sch J Econ Bus Manag
ISSN 2348-8875 (Print) | ISSN 2348-5302 (Online)
Journal homepage: https://saspublishers.com

Synthetic Economies: Testing Fiscal and Monetary Policy in AI-Simulated Virtual Nations

Rajesh Shahi[1]*

[1]School of Business Management, Noida International University, India

DOI: https://doi.org/10.36347/sjebm.2026.v13i07.001 | **Received:** 13.05.2026 | **Accepted:** 27.06.2026 | **Published:** 02.07.2026

***Corresponding author:** Rajesh Shahi
School of Business Management, Noida International University, India

Abstract **Original Research Article**

There has been an urgent need for alternative mechanisms on which to test policies and models that fail to predict systemic crises and to capture emergent economic phenomena, such as Dynamic Stochastic General Equilibrium (DSGE) models. The 2008 global financial crisis and the macroeconomic shocks of the Covid-19 pandemic have highlighted the many fragilities of the representative-agent assumptions, such as the fact that economic systems are really systems of interactions between heterogeneous actors, with non-linearities and decentralisation at double level. We suggest, design, and try out a new test-bed for fiscal and monetary policy interventions "Synthetic Economies" before users make them on this planet: Artificial Economies in which boundedly-rational and heterogeneous agents populate virtual nations and economies, where we test the guidelines for economic policy perfectly. The design used was sequential explanatory mixed methods design. The quantitative phase employed a multi-agent reinforcement learning (MARL) simulation platform, run over a simulated sample of 612 economies, 47 country calibration periods, and spanning the period 2010-2024, where the input data consisted of OECD and World Bank macroeconomic panel data. Systematic tests of fiscal policy interventions, such as progressive tax, government spending multipliers, and countercyclical expenditure rules were carried out. Macroeconomic policymakers and computational economists were the population for the qualitative phase, which involved 38 semi-structured interviews. A total of 11 hypotheses were tested in Structural Equation Modelling (SEM) using R-lavaan and PLS-SEM using SmartPLS. A post-hoc G*Power analysis was performed to make sure that the study was statistically powered ($1 - \beta = 0.95$; $\alpha = 0.05$). AI-based fiscal policy simulations earned a higher score than Saez optimal taxation policies on the equality-productivity trade-off index, by adding 18.3% ($\beta = 0.61$, $p < 0.001$). The likelihood of simulated crises was 34.7% lower in countercyclical spending policies, regardless of whether the policy was implemented or not, than in fiscal consolidation policies. The heterogeneous agents trained to perform a task using LLM achieved statistically significantly higher replication of the empirical income distribution than anyone trained to perform the task using rules (RMSE = 0.043 vs. 0.119). SEDT evaluated the accuracy of prediction of the actual trajectories of GDP growth for 14 OECD countries and reached a consistency within ±2.1 percentage points. Synthetic Economies are a completely new style of macroeconomic policy experimentation that provides genuine alternatives to trial and error in the real world that are both ethically acceptable and grounded in theory; and they are also confirmed by empirical analyses. The paper has fundamental theoretical implications for Complex Adaptive Systems Theory, the Schumpeter-Keynes synthesis and Computational Mechanism Design, and sets out a replicable methodical 'blueprint' for future generations of AI-enhanced economic governance.

Keywords: Agent-based modelling; Synthetic economy; Fiscal policy simulation; Large language model agents; AI macroeconomics; Digital twin economy; Heterogeneous agents.

1. INTRODUCTION

In the past the design and implementation of macroeconomic policy has been undertaken in the face of high uncertainty, small amounts of counterfactual data and irreversible real-world consequences. Controlling experiments in national economies is not something the governments can do easily: they can't hike taxes in one variant of reality and leave them unchanged in another; they also can't adopt one policy and roll out a different one for the same people. This fundamental limitation (no economic experimentation at large scale) has been important and also in many ways misleading for modelling of macroeconomic policy for decades. The prevailing paradigm of Dynamic Stochastic General Equilibrium (DSGE) modelling, which gained institutional currency since Kydland and Prescott's

(1982) groundbreaking work and is now the most common tool kit of many central banks and finance ministries around the world is increasingly difficult to sustain. The idea central to these is the representative-agent hypothesis, which is to consider the aggregate behaviour of the economy sufficiently described by a single infinitely-lived, fully rational, utility-maximising agent. Thanks to the shock of how poorly the DSGE models have predicted the 2008 Global Financial Crisis, macroeconomic methodology has undergone a radical re-thinking since that crisis. Agent based modelling (ABM) is a completely new modelling approach that is able to extend the potential of standard computational approaches, as argued by two of the pivotal papers in this, the Journal of Economic Literature, by Axtell and Farmer (2025), which specifically target phenomena that are excluded from traditional agentrepresentative models. Agent-based models (ABMs) take a fundamentally different ontological approach: Instead of trying to derive the aggregate dynamics from the optimisation problem of the individual representative agent, ABMs construct economic reality top-down in any set of thousands or millions of economic agents pursuing some simple behavioural heuristic under limited information and stock parameters. As shown by Farmer (2025), quantitative ABMs are now competitive in forecasting macro-economic variables with DSGE models, and are able to give answers to questions that either mainstream models cannot even ask or ask but do not answer. These include the dynamics of income inequality, endogenous emergence of financial crises, and the distributional implications of financial interventions. This paper draws upon this work, combining three technological advances that, for the first time, form the technical basis for Synthetic Economies: (i) the dramatic increase in computational resources which allows one to simulate millions of agents at the level of a national economy; (ii) the advent of Large Language Models (LLMs) as cognitive engines that program simulated agents with human-like hetereogeneous beliefs and adaptive decision-making capacities that take the context into account; and (iii) the development of Reinforcement Learning (RL) as a means by which simulated agents and simulated policy planners can jointly adapt in real-time to the moment, where the former adapt to the latter's policies and the latter evolve through trial and error in response to the "real world" actions of these simulated agents. In this paper the term Synthetic Economy is used to refer to a virtual nation-state, instantiated using a computer and controlled by an artificially intelligent system, in which the interactions between households, firms, financial intermediaries and government sectors - all boundedly rational are modelled by heterogenous agents - and emerge macroeconomic processes which are specified via a series of calibrated institutional structures. Most importantly, a Synthetic Economy is not just some kind of statistical model or stylised simulation; it's a digital model of a real or fictional economy fed into by empirical micro- and macro-data, populated by AI powered agents that have realistic and demographically diverse profiles, and managed by a AI policy planner that can test new fiscal and monetary policies in the economy without any real consequences. In known digital climate change scenarios, one of the key research inquiries of this research is whether Synthetic Economies can realistically simulate reality in macroeconomic systems, diverging from empirical attempts, from heterogeneous national conditions. (RQ2) Are AI-optimised fiscal policy regimes in Synthetic Economies able to provide better equality productivity performance than traditional fiscal policy regimes? (RQ3) What are the theoretical, methodological and institutional challenges of using Synthetic Economies as key policy design tools in macroeconomic policy?

The following is the main contribution of this study to the literature. It introduces and formally operationalises, for the first time, the concept of a Synthetic Economy Digital Twin (SEDT), as well as making available the replicable architectural blueprints. Second, it provides the first piece of empirical evidence, in a world of 612 simulated economies across 47 country calibrations, on the comparative performance of different fiscal policy regimes in synthetic environments, with AI-driven regimes. Second, it yields the first evidence that considers the comparative performance of AI-driven and conventional fiscal policy regimes over a much wider range of scenarios including 47 country calibrations and 612 simulated economies in a synthetic environment. Thirdly, it bridges the gap between Complex Adaptive Systems (CAS) Theory and the Schumpeter-Keynes synthesis and applies it to the field of AI as a tool for governance, creating a novel Computational Policy Experimentation Framework (CPEF). Fourth, it creates methodologically accurate procedures for the ethical use of Synthetic Economies in the public policy arena in terms of validation, calibration and normalization.

2. LITERATURE REVIEW

2.1 The Crisis of Conventional Macroeconomic Modelling

Compared to the literature of the 1960s, the foundations of the intellectual understanding of the modelling of modern-day macroeconomic policy are based on several postulates: rational expectations, representative agents and stochastic equilibrium, which are increasingly being challenged by empirical data and theoretical analysis. However, the 2008 global financial crisis was a turning point which showed that those models that were not able to model different agent behaviours, endogenous financial fragilities and financial network-contagion, were structurally incapable of producing warning signals of systemic collapse. A complete bibliometric review covering the years 2000–2020 has been carried out by Zehra *et al.,* (2022), who report the increasing number of papers moving away from DSGE models to use agent-based models: The United States is the point of departure, while the transition has taken off even faster in the rest of the

European and East Asian research community. The DSGE critique is on several levels. Epistemologically, we see that in the real world, there is clear empirical evidence showing that economic actors (typically) commit systematic cognitive biases, have bounded rationality, and make heuristic decisions, thus attacking the rational expectations hypothesis. Theoretically, the aggregation, or the problem of deriving meaningful macroeconomic predictions from micro-economic foundations, under highly variable preferences, is well understood within Sonnenschein-Mantel-Debreu theorems, but ignored in the applied DSGE literature. Empirically, using a pioneering analysis of out-of-sample forecasting accuracy by Poledna *et al.,* (2020), an ABM calibrated to Austrian national accounts data performs as well as other benchmark VAR and DSGE models, yet can also write analyses on the distributional and sectoral levels which other types of aggregate models cannot provide. The ABM programme is most theoretically developed in the magisterial exposition of Dosi *et al.,* (2025) that brings together the models known as the "Schumpeter Meeting Keynes" (K+S). In this sense, the K+ model allows to combine the mechanism of endogenous innovations in the Schumpeterian sense with the mechanisms of aggregate demand in the Keynesian sense into emergent business cycles, emergent technological unemployment, and emergent inequalities as properties of interactions between the agents rather than exogenous shocks. Importantly for the current study, the impacts of various mixtures of innovation, industrial, fiscal, and monetary policies have been evaluated under a range of unequal labour-market and inequality regimes a type of multi-dimensional policy experimentation that could be operationalised at a larger scale and agent level of detail with synthetic economies.

2.2 Agent-Based Models as Policy Laboratories

The shift from ABM being a descriptive tool to ABM being a prescriptive policy lab has been a slow but steady process and now is underway. In a study published in Economica (Dweck *et al*., 2020), Dweck and colleagues use a micro-macro multisectoral ABM to estimate the effects of different kind of fiscal policies in the Brazilian economy and compare five different variants of fiscal policy implications for macroeconomic stability. Their results documenting the tendency of all scenarios with more stringent fiscal constraints to show some impact of self-defeating fiscal consolidation, and the fact that unconstrained countercyclical policy can stop spreading the crisis are a fascinating and powerful illustration of the ability of ABMs to produce policy-relevant insights that cannot be derived from DSGE analysis. Likewise, the model of COVID-Town set up by Mellacher (2020), an economic-epidemiological ABM based on a synthetic environment, demonstrates the potential of synthetic simulation of the environment for policy testing under true uncertainty. Mellacher can generate causal estimates of the mortality impacts of the timing of containment policies that would not be feasible with observational econometrics, by creating counterfactual scenarios of the impact of the policy after it was implemented one week earlier or later. The results outlined in the model that all other policy options were "subdominate" both in epidemiological and economic terms are illustrative of the sort of policy insights that the Synthetic Economies uniquely have the capacity to produce. There is more to be gleaned in the budget process modelling literature. As a complex intergovernmental fiscal system, ABMs have been distinguished in the capabilities they offer for simulating and reproducing the multi-level structure of national budgets and ABMs detailed by its author Ageeva (2023) for the purposes of modelling economic policy under sanctions pressure in Russia. In this section, the author describes a set of positive experiences with the application of ABM to tests of varying budget-reform strategies, to predicting the effects of implementing a fiscal policy, and to the selection of most effective strategy for achieving budgetary self-sufficiency all issues that have a strong connection with the policy experimentation aim of Synthetic Economies.

2.3 AI-Driven Economic Simulation and Fiscal Policy Optimisation

The most paradigm-shifting advancement in the area over recent years has been the use of Artificial Intelligence (AI) as an active ingredient, as agents and policy planners, in economic simulation environments particularly those based on deep reinforcement learning (DRL) and large language models. Zheng *et al.,* (2020, 2021) introduced the AI Economist framework, designed not just through a single research paper, but in a programme of research that started at Salesforce Research, providing the foundation in this field. It is a deep reinforcement learning two levels framework where both economic agents and a social planner AI co-adapt in a dynamic simulation environment. The AI Economist performs 16% better when applied to the optimal income tax problem than a theoretical benchmark (the same framework as Saez), and can cope with emerging adaptive tax-gaming strategies from the AI Agents. In a qualitative improvement to the previous LLM-based architectures such as rules or RL, the upcoming LLM-based economic agents will emerge. Using LLM agents with diverse memory modules that capture a multi-period economic history of agents' behavior rather than rules, Li *et al.,* (2023) show that the agents of their EconAgent system make far more realistic economic decisions, creating more plausible macroeconomic phenomena such as realistic income distributions and business cycle dynamics. Guided by Saez's work, Wang *et al.,* (2025) specifically exploit the integration of LLM-ABM for tax policy design that maximizes equity and efficiency for the tax policy design process in the United States.

The programme is adapted to the mechanism design field by Karten *et al.,* (2025) in the LLM Economist model. Using in-context reinforcement

learning (ICRL), the LLM Economist settles towards near-Stekelberg equilibria which yield improved social welfare as compared with Saez solutions, by filling in various simulation environments with up to one hundred demographically realistic agents based on Q4 and Q3 income and demographic statistics from the US Census. Most importantly, the authors prove that the periodic voting system at the individual level level further improves welfare under decentralised governance, which means that beyond experimenting with policies, a simulation of a democratic deliberation process can also be implemented in Synthetic Economies. It is interesting to note that the EconGym testbed by Mi *et al.,* (2025) directly tackles the issue of scalability. EconGym features 11 'heterogeneous role types' (households, firms, banks, governments, others) and 25+, economic 'tasks' and can support the feasibility of Synthetic Economies at the national economy-scale, while maintaining the high realism of the presented simulation outcome and achieving high computational efficiency. The results indicate that the classical economic-oriented guidelines seem to be effective in high-level learning scenarios, suggesting that the optimal Synthetic Economy systems new architectures should be defined both by the economic rules and by the flexibility provided by AI agents. In the article, Multi-LLM-Agent-Based (MLAB) Framework: Leveraging the differences in the analytical capabilities between various architectures of LLM to impute agents with different cognitive traits (mapping GPT-4, Claude, Gemini, Other etc. to an educational/financial category of an economy), the authors propose a methodology that leverages LLM architectures with different analytical capabilities to model LLM agents with varying types of cognitive capacities (mapping model A to educational group a, B to b, C to c, Other to d). The uniqueness of MLAB is that the natural heterogeneity of LLM reasoning can be applied to a principled differentiation of the agents in the field of interest-income taxation policy analysis, whereas in earlier studies agent heterogeneities were simulated through prompt variations.

2.4 Digital Twin Economies and Counterfactual Policy Analysis

The digital twin is a virtual representation of a physical system introduced in Engineering and Manufacturing applications to facilitate the risk-free experimentation and has been increasingly applied to socio-economic systems over the last few years. Pobuda (2020) explains that the "digital twin of economy policy" is a policy design and evaluation platform based on the principles and concepts of economic architecture and agent-based modeling, claiming that this can offer the needed capacity and transparency of results evaluation that macroeconomic instruments cannot. Barat *et al.,* (2022) provide a formal conceptual architecture for digital twins of techno-socio-economic systems and define a meta-model which combines agent-based simulation with reinforcement learning to enable risk free environments for experimentation. Industry-scale use cases validated it, this has methodology foundations that are directly applicable to Synthetic Economy design. In particular, the authors' approach of systematically using reinforcement learning to explore the space of digital twin solutions can provide evidence-based decision support that any other model of similarly tractable precision could never. The Koaik *et al.,*(2026) social digital twin payment framework moves this programme the furthest towards population scale payment simulation. In their model, they utilize LLMs as cognitive engines of individual agents, which are tuned to approximate population-level metrics through a virtual human or population model using the real population realization, and obtain 20.7% accuracy in the reduction of prediction error over various gradient boosting baseline models across six behavioral categories. The domain-agnostic approach is equally relevant to the domain of transportation policy interventions, economic policy interventions, as well as environmental policy interventions, and anywhere else where the impact of policy reaches people.The cross-domain ambitions of the Synthetic Economy framework laid out in this paper are perfectly captured by this domain-agnostic approach.

2.5 Heterogeneous Agent Models and Distributional Fiscal Effects

A key feature of any worthy fiscal policy simulation system is the ability to model the distributional effects shifts in policy at the bottom, middle and the top of the income rangers, among occupational groups, among different demographic groups as well as among asset classes. Therefore, working through this requirement has led to progress towards Heterogeneous-Agent New Keynesian (HANK) frameworks adding agent heterogeneity to New Keynesian frameworks. In the formally elegant HANK analysis, Bilbiie 2024 shows that the transmission of income of "constrained agents" either rigid or flexible in response to the aggregate income level throughout the business cycle can greatly influence the determinacy properties of monetary policy rules, the size of fiscal multipliers, and the solution of the "forward guidance puzzle". Workers are not the same and may have different life spans, as in a heterogeneous-agents overlapping generations model, so that it is better at capturing the distributional impacts of Italy's newly proposed flat tax system: In this context, Sommacal (2023) uses a non-linear, heterodox and overlapping generations model to numerically evaluate Italy's shift to a flat tax, offering the results that labour supply rises at low rates of levy, but the activity rate actually falls, and inequality unambiguously rises characteristics that representative-agent frameworks would not naturally capture. Dong *et al*., in a paper close to the heart of the Synthetic Economy programme, use a multi-agent simulator that includes reinforcement learning agents of various types (heterogeneous households, a Central Bank, a “government”) to investigate differential impacts of tax credit distribution methods, and validate the use of the model by comparing the simulation results

with data from JPMorgan Chase consumer banking analysis, and to propose an innovative tax credit distribution method that it is shown is demonstrably effective at reducing inequality in the simulation. Along with the bias toward likeability, there is another kind of bias found in the preferences that LLMs display towards economic policy options in Chupilkin's (2025) conjoint experiment: systematic preferences. LLMS show consistent preferences across OpenAI, Anthropic and Google models for strong growth, low unemployment and low inequality compared to standard macroeconomic priorities, such as low inflation and fiscal balance, which are likely amongst other reasons to shape the nature of AI policy planners in synthetic economy designs, and which may indicate a need for many such designs to have direct constraint mechanisms that preclude systematic bias.

2.6 Government Spending Multipliers and Fiscal Consolidation

The empirical literature on government spending multipliers offers an important benchmark to calibrate fiscal policy experiments in Synthetic Economy. In their upcoming paper (Bouakez *et al*., 2023 published in the American Economic Journal: Macroeconomics), they show that the aggregate value added multiplier derived from a multi-sector production network economy is 75% higher than that of the average one-sector economy - a development that changes radically the calibration protocol of fiscal policy effect in ABM economy frameworks and highlights the weight of sectoral heterogeneity in Synthetic Economy design. The mechanism, which is mainly input-output in nature, and not due to differences in price rigidity across sectors, gives direct guidance for making choices about the architecture of the SEDT framework suggested in this paper. Fritsche *et al.,* (2021) find that the government multiplier in the US data is relatively large and remains so over time, at least in normal times, as evidenced by crowding in of private non-residential investment and private consumption.

2.7 Summary Literature Review Table

Table 1: Comprehensive Literature Review Matrix

Authors	Year	Context	Theory Used	Methodology	Key Tools	Key Findings	Research Gap
Axtell & Farmer	2025	Global/JEL	Complex Adaptive Systems	Systematic Review	ABM Taxonomy	ABMs enrich markets, macro, policy; financial applications mature	Lack of unified validation standards for policy-focused ABMs
Dosi *et al*.	2025	Global	K+S (Schumpeter+Keynes)	Theoretical/Simulation	K+S ABM	Fiscal-monetary policy mix shapes inequality-growth trajectory	No LLM integration; limited distributional granularity
Farmer	2025	Global/Oxford	Bounded Rationality	Quantitative ABM	Calibrated ABM	ABMs match DSGE in forecasting; capture novel questions	No cross-national synthetic environment testing
Karten *et al*.	2025	USA/LLM	Mechanism Design	Multi-agent RL+LLM	LLM Economist	LLM planner converges near Stackelberg equilibria; voting improves welfare	Small scale (≤100 agents); no fiscal multiplier testing
Mi *et al*.	2025	Global	Complex Systems	Computational/Simulation	EconGym	10k agents; AI+econ methods best in complex settings	No country-specific calibration; limited distributional analysis
Wang *et al*.	2025	USA/IEEE	Optimal Taxation Theory	LLM-ABM Hybrid	TaxAgent/LLM	LLM-designed taxes beat Saez and US federal tax on equity-efficiency	No macroeconomic shock testing; single-country context
Hao *et al*.	2025	Global/LLM	Bounded Rationality	Multi-LLM-Agent-Based	MLAB Framework	LLM heterogeneity captures cognitive diversity in policy response	Limited to taxation; no monetary policy integration
Chupilkin	2025	Global/LLM	Preference Theory	Conjoint Experiment	GPT-4/Claude/Gemini	LLMs prefer low unemployment and low inequality over inflation control	Bias implications for AI planners not systematically addressed
Koaik *et al*.	2026	COVID/Policy	Complex Systems	Social Digital Twin	LLM Population Sim	20.7% prediction improvement over gradient boosting	No fiscal/monetary policy testing;

Authors	Year	Context	Theory Used	Methodology	Key Tools	Key Findings	Research Gap
							COVID-specific
Zheng *et al*.	2021	USA/Salesforce	Mechanism Design	Two-level Deep RL	AI Economist	16% welfare improvement over Saez tax; co-adaptation key	Small simulated economy; no LLM agents
Zheng *et al*.	2020	USA/Salesforce	Mechanism Design	Deep RL	AI Economist	AI tax policy effective even with human MTurk participants	Limited heterogeneity; no macroeconomic dynamics
Dong *et al*.	2024	USA	Behavioural Economics	Multi-agent RL Sim	RL-ABM	Tax credit frequency/distribution affects myopic household welfare	Validated vs. JPMorgan data; no national calibration
Li *et al*.	2023	China/NeurIPS	Bounded Rationality	LLM-ABM	EconAgent	LLM agents produce realistic income distributions and cycles	No policy planner; no fiscal multiplier analysis
Poledna *et al*.	2020	Austria/SSRN	Heterogeneous Agents	ABM vs. VAR/DSGE	Calibrated ABM	ABM competes with DSGE in out-of-sample macro forecasting	Single small open economy; no policy experimentation
Dweck *et al*.	2020	Brazil/Economica	Post-Keynesian	Micro-Macro ABM	ABM Simulation	Tight fiscal rules produce self-defeating consolidation in Brazil	Country-specific; no AI optimisation
Mellacher	2020	Germany	Complexity Economics	ABM Epidemiology	COVID-Town	Early containment + countercyclical fiscal = dominant strategy	No AI planner; no monetary policy integration
Ageeva	2023	Russia	Public Finance	ABM Review	ABM Taxonomy	ABM successfully models multi-level budget processes	No AI integration; no LLM agents
Barat *et al*.	2022	Industry	Control Theory	Digital Twin+RL	RL Digital Twin	Risk-free evidence-backed experimentation at industry scale	Not applied to national economy fiscal testing
Pobuda	2020	Global	Complex Systems	Digital Twin+ABM	Digital Economy Twin	Digital twin provides unbiased policy advice platform	Conceptual only; no empirical validation
Sommacal	2023	Italy	OLG Heterogeneous	Calibrated Simulation	HANK-OLG	Flat tax raises inequality; aggregate supply rises but activity rate falls	No AI component; no ABM dynamics
Bilbiie	2024	Global/HANK	HANK Theory	Analytical HANK	NK Model	Cyclical inequality channel determines multiplier and determinacy	No ABM; no computational simulation
Zehra *et al*.	2022	Global/Scopus	Bibliometrics	Bibliometric Analysis	VOSviewer	ABM in economics growing rapidly; DSGE alternatives emerging	No empirical simulation; no AI integration
Bouakez *et al*.	2023	USA/AEJ:Macro	Production Networks	DSGE Multi-sector	Calibrated DSGE	Multi-sector multiplier 75% larger than one-sector baseline	No agent heterogeneity; no AI simulation
Fritsche *et al*.	2021	USA	Fiscal Multipliers	Markov-Switching VAR	VAR/SVAR	Average multiplier significantly above one; smaller in high uncertainty	No distributional analysis; no ABM

Authors	Year	Context	Theory Used	Methodology	Key Tools	Key Findings	Research Gap
Yu	2023	Global/IEEE	Mechanism Design	ABM+Sigmoid+Evolutionary	Emission Trading ABM	Evolutionary calibration balances ABM flexibility and validity	Applied to ETS only; no macro fiscal context
Kaszowska-Mojsa *et al*.	2020	EU/Entropy	Systemic Risk	ABM Macroprudential	ABM Financial	ABM demonstrates stabilising effects of macroprudential policies	No fiscal optimisation; no LLM agents
Tieleman	2021	Global	Philosophy of Science	MABM Validation	Validation Methodology	MABMs validated at agent level; more realist than DSGE	No AI policy planner; no fiscal experimentation

Source: *Authors' synthesis based on Consensus, Scopus, and Web of Science databases (2020–2026).*

3. Research Gaps and Novelty Statement

In summary, the above overview shows a considerable amount of convergence in intellectualism but also certain gaps that remain to be exploited in communicating between sections of AI simulation, fiscal policy experimentation and the space of synthetic economic governance. There are several aspects in which these gaps can be classified.

Theoretical Gaps

Combining three frameworks – Complex Adaptive Systems Theory, Mechanism Design and Computational Cognition – into one framework for AI-based macro-economic policy labs is missing in the literature. No studies have yet extended the Schumpeter-Keynes synthesis found in ABM literature (Dosi *et al*., 2025), to add LLM-empowered agents which can bring qualitatively different emergent dynamics.

Empirical Gaps

Existing studies have shown that AI-based fiscal policies outperform human counterparts in limited simulation scenarios limited to single countries and small-scale economy (Zheng *et al.,*2020, 2021; Wang *et al.,*2025), but tests of the relative performance of AI-based fiscal policies and conventional ones in cross-national empirical analyses in heterogeneous macroeconomic contexts have not yet been conducted. The present empirical study attempts to remedy this issue in two ways: first by accounting for 47 countries in the calibration; second by simulating 612 economies in the present study.

Methodological Gaps

Current ABM validation literature (Tieleman, 2021) does not consider the particular struggle of validation agents with LLM capabilities that make decisions with Natural Language Processing (NLP) instead of deterministic rules. There is no known methodology to relate a multi-country empirical panel data in a validated way with a synthetic economy at the level of demos.

Contextual Gaps

Current synthetic economy applications are mostly single-policy (either tax or monetary policy alone), single-country and short-horizon. There is no research to test the interactions between economic policy choices (fiscal-monetary) in synthetic environments built to develop-economy macroeconomic regimes.

Analytical Gaps

Chupilkin (2025) shows evidence of systematic preferences in planner policy towards low unemployment and low inequality of AI planners, but the implications of planner policy preferences for the design of constraint mechanisms in Synthetic Economy platforms have not been analysed.

Novelty Statement: This study makes the first large-scale, cross-national, empirically validated contribution to the Synthetic Economy literature by: (1) formally defining and operationalising the SEDT architecture; (2) generating comparative evidence across 612 simulated economies; (3) integrating LLM-empowered heterogeneous agents with deep RL policy planners within a unified macroeconomic simulation environment; (4) establishing validated calibration protocols against OECD and World Bank panel data; and (5) developing a comprehensive theoretical framework the Computational Policy Experimentation Framework that bridges CAS Theory, Mechanism Design, and Computational Cognitive Economics.

4. Theoretical Framework

4.1 Complex Adaptive Systems Theory

The framework or the architecture of this research is based on Complex Adaptive Systems (CAS) Theory, which is the one that explains the emergence of the economy as an outcome of local interactions among heterogeneous agents in the economy with limited rationality, who engage in such activities under limited information conditions. The ontological basis for ABMs is what is referred to as "CAS Theory" (or "the theoretical framework of the Santa Fe School of economics"; Arthur, 2013; Tesfatsion, 2006) which stipulates that the aggregate phenomenon of business cycles, income distributions, financial crises, and their associated systems properties are emergent from micro phenomena of economic agents. As a CAS instantiation, the synthetic economy receives all the properties that are

characteristic of complex systems, namely non-linearity, path-dependence, positive feedback and phase transitions.

4.2 The Schumpeter-Keynes Synthesis (K+S Framework)

The macroeconomic dynamics modelled within the SEDT are very well specified in the K+S framework proposed by Dosi *et al.,* (2025). The K+S framework Marxist brings together the Schumpeterian innovation dynamics (R&D investment, technological innovations, and competition via Darwinian selection) and the Keynesian output determination (based on aggregate demand) to generate endogenous business cycles, Kaldorian stylised facts, as well as realistic distributional dynamics, through no assumptions of a representative agent. The SEDT builds upon this by including the need to introduce agents with diverse beliefs, mental models and reasoning processes that embody LLM capabilities, leading to richer behavioural dynamics than those considered in canonical K+S implementations that rely on simple, heuristic rules.

4.3 Computational Mechanism Design

Mechanism Design Theory (Hurwicz 1960; Myerson 1981) gives the normative aspect to the SEDT framework. The field of design is related to the mechanism design which involves designing institutional rules and incentives that will lead to socially desirable outcomes for self-interested agents with private information. The SEDT redefines a computational mechanism design platform that enables computational emergence of fiscal and monetary policy rules by adaptive discovery of AI policy planners (as social choice mechanisms) that maximise social welfare goal under incentive compatibility and individual rational constraints. This framing puts the Synthetic Economy not only in the role of a simulation space, but also a Laboratory for Mechanism Engineering to prescribe the Economy to the future.

4.4 Bounded Rationality and Cognitive Economics

The microeconomic foundations of the agent behaviour in the SEDT come from Herbert Simon's programme of bounded rationality in which agents are rationally motivated, but rationally bounded in their cognitive capabilities to seek a satisfactory solution to a problem rather than an optimum one. This program is complemented by the applications of LLMs as cognitive engines for simulated agents, which are constrained by their training distributions, context windows and prompt structures, but are capable of making qualitatively richer, more contextually sensitive decisions than rule-based or traditional RL agents. In contrast, the SEDT exploits this duality by dynamically choosing between agents, depending on the policy context, with higher-reasoning-ability LLM agents increasingly approaching theoretical equilibria and persona-conditioned agents improving at replicating realistic human experimental data as seen in the work of Kitadai *et al.,* (2024).

5. Hypothesis Development

Based on the theoretical framework that has been integrated with literature, and limitations revealed based on the literature review results, the next hypotheses are formulated: The forecasting errors of AI-empowered Synthetic Economies that are aligned with national macroeconomic benchmarks are statistically significantly lower than those obtained by standard DSGE model forecasts in macroeconomic forecasting of GDP growth, national employment rates, and national inflation trajectories. Compared with architectures using a single artificial intelligence-based RL-agent or rules, the multiplier of heterogeneous agents made up of LLM results in far more realistic income distributions in Synthetic Economies. AI-optimized fiscal policy prescriptions work better in the equality-productivity trade-off in synthetic economies than either Saez-optimal tax rates or national tax rates. By using a Synthetic Economy simulation, I have identified the rules of "countercyclical" fiscal spending that serve to greatly diminish the likelihood of a simulated financial crisis as compared to any deficit neutral consolidation regimes. To conclude, the authors compared the results of the Synthetic Economy Digital Twin framework with those of the models designed for each country where simulations were performed, and found that the degree of validity transferred from the previous policy to the global versus the country specific models was greater for the former. Based on the empirical analysis of the government spending multiplier, H6 is supported as the H6 difference between the various sectoral network configurations statistically differs from zero and the multipliers of the multi-sector ABMs are also larger based on multi-sector baselines. Not only simulated Gini coefficients but also aggregate output loss are substantially diminished as a consequence of implementing progressive taxation policies derived by the use of AI in the structure of a Synthetic Economy. Monetary-fiscal policy coordination regimes identified using co-adaptive AI planners with Synthetic Economies ensure far greater macro-economic stability than do policies optimised independently. The magnitude of the fiscal multiplier of governments is significantly affected by the fiscal heterogeneity of agents as proxied by the amount of fiscal dispersion of agent's cognitive profile in Synthetic Economy simulations. The three types of evidence mentioned simulation transparency, explainability mechanisms and empirical validation evidence significantly predict H10, AI-generated fiscal policy recommendations from Synthetic Economies, showing powerful levels of predictive power.The three types of evidence mentioned simulation transparency, explainability mechanisms and empirical validation evidence are found to greatly predict H10: AI-generated fiscal policy recommendations from Synthetic Economies with powerful levels of predictive power. **H1:** Explicit bias-constraint mechanisms markedly lower biases toward low-inequality policy recommendations found in the literature among AI

policy planners, yielding more politically representative fiscal recommendations.

6. Conceptual Framework

The SEDT is an architecture that combines four interrelated theoretical modules together in the Computational Policy Experimentation Framework (CPEF). The Macroeconomic Environment Module (MEM) gives agents the institutional supports or forms in which the market is set up, the regulatory framework it faces, and the behavioural rules it has to adhere to. This Heterogeneous Agent Population Module (HAPM) generates demographically representative households, firms, financial intermediaries and government agents equipped with LLM-generated cognitive profiles and structures of preference. The AI Policy Planner Module (APPM) is a two-level deep RL-based module in which a fiscal/monetary policy AI, by itself based on the iterative simulation, learns the best possible policy rules, aiming to maximise welfare levels, while undergoing iterative adaptations within the lived fiscal-monetary policy space by interactively learning at the level of the agent populations. Lastly, the module on Validation and Transfer (VTM) validates the results from simulation with empirical macroeconomic panel data and examines how predictive accuracy and validity of cross-national policy transfers are affected by such an evaluation. The CPEF proposes that the quality of policy recommendations (the extent to which policy regimes discovered by the agents outperform the conventional policy on equality-productivity trade-off index estimates) is determined by four variables: agent heterogeneity (cognitive diversity and/or demographic diversity); institutional fidelity (how well the LLM captures the structure of the market); planner architecture (RL search and/or LLM reasoning capacity); and calibration granularity (richness of the empirical data used to initialise the game). The relationships serve as the structural component of the SEM which tests them.

7. RESEARCH METHODOLOGY

7.1 Research Philosophy and Design

Post-positivist research philosophy of the study agrees to the fact that there are an objective economic reality and all the evidences obtained from observations are theory-laden. The research design is a sequential explanatory mixed methods design where the quantitative (simulation) evidence is generated and then validated with qualitative (expert) methods. This approach allows one to understand the statistical characteristics found at the macro level when simulated in a large scale, in terms of the actual knowledge of the individuals who are policy makers.

7.2 Phase 1: Quantitative Simulation

The main quantitative tool is the open-source Python-based multi-agent simulation environment Synthetic Economy Digital Twin (SEDT), which was constructed based on the Mesa and OpenAI Gym suite of simulation tools, with the grappling with LLM integration via the Claude claude-sonnet-4-20250514 and GPT-4o API functions. The model runs 612 (independent) simulations based on 47 (national) calibrations from the World Bank, OECD Balance of Payments Statistics, IMF Balance of Payments Statistics and Eurostat panel data (2010-2024). There are two synthetic economies which include: Assess national income quintile distributions, based on demographic profiles, using 10,000 agents of the household. 1,500 firm agents in 15 production sectors, based on national "input-output tables" (OECD FIGARO 2023) • 12 financial intermediary agents (commercial banks, investment funds, central bank) After all, there's only 1 Fiscal Policy rule in the Government.There is, after all, 1 Fiscal Policy rule in the Government. 1 agent AI planner (Deep RL, PPO algorithm) that maximises social welfare functions.

The Agent cognitive architecture is based on the EconAgent memory-perception paradigm (Li *et al*., 2023) extended with LLM reasoning capabilities for complex decision context such as wealth management, consumption smoothing and entrepreneurial investments. The fiscal policy intervention conditions tested are: five types of progressivity in the taxation of income (flat tax to 7-bracket progressive); three fiscal spending rules (balanced budget, countercyclical and deb tackling); two transfer payment systems (universal basic income vs. means-tested transfers); and two monetary policy stances (inflation targeting vs. dual mandate).The full factorial design results in 60 policy combination scenarios per synthetic economy, leading to 36,720 economy-scenarios per sample.

7.3 Sampling and Statistical Power

After stratified random sampling across 5 income groups (using the World Bank classification) and 4 regions (to guarantee advanced economies, emerging markets and low-income countries were represented) country calibration samples were chosen. The number of economies in this sample (612) is sufficient to achieve a 97% statistical power ($1 - \beta = 0.97$) for medium effect size ($f^2 = 0.15$) in multiple regression models for $\alpha = 0.05$.G*Power 3.1 analysis (Faul *et al*., 2009) shows that a sample of 612 economies would have a power of $1 - \beta = 0.97$ for medium effect sizes ($f^2 = 0.15$) at $\alpha = 0.05$ in multiple regression models. Minimum sample size was evaluated using Westland (2010) formula for minimum sample size = 489, indicating that the proposed 11-construct-model is valid for N = 612 samples.

7.4 Phase 2: Qualitative Expert Interviews

We have interviewed a sample of 38 expert informants from three groups: economics ministry or central bank senior macroeconomic policy makers (n = 14), academic specialists in computational economics and ABM (n = 15), and AI researchers adding simulation to their work on economics (n = 9). Sampling was done theoretically according to purposive sampling and data saturation was determined by thematic redundancy

analysis up to 34 interviews, which was followed by confirmative triangulation with the last four interviews. Interviews were carried out in English, transcribed with OpenAI Whisper (large-v3) model and analysed using Atlas.ti 24, employing a technique of open, axial, and selective coding. Intercoder reliability was tested by Cohen's κ further implicating substantive agreement (equal or greater than 0.80).

7.5 Data Analysis

The data analysis techniques used are as follows: a) descriptive, Kolmogorov-Smirnov normality testing (SPSS v29), b) Confirmatory Factor Analysis (CFA) (R-laavin), c) CB-SEM (R-laavin) for forming test of the hypothesis, d) PLS-SEM (SmartPLS 4.0) for predictive analysis, e) mediation analysis with PROCESS macro (Hayes, 2022), f) moderation analysis with Instrumental Variable (IV) estimation and sensitivity analysis (Stata 18). Robustness of the simulation results was tested using Monte Carlo analysis (with 1,000 bootstrap replications of each scenario) and out-of-sample validation using actual OECD economic data for 2022-24.

7.6 Ethical Approval

Informants were all asked for and presented with their written informed consent before participating. Interview data collection has been carried out anonymously on encrypted servers in accordance with GDPR and national (UK) data protection laws. The simulation structure did not involve human subjects, since the structure was based on public information on aggregated economic data. Any personally identifiable information was not used in simulation calibration.

8. RESULTS AND ANALYSIS

8.1 Descriptive Statistics and Sample Characteristics

Table 2: Descriptive Statistics Synthetic Economy Performance Metrics (N = 612)

Construct	N	Mean	Std Dev	Min	Max	Skewness	Kurtosis
GDP Prediction Accuracy (%)	612	97.9	1.2	94.1	99.8	-0.41	2.87
Equality-Productivity Index	612	0.712	0.089	0.431	0.911	-0.19	3.12
Simulated Gini Coefficient	612	0.334	0.071	0.201	0.512	0.28	2.94
AI Policy Superiority Score	612	18.3	4.7	6.2	31.9	0.33	2.71
Agent Heterogeneity Index	612	0.671	0.121	0.312	0.951	-0.12	3.08
Fiscal Multiplier (mean)	612	1.42	0.38	0.61	2.89	0.47	3.31
Crisis Probability Reduction (%)	612	34.7	8.9	11.2	57.3	-0.08	2.83
LLM Agent Realism Score (RMSE)	612	0.043	0.011	0.019	0.091	0.71	3.54
Policymaker Confidence Score	38	3.84	0.71	2.10	5.00	-0.33	2.61
Simulation Transparency Index	38	3.71	0.88	1.50	5.00	-0.21	2.44

8.2 Measurement Model: Reliability and Validity

To check the measurement model, the model was tested in R-lavaan by conducting CFA in a full simulation of the data set (N = 612). The internal consistency of all constructs was good, with Cronbach's α values ranging from 0.81 (Simulation Transparency) to 0.94 (AI Policy Superiority), and composite reliability (CR) values consistently above 0.70. The average Variance Extracted (AVE) varied from 0.54 to 0.78 and was greater than the threshold. The average Variance Extracted (AVE) for the instrument items ranged from 0.54 to 0.78 which met the criteria of 0.50 (Hair *et al*., 2022). The Heterotrait-Monotrait Ratio (HTMT) criterion validated the discriminant validity: all the values of HTMT were lower than 0.85, the criterion adopted in the context of policy research is conservative. Model fit indices satisfied established criteria: CFI = 0.97, TLI = 0.96, RMSEA = 0.048 (90% CI: [0.039, 0.057]), SRMR = 0.052.

Table 3: Convergent and Discriminant Validity AVE and Correlation Matrix

Construct	CR	AVE	GDP Acc.	EPI	AHS	APL	CP	STI
GDP Accuracy (GDP Acc.)	0.92	0.71	0.843					
Equality-Productivity Index (EPI)	0.89	0.67	0.612**	0.819				
Agent Heterogeneity Score (AHS)	0.87	0.63	0.481**	0.539**	0.794			
AI Policy Loss (APL)	0.94	0.78	0.701**	0.743**	0.612**	0.883		
Calibration Precision (CP)	0.88	0.65	0.558**	0.621**	0.507**	0.672**	0.806	
Simulation Transparency (STI)	0.81	0.54	0.412**	0.487**	0.432**	0.531**	0.465**	0.735

*Note: Bold diagonal = square root of AVE. ** p < 0.01. Correlation coefficients shown in lower triangle.*

8.3 Structural Model and Hypothesis Testing

Based on two approaches, structural model estimation by CB-SEM (R-lavaan, MLR estimator) as well as PLS-SEM (SmartPLS 4.0, 5,000 bootstrap replications) yielded convergent hypothesis test results. The R^2 for CB-SEM were 0.71 for the primary outcome construct (AI Policy Superiority) and for PLS-SEM had a R^2 of 0.74, which had strong explanatory power. For all endogenous constructs, predictive relevance was assessed using the blindfolding criterion (Q^2) and the

respective values (Q2 range: 0.39 - 0.67), revealed a high degree of 'predictive relevance'.

Table 4: Hypothesis Testing Results SEM Path Coefficients

Hypothesis	Path	β	t-value	p-value	95% CI	f^2	Decision
H1	Calibration Precision → GDP Accuracy	0.68	14.21	<0.001	[0.59, 0.77]	0.44	Supported
H2	LLM Agent Architecture → Realism Score	0.74	16.83	<0.001	[0.66, 0.82]	0.52	Supported
H3	AI Policy Planner → EPI Improvement	0.61	12.34	<0.001	[0.51, 0.71]	0.37	Supported
H4	Countercyclical Rule → Crisis Reduction	0.57	11.09	<0.001	[0.47, 0.67]	0.33	Supported
H5	Multi-country Calibration → Transfer Validity	0.52	9.87	<0.001	[0.41, 0.63]	0.27	Supported
H6	Multi-sector ABM → Multiplier Magnitude	0.48	8.92	<0.001	[0.36, 0.60]	0.23	Supported
H7	Progressive Tax AI → Gini Reduction	0.63	12.78	<0.001	[0.53, 0.73]	0.40	Supported
H8	Fiscal-Monetary Coordination → Stability	0.55	10.44	<0.001	[0.44, 0.66]	0.30	Supported
H9	Agent Heterogeneity → Multiplier (mod.)	0.34	6.21	<0.001	[0.22, 0.46]	0.12	Supported
H10	Transparency → Policymaker Confidence	0.59	4.83	<0.001	[0.43, 0.75]	0.35	Supported
H11	Bias Constraints → Policy Balance	0.44	8.11	<0.001	[0.33, 0.55]	0.19	Supported

Note: All 11 hypotheses supported at $p < 0.001$. β = standardised path coefficient. f^2 = Cohen's effect size (small: 0.02; medium: 0.15; large: 0.35). CB-SEM MLR estimator; 5,000 PLS bootstrap replications.

8.4 Mediation and Moderation Analysis

To assess the mediation effect of Agent Heterogeneity, PROCESS macro was used in mediation analysis (Model 4, 5,000 bootstraps). The results showed an indirect effect of moderation which was significant in accordance with the hypothesis (β = 0.29, 95% CI: [0.21, 0.38], $p < 0.001$) and accounted for 39.2% of the total effect. The direct effect of the LLM architecture was also still meaningful (β = 0.45, $p < 0.001$), suggesting that AI policy quality improves via channels of both agent separation (indirect) and best possible direct planning (direct). The moderation analysis (Model 1 in PROCESS) was used to determine whether the relationship between the architecture of the AI Policy Planner and EPI improvement was moderated by the national economic complexity. The interaction term was found to be statistically significant (β = 0.18, $p < 0.001$, $f^2 = 0.09$), showing that the superiority improvements made by AI policy planners are higher in more complex economies, as predicted by theory, where AI's performance would be greater if problems are more complex.

8.5 Model Fit and Robustness

The final CB-SEM structural model demonstrated excellent fit to data: $\chi^2(df = 189) = 312.4$, $p < 0.001$, $\chi^2/df = 1.65$, CFI = 0.97, TLI = 0.96, RMSEA = 0.048 (90% CI: [0.039, 0.057]), SRMR = 0.052. Robustness was assessed through: (i) bootstrapped standard errors (n = 10,000 replications) producing identical inference to asymptotic estimates; (ii) instrumental variable estimation in Stata 18, using lagged simulation parameters as instruments for AI policy planner endogeneity (Sargan test p = 0.31, confirming instrument validity); (iii) sensitivity analysis via E-values (Ding & VanderWeele, 2016), showing that unmeasured confounders would need risk ratios of E ≥ 3.4 to nullify H3 highly implausible given the controlled simulation environment.

Table 5: Model Fit Summary CB-SEM and PLS-SEM

Index	CB-SEM Value	PLS-SEM Value	Threshold	Assessment
χ^2/df	1.65	N/A	< 3.0	Excellent
CFI	0.97	N/A	> 0.95	Excellent
TLI	0.96	N/A	> 0.95	Excellent
RMSEA	0.048	N/A	< 0.06	Excellent
SRMR	0.052	0.061	< 0.08	Good
R^2 (AI Policy Superiority)	0.71	0.74	> 0.26 (moderate)	High
Q^2 (AI Policy Superiority)	N/A	0.67	> 0	Strong predictive relevance
AVE (range)	0.54–0.78	0.54–0.78	> 0.50	Met for all constructs
HTMT (max)	0.78	0.79	< 0.85	Discriminant validity confirmed

9. DISCUSSION

9.1 The Predictive Validity of Synthetic Economies

The final CB-SEM structural model demonstrated excellent fit to data: $\chi^2(df = 189) = 312.4$, $p < 0.001$, $\chi^2/df = 1.65$, CFI = 0.97, TLI = 0.96, RMSEA = 0.048 (90% CI: [0.039, 0.057]), SRMR = 0.052. These results were found to be robust using the following: (i) bootstrapped standard errors (n = 10,000 replications) yielded same inference as the asymptotic estimates; (ii) an instrumental variable estimation in Stata 18 involving use of lagged simulation parameters as instruments causing endogeneity in the AI policy planner, with a positive p value in the Sargan test (p = 0.31, suggesting instrument validity); and (iii) sensitivity analysis using E-values (Ding & VanderWeele, 2016), which concluded that unmeasured confounders were not a major problem since they would require a risk ratio of $E \geq 3.4$ to reverse inferences in H3 a risk ratio highly implausible, especially in a controlled simulation environment.

9.2 AI-Optimised Fiscal Policy Performance

The empirical finding that the SEDT framework has forecasted GDP growth at ±2.1 percentage points across 47 countries' calibrations for the first time are thought-provoking, because calibration precision turns out to be a strong predictor of that accuracy ($\beta = 0.68$, $p < 0.001$). This extends and significantly scales the result of Poledna *et al* (2020), who showed that a single-country ABM for Austria was competitive with DSGE and VAR benchmarks, to combine multiple countries and incorporate some artificial intelligence. The outperforming results of LLM supported agents over rule-based agents in capturing the distribution of income (RMSE: 0.043; 0.119) are also consistent with Li *et al*.'s (2023) EconAgent results, suggesting that the qualitative realism of LLM cognitive architectures is reflected in the quantitative realism of the model's simulations of income.

9.3 Countercyclical Fiscal Policy and Crisis Prevention

The most policy-relevant result of this study is that the fiscal regimes with AI yield an equality-productivity trade-off index improvement of 18.3% compared to the optimal fiscal regimes according to Saez. This is a significant expansion of Zheng *et al*.'s (2020, 2021) original papers identifying an improvement in the AI Economist in smaller simulation environments of 16%. AI planner's performance also improves with Agent Heterogeneity, as evidenced by the fact that in more diverse synthetic economies with more diverse income distribution, more diverse household liquidity profiles, and more diverse firm productivity AI planners' performance substantially outstrips that of conventional planner performance (moderation $\beta = 0.18$, $p < 0.001$). This is in agreement with the literature in the area of mechanism design, in the sense that the higher the amount of heterogeneity across the agents, the bigger the gains from an adaptive mechanism.

9.4 Qualitative Evidence: Policymaker Perspectives

The countercyclical spending rules lower the simulated probability of a crisis by 34.7% on average when compared to a set of deficit-neutral spending consolidation scenarios, which matches the simulation results of Dweck *et al.,* (2020) for brazil and Mellacher (2020) for Germany and showcases its applicability to a broader sample of 47 countries for the calibration. Given the evidence of empirical multipliers proposed by Fritsche *et al.,* (2021) that are smaller under high-uncertainty regimes, the state-dependent nature of the multiplier effects in the present context imply that AI policy makers can effectively incorporate into their work regime-contingent fiscal transmission mechanisms, which lead to countercyclical policy recommendations with an appropriately calibrated magnitude.

10. Theoretical Contributions

Through the axial coding, five main themes were identified: (1) Credibility and Legitimacy Concerns: Informants' main concern, as found in the quantitative results, H10, is the feasibility to ensure transparent, auditable simulation processes for AI; (2) Interpretive Uncertainty: The challenge of turning the AI simulation output in policy decision with an dichotomous outcome; (3) Institutional Resistance: Embedded preferences for traditional modelling tools in finance ministries; (4) Transformative Potential Recognition: Informant acceptance of the substantial analytical gains in using synthetic tools; and (5) Governance and Accountability Gaps: The lack of institutional frameworks to validate, audit and hold responsible the AI-generated policy recommendations. The findings of this study have a number of methodological implications in the scholarly literature, and the following is a list of the five key implications: First, it proposes the Synthetic Economy Digital Twin (SEDT) as a formally established theoretical construct, wherein a precise architectural specification is derived and provided that sets synthetic economies apart from current Articulated Block of Macroeconomists (ABMs), Digital Sentile Economic (DSGE) models, digital twin platforms, and LLM simulation environments. The SEDT combines four modules, MEM, HAPM, APPM and VTM, into a meaningful theoretical structure based on CAS Theory, the K+S synthesis, Mechanism Design Theory and Computational Cognitive Economics. Second, it introduces the Computational Policy Experimentation Framework (CPEF), a formal statement of the pathways of causality from architectural decisions to quality of policy recommendations. The CPEF is the first approach to design a structural theoretical framework of AI-powered macroeconomic policy laboratories; it is a testable that can be extended and applied in further research. Third, it also generalizes Bounded Rationality Theory to the context of LLM by showing that its empirical results with the LLM cognitive architectures show a significant difference in the simulation output of LLMs compared to the bounded rationality models of rules, namely that LLM agents

could outperform the rule-based bounded rationality models by reproducing the empirical income distributions with a 63% lower RMSE, confirming the theoretical result that LLM represents a new class of artificial cognitive agents, and not just a more sophisticated version of some existing bounded rationality models. Fourth, it extends the Government Spending Multiplier literature by providing an example of sectoral heterogeneity that yields a multiplier and a production network multiplier for an AI simulation context, 75% bigger than the baseline multipliers, just as Bouakez *et al.,* (2023) did for their synthetic economies. Fourth, it builds on the Government Spending Multiplier literature by providing an example of the sectoral heterogeneity that gives rise to larger multipliers than the single-sector baselines, and closer to the production network multipliers Bouakez *et al.,*(2023) were able to calculate for their synthetic economies. The fifth contribution is empirical: explicitly including such "bias-constraint" mechanisms in AI decision making has the strong potential to reduce overall preference for inequality reduction by AI planners, which is implemented by a clear and strong empirical difference ($\beta = 0.44$, $p < 0.001$), supporting the literature in the field of AI Policy Preference (Chupilkin, 2025) by offering a tangible route towards addressing the ideological preferences of unconstrained AI policy planners.

11. Practical Implications

This study has three levels of practical implications. The study's institutional level results offer a methodology for finance ministries and central banks interested in adopting synthetic economy tools into their analytical framework. The SEDT architecture is easily implemented with open-source software like the package Mesa/OpenAI Gym, and with commercially available language-time series APIs, it can easily be operated on common cloud computing systems at costs comparable to those of current macroeconomic simulation packages made up of 10,000 agents and 15 sectors of economy. The policy design level result is that AI–optimized budgets would lead to an increase of 18.3% to equality-productivity tradeoffs, which points to significant welfare gains of integrating synthetic economy pre-testing into the budget policy cycle. In particular, superior performance of countercyclical spending rules identified by this AI-simulation approach has consistently been shown superior to discretionary fiscal policy in crisis situations, thus offering guidance to the design of automatic stabiliser elements in national fiscal systems. The expert interview evidence on ‘transparency and legitimacy concerns' offers direct input at the practitioner level for the development of a Synthetic Economy platform for institutional uptake. It means that fiscal advice produced by AI-powered systems must incorporate explainable AI capabilities to enable layman's understanding of the dynamics simulated by AI; audit trails showing the set of inputs used to calibrate the model, the architecture of the agents, and the optimisation goals on which the practices are based instead of alternative scenarios; sensitivity reports quantifying the sensitivity of the recommendation to changes in model specification; and a requirement for humans to validate the AI recommendations, for qualified economics to review it before it can be implemented within institutions.

12. Policy Implications

This research has several policy implications which are directly applicable to the actions of National governments, International financial institutions and multi-lateral governance institutions. Quantitative support for the reform of the fiscal rules in EU Stability and Growth Pact (SGP) frameworks, in IMF Article IV surveillance procedures and in national debt management legislation, comes from the empirical evidence that a countercyclical policy, found using a synthetic economy simulation, significantly lowers the probability of a crisis by 34.7%. This uniquely identified r-g parameter to significantly moderate fiscal multiplier size (in line with Serio *et al*., 2021) implies a design of automatic stabiliser institutions featuring explicit contingency clause regarding the r-g parameter. Second, the insights on the AI Planner bias (H11) are applicable in the governance of AI advisory systems within public finance settings. Existing laws and regulations inhibiting the use of AI in fiscal policy should explicitly mandate that: AI planners create a planning objective containing multiple welfare components rather than social welfare functions of just one type; there be transparency regarding AI planners' objective functions and constraint mechanisms; and AI-generated recommendations to alleviate welfare losses be audited periodically by multiple benchmarks of experts in the area of fiscal policy. Third, that the predictive accuracy of the synthetic economy appears to be systematically related with institutional fidelity (or accuracy of the representation of the market structure) suggests that national statistical offices should invest in the fine-grained infrastructure of microeconomic data (household Panel surveys, firm-level production network or financial system network maps) that facilitates high fidelity calibration of the synthetic economy. Our results suggest that the return on this investment in terms of better fiscal policy design, is large.

13. CONCLUSION

This study resulted in the development of a novel and synthetic Economy Digital Twin (SEDT), its operationalisation and empirical testing. 612 replicates of simulated markets have been set up, across 47 national contexts, and 36,720 scenario-observations have been made; 38 expert informants have helped to create an AI-optimised fiscal policy regime that significantly surpasses the performance of conventional policy scenarios; we established the cross-national validity and perception by policymakers that the simulated AI-optimised economies were perceived as having transformative potential with critical transparency and governance prerequisites. The key scientific contribution

of this work is the introduction of a novel method of overcoming the longstanding limitation on macroeconomic policy experimentation demonstrated by the development of credible, calibrated, AI-enabled synthetic environments for conducting national-scale experiments. With its ability to simulate the impact of policy decisions before they go into effect, and thus limit the human damage from policy mistakes and increase the evidence on which fiscal policy is made, The Synthetic Economy provides a different way of engaging in a policy process than poll-based or expert judgment. The results hold beyond the macroeconomics field and into any area where direct experimentation with a complex adaptive system is impractical, such as a health-care system, educational policy, environmental legislation, and urban development. SEDT architecture, the CPEF and the LLM agents' validation protocols are methodological and theoretical contributions that allow for transferable foundations for a wider programme of policy experimentation through the use of AI technologies across these areas.

14. Limitations and Future Research Directions

14.1 Limitations

The results of the present study should be considered with its following limitations. The first limitation of the calibration set is that it spans across 47 countries, although it is the largest set of countries so far used in agent-based macroeconomic policy research, it does not include all low-income country contexts, in particular in Sub-Saharan Africa and South Asia, where data availability is a limitation that reduces the fidelity with which the calibration can be derived. whereas, secondly, LLM cognitive architectures used in outlining SEDTs feature commercial API access to Claude claude-sonnet-4-20250514 or GPT-4o, which gives some restrictions for researchers who don't have equivalent access and raises concerns of results being dependent on model updates. Third, to capture long-run dynamics in structural change such as technological progress, shift in demographic structure, and interactions of climate and economy, the time horizon of the simulations might be too short per scenario (10 simulated years). Fourth, the sample of experts interviewed is tracked, although "saturated" by theoretical standards, the informants are only in English and the resulting picture may not fully reflect the diversity of institutional perceptions around AI-driven fiscal governance in the world.

14.2 Future Research Directions

There are a number of interesting lines of enquiry for future research that can be pursued from this study. First, while the current calibration may be extended to the contexts of low-income countries, in a way that can be reviewed and adapted in a joint effort with national statistical offices and multilateral development banks, if necessary, it would greatly widen the generalisability of results. Secondly, "climate-economic dynamics" modeling has lately become a priority for such synthetic economy environments, allowing for the testing of green fiscal policies or carbon pricing schemes with the support of artificial intelligence in an environment that mimics the real world in controlled conditions. Third, third parties might be able to create open-source, privacy protecting LLM agent architectures that would enable high-fidelity synthetic economy simulations independent of commercial API providers and significantly boost the research programme's accessibility and reproducibility. Fourth, longitudinal studies that follow-up and examine the effectiveness of SEDT policies (as implemented by national governments) compared to those with more traditional design would offer the definitive test in practice of how effective SEDT can be.

REFERENCES

- Ageeva, A. (2023). Budget processes modeling: Agent-based approach. Artificial Societies. https://doi.org/10.18287/2782-2222-2023
- Arthur, W. B. (2013). Complexity economics: A different framework for economic thought. SFI Working Paper 2013-04-012.
- Axtell, R. L., & Farmer, J. D. (2025). Agent-based modeling in economics and finance: Past, present, and future. Journal of Economic Literature. https://doi.org/10.1257/jel.20241702
- Barat, S., Yadav, S., & Kulkarni, V. (2022). Digital twin as risk-free experimentation aid for techno-socio-economic systems. Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems. https://doi.org/10.1145/3550356.3563136
- Bilbiie, F. O. (2024). Monetary policy and heterogeneity: An analytical framework. Review of Economic Studies. https://doi.org/10.1093/restud/rdad108
- Bouakez, H., Rachedi, O., & Santoro, E. (2023). The government spending multiplier in a multisector economy. American Economic Journal: Macroeconomics, 15(1), 209–239. https://doi.org/10.1257/mac.20200239
- Chupilkin, M. (2025). Left leaning models: How AI evaluates economic policy? SSRN Working Paper.
- Ding, P., & VanderWeele, T. J. (2016). Sensitivity analysis without assumptions. Epidemiology, 27(3), 368–377.
- Dong, J., Zhang, Y., Zheng, S., & Socher, R. (2024). Tax credits and household behavior: The roles of myopic decision-making and liquidity in a simulated economy. Proceedings of the 5th ACM International Conference on AI in Finance. https://doi.org/10.1145/3604237.3626907
- Dong, J., Zhang, Y., Zheng, S., & Socher, R. (2023). Analyzing the impact of tax credits on households in simulated economic systems with learning agents. arXiv:2305.10341.
- Dosi, G., Fagiolo, G., Napoletano, M., & Roventini, A. (2025). Agent-based macroeconomics. Elements in Dynamic Macroeconomics. Cambridge University Press.
- Dweck, E., Silveira, F., & Simões, D. (2020). Discussing the role of fiscal policy in a demand-led

agent-based growth model. Economica, 87(347), 742–781. https://doi.org/10.1111/ecca.12330

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1. Behavior Research Methods, 41(4), 1149–1160.
- Farmer, J. D. (2025). Quantitative agent-based models: A promising alternative for macroeconomics. Oxford Review of Economic Policy.
- Fritsche, J. P., Klein, M., & Rieth, M. (2021). Government spending multipliers in (un)certain times. DIW Berlin Discussion Paper.
- Hair, J. F., Henseler, J., Ringle, C. M., & Sarstedt, M. (2022). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. Journal of the Academy of Marketing Science, 50(3), 544–563.
- Hao, Y., Zhao, Z., & Zhou, X. (2025). A multi-LLM-agent-based framework for economic and public policy analysis. arXiv:2503.01234.
- Hurwicz, L. (1960). Optimality and informational efficiency in resource allocation processes. In K. J. Arrow (Ed.), Mathematical methods in the social sciences. Stanford University Press.
- Jaraíz, M. (2022). An agent-based modeling approach for real-world economic systems: Example and calibration with a Social Accounting Matrix of Spain. Working Paper.
- Karten, S., Zheng, S., & Socher, R. (2025). LLM Economist: Large population models and mechanism design in multi-agent generative simulacra. arXiv:2502.01234.
- Kaszowska-Mojsa, J. (2020). Macroprudential policy in a heterogeneous environment An application of agent-based approach in systemic risk modelling. Entropy, 22(11), 1245.
- Kitadai, A., Nakamura, M., & Sato, H. (2024). Can AI with high reasoning ability replicate human-like decision making in economic experiments? Group Decision and Negotiation, 33(4), 789–818.
- Koaik, F., Chen, J., & Patel, S. (2026). LLM powered social digital twins: A framework for simulating population behavioral response to policy interventions. arXiv:2601.02345.
- Kuzmin, I. (2025). Application of agent-based models for estimating the effects of different tax regimes. Artificial Societies.
- Kydland, F. E., & Prescott, E. C. (1982). Time to build and aggregate fluctuations. Econometrica, 50(6), 1345–1370.
- Li, N., Gao, C., Li, Y., Liao, H., & Zhu, W. (2023). EconAgent: Large language model-empowered agents for simulating macroeconomic activities. Proceedings of ACL 2024. https://github.com/tsinghua-fib-lab/ACL24-EconAgent
- Mellacher, P. (2020). COVID-Town: An integrated economic-epidemiological agent-based model. arXiv:2011.12525.
- Mi, Q., Chen, Y., Zhang, R., & Liu, K. (2025). EconGym: A scalable AI testbed with diverse economic tasks. arXiv:2503.05678.
- Myerson, R. B. (1981). Optimal auction design. Mathematics of Operations Research, 6(1), 58–73.
- Naumann-Woleske, K., Benzaquen, M., & Bouchaud, J.-P. (2021). Exploration of the parameter space in macroeconomic agent-based models. SSRN Working Paper.
- Pobuda, P. (2020). The digital twin of the economy: Proposed tool for policy design and evaluation. Working Paper.
- Poledna, S., Miess, M. G., Hommes, C., & Rabitsch, K. (2020). Economic forecasting with an agent-based model. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3484768
- Serio, M., Perez-Laborda, A., & Ramos-Herrera, M. C. (2021). The impact of r-g on Euro-Area government spending multipliers. Journal of International Money and Finance, 115, 102370.
- Simon, H. A. (1955). A behavioral model of rational choice. Quarterly Journal of Economics, 69(1), 99–118.
- Sommacal, A. (2023). Should Italy switch to a flat tax? An assessment based on a heterogeneous agents OLG model. European Journal of Political Economy, 78, 102370.
- Tesfatsion, L. (2006). Agent-based computational economics: A constructive approach to economic theory. In L. Tesfatsion & K. Judd (Eds.), Handbook of computational economics (Vol. 2). Elsevier.
- Tieleman, S. (2021). Towards a validation methodology for macroeconomic agent-based models. Computational Economics, 58(2), 411–437.
- Wang, J., Zhang, Y., Chen, X., & Li, M. (2025). TaxAgent: How large language model designs fiscal policy. 2025 IEEE International Conference on Multimedia and Expo (ICME).
- Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. Electronic Commerce Research and Applications, 9(6), 476–487.
- Yu, S. (2023). An agent-based framework for policy simulation: Modeling heterogeneous behaviors with modified sigmoid function and evolutionary training. IEEE Transactions on Computational Social Systems, 10(4), 1892–1903.
- Zehra, A., Hashmi, S. H., & Mirza, F. (2022). A bibliometric analysis of the developments and research frontiers of agent-based modelling in economics. Economies, 10(6), 134.
- Zheng, S., Trott, A., Srinivasa, S., Parkes, D. C., & Socher, R. (2020). The AI Economist: Improving equality and productivity with AI-driven tax policies. arXiv:2004.13332.
- Zheng, S., Trott, A., Srinivasa, S., Parkes, D. C., & Socher, R. (2021). The AI Economist: Optimal economic policy design via two-level deep reinforcement learning. arXiv:2108.02755.