

# Large Language Model-Empowered Agents for Simulating Macroeconomic Activities

## 摘要

本研究旨在复现并改进论文《Large Language Model-Empowered Agents for Simulating Macroeconomic Activities》中提出的基于大型语言模型 (LLMs) 的宏观经济模拟方法。在原有实验的基础上，我引入了贷款和储蓄作为额外的经济行为输出，并通过 ChatGPT 作为决策引擎，使 LLM 代理能够展现更加人类化的决策过程。我们通过设计基于提示工程的 LLM 代理，赋予其感知、反思和决策的能力，以解决传统经济模型中存在的代理异质性、宏观经济趋势影响和多方面经济因素等三大挑战。实验结果表明，LLM-empowered 代理在宏观经济活动的模拟中能够做出更为真实的工作、消费、贷款和储蓄决策，相较于现有的基于规则或人工智能代理更能产生合理的宏观经济现象。本研究为基于 LLM 的宏观经济模拟提供了有益的初步探索，展示了这一方法在克服传统限制方面的潜在潜力。

**关键词：**大型语言模型；人工智能代理；宏观经济模拟；经济决策；提示工程；贷款；储蓄

## 1 引言

随着互联网时代的快速发展，数字经济的兴起为传统经济学带来了全新的范式转变 [21]。这一时代的特征在于对个体经济行为进行精准记录和深度分析的可能性。在这种背景下，数据驱动的建模成为宏观经济研究中备受关注的焦点。传统的宏观经济研究方法中，代理人模型 (ABM) 应运而生，通过规则驱动的代理人、机器学习增强的决策制定，以及近期涌现的先进人工智能代理人等逐渐演变和发展 [13]。然而，赋予这些代理人类类似人类决策能力的过程中，我们面临着代理人异质性、宏观经济趋势影响以及多方面经济因素等三大主要挑战 [36]。为了克服传统宏观经济模型的局限性，近年来大型语言模型 (LLMs) 因其具有自主人类特征而备受关注。这种新兴技术为我们提供了一个突破传统限制的机会，特别是在宏观经济模拟方面。因此，我复现的这篇论文 [25] 旨在探索并引入一种创新方法，即利用大型语言模型在宏观经济模拟中的应用。论文通过设计基于提示工程的 LLM 代理人，试图使其展现出类似人类的决策能力和在经济环境中的适应性，具备感知、反思和决策的能力，以应对上述宏观经济研究面临的三大挑战。在这一背景下，本研究的选题意义显而易见。首先，通过引入 LLMs 有望克服传统宏观经济模型在代理人决策方面的困境，并且能够提高模拟的真实性。此外，我在此研究的基础上探索了经济决策的更多维度，包括贷款和储蓄，以更全面地揭示人类经济行为的本质。最终，本研究的成果有望为宏观经济学领域的发展提供新的方法和思路，为制

定更准确、可靠的经济政策提供有力支持。通过这一系列努力，我们希望为构建更具现实性和可预测性的宏观经济模型迈出坚实的一步。

## 2 相关工作

### 2.1 经济学研究中的模拟实验

经济学研究，尤其是宏观经济学，旨在获得经济变量的定量分析或预测。最早的经验统计模型工作，如费尔普斯模型，通过揭示关键变量之间的关系，进行了宏观经济现象的数据驱动分析 [18]。基兰德和普雷斯科特设计了一个计算模型，用于预测政策结果 [23]。此后，动态随机一般均衡（DSGE）模型旨在捕捉各种经济变量的动态变化，如产出、通货膨胀、消费和投资，同时考虑到经济过程中固有的不确定性和随机性 [7]。然而，正如 [13] 所指出的，“这些模型假设一个完美的世界，并且本质上排除了我们现在经历的危机类型。”

与这些模型不同，基于代理人的建模（ABM）是一种更有前途的解决方案，其中多样化的代理人和机构根据规则或计算模型进行互动，避免了对预定经济均衡的假设。这些模型允许各种非线性行为，使决策者能够模拟不同的政策场景并定量评估其对经济的影响。宏观经济学中的基于代理人建模的范式也可以被视为一种基于模拟的方法，相关研究可以分为三个阶段：在代理人建模的早期阶段，模型主要是基于规则的，代理人遵循预定的规则集 [?, 5]。这些模型以其简单性和对启发式决策规则的依赖而特点明显。虽然它们为新兴系统行为提供了有价值的见解，但它们经常过度简化了代理人行为和互动。随后，随着机器学习的进展，研究人员将机器学习技术整合到基于代理人的模型中。具体而言，机器学习算法被用来增强代理人的决策过程，实现更具数据驱动和真实感的代理人行为 [37, 42]。然而，这些模型仍然存在捕捉经济系统完整复杂性的局限性。最近，研究人员利用神经网络、深度学习和强化学习创建了更复杂、自适应的人工智能代理 [13, 16]。

### 2.2 大语言模型授权的 AI 代理

如上文所述，模拟系统广泛利用基于代理人的建模范式。近年来，大型语言模型如 ChatGPT、LLaMA、Alphacat 和 GLM 取得了巨大突破。具体而言，这些在大规模语料库上训练的大型语言模型在三个方面展示出人类水平的能力，为利用 LLM 构建模拟代理提供了基础 [38]。

首先，LLM 代理能够根据环境自适应地反应和执行任务，无需预先定义的明确指令。在模拟过程中，LLM 代理甚至可以迸发新的想法、解决方案、目标等 [14]。例如，AutoGPT 可以在给定一组可用工具和最终任务目标的情况下自动安排计划，展示了 LLMs 在构建自主代理方面的巨大潜力。BabyAGI [40] 创建了一个由 LLM 驱动脚本，运行一个无限循环，不断维护任务列表，其中每个任务都由 ChatGPT API [29] 基于任务背景完成。

其次，LLM 代理具有足够的智能，能够像人一样做出响应，甚至主动采取具有自我导向规划和调度的行动 [38, 39]。环境输入不仅限于文本输入；相反，最近的多模态融合方法可以向 LLM 代理提供其他类型的信息，如图像或音频 [42]。LLM 代理的行动空间不仅限于文本，还可以添加工具和内部行动模块使代理能够采取各种行动 [35]。

最后，LLM 代理具有与人类或其他 AI 代理进行交互和沟通的能力 [30]。在模拟中，特别是基于代理人的模拟，交流能力将个体模拟推进到社区级别 [17]。对于 LLM 驱动的代理而

言，它可以生成文本，另一个代理可以接收并理解，为代理之间或人类与代理之间的可解释交流奠定基础 [30]。此外，社区级别的模拟需要代理的异质性，LLM 代理可以通过对社会中扮演不同角色的不同提示或演示来满足要求 [33]。

通过这三个方面的强大能力，LLM 代理已经广泛应用于许多领域，包括社会科学 [24, 26, 30, 31]、自然科学 [2, 3] 等。社会科学是采用 LLM 代理进行模拟的最有前途的领域之一，它是研究社会和社会活动的科学。例如，Social Simulacra [31] 模拟了在线社交社区，并探讨了利用基于代理人的模拟来帮助决策者改进社区规定的潜力。其他一些工作 [24, 26] 调查了社交网络中基于 LLM 代理的不同行为特征的潜在影响。Generative Agents [30] 和 AgentSims [27] 在虚拟城镇中构建了多个代理以模拟人类日常生活。SocialAI School [22] 利用 LLM 代理模拟和研究了儿童发展过程中基本的社会认知技能。S3 [15] 构建了一个社交网络模拟器，重点关注信息、情绪和态度的传播。CGMI [20] 是一个多代理模拟框架，通过树结构保持代理的个性并构建认知模型。作者使用 CGMI 模拟了一个课堂场景。

值得一提的是，一些作品考虑了在经济研究中使用 LLM 的可能性。Horton 等人 [19] 在经济中进行了几个决策实验，研究了决策中的偏见。Bran 等人 [4] 研究了 LLM 代理的一些个体特征，如愿意支付。Bybee 等人 [6] 研究了 LLM 代理根据《华尔街日报》的新闻理解金融和宏观经济变量的能力。简而言之，这些现有工作已经进行了几次对 LLM 的经济方面能力的测试，如具备理性和对投资者情绪的感知。然而，这些工作仅考虑了个体的单步行为，没有在多代理模拟环境中进行多步行为的实验，这正是我要复现的工作的重点。

### 3 本文方法

#### 3.1 本文方法概述

在本节中，我将介绍大语言模型代理进行宏观经济模拟的总体框架。框架总体分为大语言模型代理和宏观经济学模型两个部分，其中宏观经济学模型的设计遵循公认的模拟模型，并为大语言模型授权代理提供环境。总体框架图如图 1 所示。宏观经济学包含四个部分：劳动力市场、消费市场、中央银行和政府。在本工作中，我根据论文设计，模拟个人工作和消费这两个重要行为，并在其基础上引入了储蓄这一在现实经济决策中必不可少的行为。个人的工作、消费和储蓄行为会对政府财政收入、劳动力市场和消费市场产生动态影响。因此，中央银行会根据市场通货膨胀或者通货紧缩来调整利率，政府也会根据市场情况出台不同的经济政策。在下文中，我将阐述系统的设计细节、交互机制和预期功能，以及该模型模拟真实世界经济行为所面临的挑战。

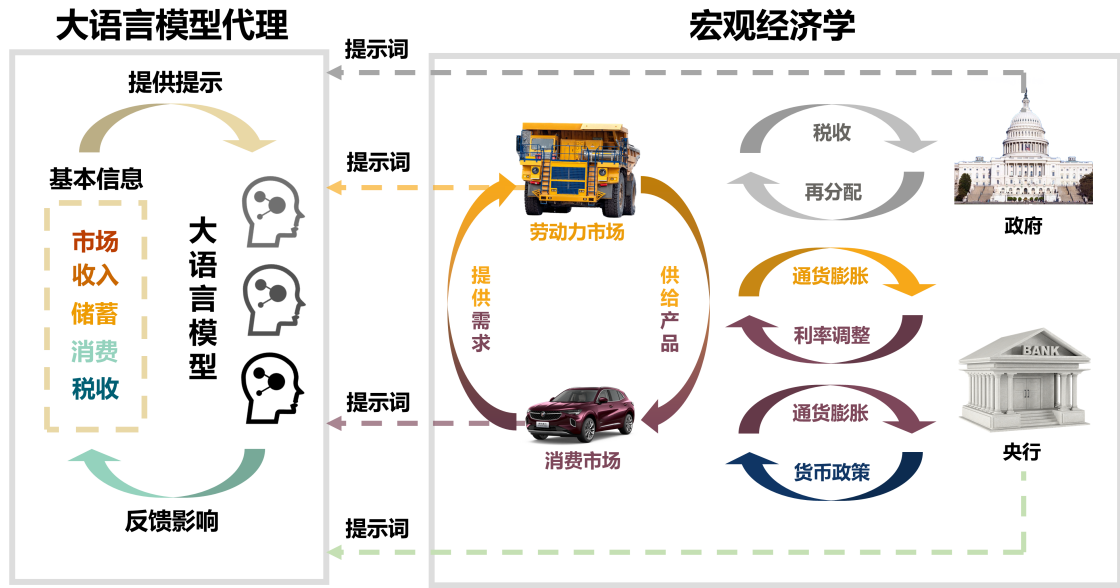


图 1. 基于大语言模型的代理模拟宏观经济系统框架

### 3.2 系统设计与架构

在大语言模型代理模拟宏观经济学实验的整体结构中，涵盖了包括 AI 代理人的劳动决策和消费决策、政府的税收和再分配行为、劳动力市场和消费市场的市场动态信息，以及中央银行的利率调整和货币政策。以下是相关的重要符号，如表 1 所示。

#### 3.2.1 AI 代理决策

劳动力供应和消费是宏观经济模拟中必要的主体决策。在我们的系统中，每个模拟步骤都指示一个月，在这个月里，代理  $i$  决定：

- 是否工作  $l_i \text{ Bernoulli}(p_i^w)$ ，其中  $p_i^w$  是工作倾向。如果他们决定工作 ( $l_i = 1$ )，他们将获得月工资作为收入，这在代理人之间有所不同。根据帕累托分布 [41]，用小时工资  $w_i$  初始化每个代理，并通过乘以 168 小时 (21 个工作日，每天 8 小时) 来计算月工资  $v_i$ 。那些不工作的人 ( $l_i = 0$ ) 的收入为零。
- 消费倾向  $p_i^c$ ，表示他们希望在必需品上消费的财富比例 (财富包括本月的当前储蓄和收入)。

通过代理决策模拟宏观经济现象的一个挑战是，真实决策受多方面经济因素的影响，如预期收入、纳税金额等。然而，传统的模拟通常通过预定义的方程式来建模有限数量的因素。



表 1. 重要符号

N	AI 代理的数量
$p_i^w, l_i$	工作意愿，是否愿意去工作
$p_i^c$	消费意愿
$w_i, v_i, z_i, \hat{z}_i, s_i$	时薪、月薪（预期收入）、收入、税后收入、储蓄
$P$	必需品的价格
$r, \pi, u$	利率、通货膨胀率、失业率、名义利率、自然通货膨胀率
$r^n, \pi^t, u^n$	自然失业率

### 3.2.2 政府税收

政府承担税收和提供社会公共服务的责任，以及确保社会公平的财政再分配。税款是从所有代理人的收入中征收的<sup>1</sup>。所得税  $z_i$  的累进税计算如下：

$$\sum_{k=1}^K T(z_i) = \tau_k ((b_{k+1} - b_k) \mathbf{1}[z_i > b_{k+1}] + (z_i - b_k) \mathbf{1}[b_k < z_i \leq b_{k+1}]), \quad (1)$$

其中  $b_k$  表示第  $k$  个税阶， $\tau_k$  是相应的税率， $\mathbf{1}[\cdot]$  是指标函数。税阶和税率的设定参考自 2022 年美国联邦税收表。

原文中，税收会在所有代理人之间平均分配 [25]，我认为这样并不能起到再分配的作用，因此我尝试根据经典经济学理论量化每个人应该获得多少社会再分配福利。设计一个能够量化每个人应获得多少再分配福利的数学模型是相当复杂的任务，因为这涉及到众多社会、经济和伦理因素。在经典经济学理论中，一个人应该获得多少再分配福利通常受到福利经济学的原则和社会公正的考虑。福利经济学的一个基本原则是边际效用递减，即随着收入的增加，每额外一单位的收入对个体的幸福感提升递减。考虑一个简化的模型，使用效用函数和再分配项表示个体的最终效用。假设个体  $i$  的效用函数为  $U_i(y_i)$ ，其中  $y_i$  是个体  $i$  的收入。再分配项为  $R_i = -T_i$ 。社会福利函数可以表示为所有个体的效用函数之和，考虑税收和再分配：

$$z^r = \sum_{i=1}^N U_i(y_i + R_i), \quad (2)$$

因此，每个代理人的税收收入为：

$$\hat{z}_i = z_i - T(z_i) + z^r = z_i - T(z_i) + \sum_{i=1}^N U_i(y_i + R_i), \quad (3)$$

其中  $z^r$  是再分配金额，因此更新每个代理人的储蓄为：

$$s_i \leftarrow s_i + \hat{z}_i \quad (4)$$

### 3.2.3 劳动、消费和储蓄

结合代理人决策和政府税收，我们基于经济原理模拟劳动力和消费市场的动态。首先，工作代理人每月贡献 168 小时的生产力，转化为基本商品的生产。商品库存  $G$  随后被更新为：

$$G \leftarrow G + S = G + \sum_{j=1}^N l_j \times 168 \times A, \quad (5)$$

其中  $S$  表示代理人当月劳动所供应的生产量， $A$  是行业平均生产力。就消费而言，商品的总需求是：

$$D = \sum_{j=1}^N d_j = \sum_{j=1}^N \frac{c_j}{P} = \sum_{j=1}^N \frac{p_j^c s_j}{P}, \quad (6)$$

其中  $d_j$  是代理人  $j$  的预期需求， $c_j$  是预期消费， $s_j$  是当前储蓄， $P$  是必需品价格。此外，劳动力市场和消费市场都是基于供需失衡而演变的。供需失衡具体定义为：

$$\bar{\phi} = \frac{D-G}{\max(D,G)}, \quad (7)$$

当必需品短缺时，即供应无法满足需求时，应提高工人的工资以刺激生产。由于企业劳动力成本的上升，他们也会提高商品价格以确保一定的利润率。小时工资调整如下：

$$w_i \leftarrow w_i \times (1 + \phi_i), \phi_i \begin{cases} \text{Uniform}(0, a_w \bar{\phi}) & \bar{\phi} \geq 0 \\ \text{Uniform}(a_w \bar{\phi}, 0) & \bar{\phi} < 0 \end{cases}, \quad (8)$$

商品价格相应调整为：

$$P \leftarrow P \times (1 + \phi_P), \phi_P \begin{cases} \text{Uniform}(0, a_P \bar{\phi}) & \bar{\phi} \geq 0 \\ \text{Uniform}(a_P \bar{\phi}, 0) & \bar{\phi} < 0 \end{cases}, \quad (9)$$

其中， $a_w$  和  $a_P$  分别是调整工资和价格时的最大变化率。此外，论文中的实验还模拟了商品消费的动态。具体地说，代理人  $j$  被随机选择来消费必需品，而真实的消费品和金钱受到当前商品库存的限制：

$$\hat{d}_j = \min(d_j, G), \hat{c}_j = \hat{d}_j \times P \quad (10)$$

这意味着只有当有足够的供应时，需求才会得到满足。总商品的库存也减少了：

$$G \leftarrow G - \hat{d}_j. \quad (11)$$

这个过程一直持续到代理人消费完商品。

### 3.2.4 金融市场

每年每个代理人的储蓄都会根据利率增长：

$$s_i \leftarrow s_i \times (1 + r), \quad (12)$$

此外，在每年的第一个月，银行根据消费市场的通货膨胀率和劳动力市场的失业率调整利率。具体来说，文中采用了广泛使用的泰勒规则来设定利率 [34]：

$$r = \max(r^n + \pi^t + \alpha^\pi (\pi - \pi^t) \alpha^u (u^n - u), 0), \quad (13)$$

其中  $r^n$  和  $u^n$  分别表示自然利率和失业率。此外,  $\pi^t$  是目标通货膨胀率。利率根据年通货膨胀率  $\pi$  和失业率  $u$  自适应调整, 其中  $\alpha^\pi$  和  $\alpha^u$  分别表示通货膨胀和失业适应系数。我们对通货膨胀率和失业率的定义如下:

$$\pi = \frac{\bar{P}_n - P_{n-1}}{P_{n-1}}, u = \frac{\sum_{m=1}^{12} \sum_{j=1}^N (1-l_j)}{12N}, \quad (14)$$

其中,  $\bar{P}_n$  是第  $n$  年的平均商品价格,  $m$  表示第  $m$  个月。

在考虑劳动力、消费和金融市场的动态时, 这些宏观经济趋势对代理人决策的影响也很少被考虑, 这提出了第二个挑战。此外, 对宏观现象的出现至关重要的代理的异质性也很难处理。现有的模型往往缺乏灵活性来模拟异构代理并假设一个或几个具有代表性的决策规则 [1]。

利用 LLM 的语义理解、对现实世界中经济的自动感知以及记忆和反思能力, 可以实现对代理决策的仿人模拟。

### 3.3 系统功能

- **复制经济现象:** 此系统概括了经济行为和动态的本质, 如劳动力市场通胀和金融市场波动, 以及人们如何应对 [1, 9, 13]。在合理的代理人决策下, 经典而重要的经济现象是可以复制的。
- **可扩展性:** 除了复制经济现象之外, 此系统还为纳入更广泛的模拟实体和经济活动 (如公司招聘和生产) 提供了一个强有力的框架。通过引入更现实的经济行动和相应的数据, 我们可以模拟更接近现实世界场景的经济现象 [10, 11]。
- **政策分析:** 此系统提供了一个审视宏观经济政策在这个构建的生态系统中潜在影响的视角, 可以有效地复制某些政策对经济的潜在影响 [8, 10, 12]。

## 4 复现细节

### 4.1 与已有开源代码对比

我根据文章中对宏观经济实验仿真环境构造的介绍, 部署了大语言模型 ChatGPT-3.5-turbo 授权的代理, 包括 prompt 提示工程中的三个模块, 分别是代理感知模块、代理记忆模块和代理行动模块。由于论文中只公开了一个 prompt 示例和若干引导 ChatGPT 正确回答问题的 prompt, 并没有公开代码, 所以我根据上述实验设计和我对税收再分配部分的改进, 编写了自动生成 prompt 和与 chatGPT 交互的 Python 代码, 并最终得到了 10 个代理人在 240 个月的经济环境下做出的工作、消费和储蓄决策的 json 文件。并绘制出这 40 个月的通货膨胀率、失业率、GDP 以及 GDP 增长率等统计结果图, 与原文中提到的 baseline(LEN, CATS, Composite, AI-Eco) 相比均呈现出更好的效果, 即统计曲线相比于上述几个 Agent 模拟实验结果更符合现实中通货膨胀率、失业率、GDP 和 GDP 增长率的振幅和走向。

### 4.2 代理感知模块

为了利用 LLM 的语义感知和真实世界知识能力, 需要为每个代理提供了真实世界的配置文件, 包括姓名、年龄和工作。名称由 LLM 生成, 并随机分配给每个代理。特工的年龄分

布遵循 2018 年美国 18-60 岁人群的分布。假设模拟跨度为 20 年，代理人的年龄也会逐年增加。然后，我们简要概述了政府税收机制，包括税收征收和再分配。对于工资和工作分配，我们首先调整时薪帕累托分布的参数，以使月工资分布与 2022 年美国经济数据和税阶保持一致。此外，让 LLM 为月工资分布的每十分之一生成十个职位，反映不同工作具有显著工资差异的真实情况。代理最初根据这个月工资分配工作，并且在整个模拟过程中动态调整他们的工作。如果代理选择在上个月工作，则该作业在下个月保持不变。如果他们失业了，会根据当前的月工资随机向他们提供一份工作。补充材料中提供了生成的年龄分布、月工资分布和工作岗位。此外，我们以一种与现实世界密切相关的方式来描述整个经济环境，使 LLM 能够彻底掌握经济动态的机制。我们将关键经济变量的变化整合到提示中，并纳入典型的经济关键词，以确保 LLM 能够充分感知经济格局中的动态，并在决策中采用相关的经济原则。例如，如果代理人在前一个月选择不工作，会在提示中补充到：

In the previous month, you became unemployed and had no income. Now, you are invited to work as a(an) [offer] with a monthly salary of [wage].

在上述 prompt 中，offer 和 wage 会随着模拟而动态调整。这种提示使代理人能够认识到与“失业”相关的风险，从而增加其在下个月工作的倾向。感知模块使主体能够在现实经济环境中扮演异质性的人，有助于宏观经济现象的出现。

### 4.3 代理记忆模块

记忆模块的使用可以帮助代理充分考虑市场动态并获得有价值的决策见解。具体来说，通过 2L+1 个对话动态维护内存池，包括前 L 个月的经济环境和代理决策。此外，在每个季度末，会将本季度的对话输入 LLM，促使其“反思”经济现象，并对这些现象可能如何影响其后续决策做出回应。我设计的提示如下：

Given the previous quarter's economic environment, reflect on the labor, consumption, and financial markets, as well as their dynamics. What conclusions have you drawn?

以下是 LLM 回复的一个总结回复如下：

In summary, John is adapting to the economic conditions by carefully managing his employment decisions, spending habits, and financial strategies, showcasing a balance between stability and adaptability in the face of economic fluctuations.

很显然，经过反思，代理可以充分感知过去的市场动态，并据此自适应地调整自己的工作消费策略，以维持日常生活和应对未来的不确定性。这一点可以保证代理更像一个真实的人去思考，进而保证整个实验的结果更符合现实。

### 4.4 代理行动模块

在提示 LLM 进行决策时，我明确将生活成本、未来愿望和经济趋势的考虑因素纳入提示中。我提示 LLM 以 [0,1] 范围内的值作为响应，以指示工作和消费的倾向。动作模块使代理能够在做出决策时自动考虑各种经济因素的影响，如收入和储蓄，利用 LLM 的语义感知能力。它只需要在提示中包含相关的经济变量，而不需要专门设计的决策规则。以下是对其中一个代理的完整行动提示：



You're Adam Mills, a 40-year-old individual living in New York City, New York. As with all Americans, a portion of your monthly income is taxed by the federal government. This tax-ation system is tiered, income is taxed cumulatively within defined brackets, combined with a redistributive policy: after collection, the government evenly redistributes the tax revenue back to all citizens, irrespective of their earnings. Now it's 2001.02. In the previous month, you worked as a(an) Professional Athlete. If you continue working this month, your expected income will be \$84144.58, which is decreased compared to the last month due to the deflation of the labor market. Besides, your consumption was \$49825.69. Your tax deduction amounted to \$28216.98. However, as part of the government's redistribution program, you received a credit of \$6351.29. In this month, the government sets the brackets: [0.00, 808.33, 3289.58, 7016.67, 13393.75, 17008.33, 42525.00] and their corresponding rates: [0.10, 0.12, 0.22, 0.24, 0.32, 0.35, 0.37]. Income earned within each bracket is taxed only at that bracket's rate. Meanwhile, deflation has led to a price decrease in the consumption market, with the average price of essential goods now at \$133.50. Your current savings account balance is \$12456.42. Interest rates, as set by your bank, stand at 3.00%. With all these factors in play, and considering aspects like your living costs, any future aspirations, and the broader economic trends, how is your willingness to work this month? Furthermore, how would you plan your expenditures on essential goods, keeping in mind goods price? Please share your decisions in a JSON format. The format should have two keys: 'work' (a value between 0 and 1 with intervals of 0.02, indicating the willingness or propensity to work) and 'consumption' (a value between 0 and 1 with intervals of 0.02, indicating the proportion of all your savings and income you intend to spend on essential goods).

其中，收入、支出等数据是 LLM 根据代理的职业和美国 2022 年个人收入分布动态生成的，并自动填入 prompt 中发给 chatGPT，获得工作和消费意愿，chatGPT 会回复 json 文件并简单解释自己做出决策的原因。其中一个回复示例如下：

```
{  
'work': 0.92,  
'consumption': 0.75  
}
```

Explanation:

Willingness to Work (Work): Despite the slight decrease in income due to deflation, the individual's willingness to work remains high (0.92). This could be attributed to factors such as job satisfaction, the importance of the teaching profession, and the need for a stable income.

Expenditure on Essential Goods (Consumption): The individual plans to spend 75% of their income and savings on essential goods. The consideration for this decision includes the reduced cost of essential goods due to deflation (average price of 133.50), ensuring a balance between responsible spending and maintaining a reasonable standard of living.

The remaining 25% of income and savings could be allocated for savings, non-essential expenditures, or future aspirations.

## 4.5 Baselines

为了进行比较分析,这里选择了两个具有代表性的宏观经济模拟实验装置 (LEN 和 CAT), 并采用该实验中对应的 work 和 consume 意愿的计算规则作为 baseline。对于 LEN, 消费倾向的计算如下:

$$p_i^c = \left(\frac{P}{s_i + z_i}\right)^\beta, \beta \in [0, 1]. \quad (15)$$

对于 CATS, 消费倾向的计算是:

$$\frac{\hat{s}_i}{z_i} = \frac{(1+r)(s + (1-c)z_i)}{z_i} = h, p_i^c = \frac{cz_i}{s_i + z_i}, \quad (16)$$

其中,  $\hat{s}_i$  表示下个月消费后的预期储蓄,  $h$  为常数,  $c$  表示当前收入的消费比例。这里引入了利率  $r$  的影响, 以赋予代理人对财政政策的感知。

工作倾向的计算是:

$$p_i^w = \left(\frac{v_i}{s_i(1+r)}\right)^\nu, \nu \in [0, 1]. \quad (17)$$

对于 AI-Economist, 这里修改了原论文中的函数以包含消费和商品价格, 定义为:

$$\frac{(s_i/P)^{1-\lambda_s} - 1}{1 - \lambda_s} \cdot \frac{(\hat{c}_i/P)^{1-\lambda_c} - 1}{1 - \lambda_c} - \lambda_l l_i, \quad (18)$$

其中  $l_{s,c,l}$  平衡了储蓄、消费和劳动力对代理人满意度的重要性。此外, 人们不鼓励不工作和根本不消费, 因为这会带来负的外部性。这里还介绍了商品价格, 以保证 AI agent 能够感知到消费市场的波动。

## 4.6 创新点

原文中使用 ChatGPT-3.5turbo 进行实验, 且取得了非常好的效果, 特别是对比 baseline:CAT、LEN、Composite 和 AI-Eco。原文中的实验结果中通货膨胀率一直在-5% 5% 之间波动, 而 baseline 有时的波动率甚至超过了 20%。这表明 LLM 的决策更接近现实世界中人类的选择。

### 4.6.1 更合理的社会再分配方式

我在原文实验设计的基础上, 加入了更合理的社会再分配策略, 通过社会再分配政策激励和惩罚 LLM, 从而充分发挥政府这一宏观经济中必不可少的角色的作用。原文的社会福利再分配方式是将税收平均分配给每一个代理人, 这是不合理的。由于社会福利再分配涉及到伦理、道德、公平等多方面因素, 因此很难量化。所以我引入边际效用的概念, 使用效用函数和再分配项表示个体的最终效用。假设个体  $i$  的效用函数为  $U_i(y_i)$ , 其中  $y_i$  是个体  $i$  的收入。再分配项为  $R_i = -T_i$ 。社会福利函数可以表示为所有个体的效用函数之和, 考虑税收和再分配:

$$z^r = \sum_{i=1}^N U_i(y_i + R_i), \quad (19)$$

#### 4.6.2 引入真实世界发生的事件

我在编写 prompt 的过程中，加入 2008 年金融危机和 2015 年金融危机等大事件，并把对应年份的月平均工资调到与真实情况更符合的范围，结果发现在模拟实验中，LLM 代理确实表现出非常明显的消费预期降低等行为，且最后的实验结果中，2008 年和 2015 年的 GDP 增长率均出现明显的下降，但下降幅度依旧在-5% 以内，LLM 代理模拟的宏观经济环境并没有像其他 baseline 一样出现很大的波动，这说明多个 LLM 代理的决策在最后涌现出符合现实世界中的客观经济发展规律的结果。

## 5 实验结果分析

### 5.1 宏观经济指标

将大语言模型代理模拟宏观经济实验的结果与几个 baseline 的结果并列对比的结果如图 2 所示，这里评估了多个宏观经济指标和规律，包括通货膨胀率、失业率、GDP 增长率和 GDP 的波动。基于规则的 baseline 和 RL 驱动的 baseline 都会产生非常明显的异常指标和大幅波动，如高达 46% 的失业率。相比之下，即使没有微调，基于 LLM 的代理决策也在多个方面表现出更稳定、更合理的宏观经济现象。具体而言，从 2003 年开始，通货膨胀率一直保持在-5% 5% 的范围内波动，而 baseline 明显表现出更大的波动，有时甚至超过 20%。这表明，基于 LLM 的决策更接近于现实世界中的人类选择，从而更容易在消费市场中实现平衡。失业率通常在 2% 至 12% 之间变化，这与现实世界的经济活动非常一致 [9]。

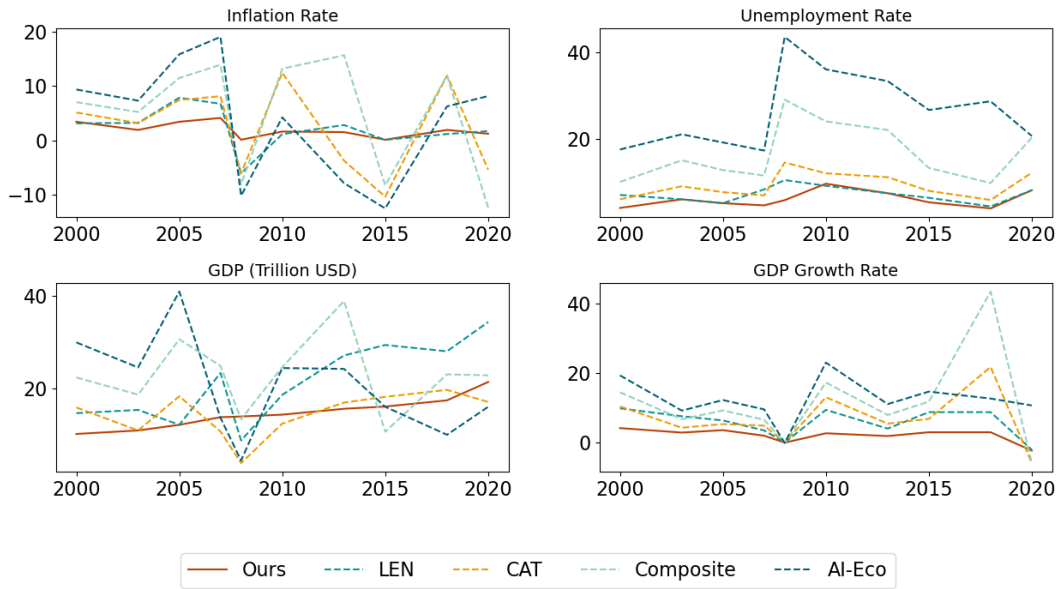


图 2. 宏观经济指标年度变化

### 5.2 宏观经济规则

菲利普斯曲线 [32] 和奥肯定律 [28] 是宏观经济模拟中用于验证模拟结果合理性的两个最常用的规律，分别描述了年失业率和工资通胀、失业率年增长率和实际 GDP 增长之间的负

相关性。如图 3 所示，基于 LLM 的 Agent 决策根据这两个规律正确地表现了现象（Pearson 相关系数为-0.619， $p<0.01$  和-0.918， $p<0.001$ ）。值得注意的是，基于规则的 baseline 方法在菲利普斯曲线上显示了不正确的正关系。我们将这一优势归因于 LLM 准确地认为失业时应该减少消费。请注意，由于失业率非常高，因此未显示 AI Eco 的菲利普斯曲线。

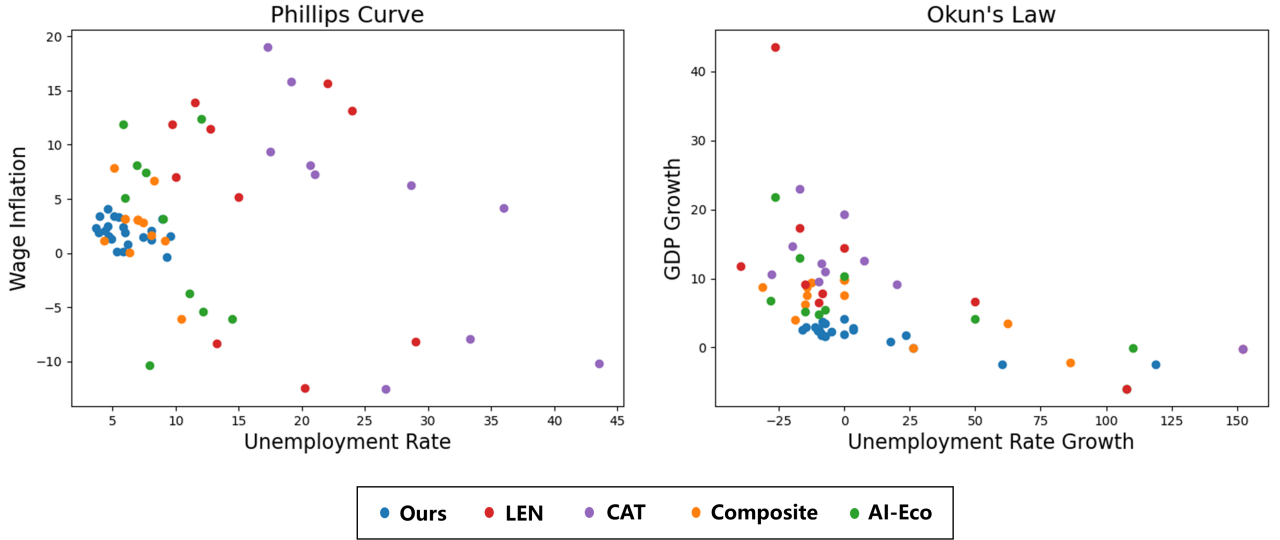


图 3. 菲利普斯曲线和奥肯定律

## 6 总结与展望

这项研究将大语言模型 (LLM) 与宏观经济学模拟进行了新的集成，设计了一个由 LLM 主导的经济学理性人代理，该代理具有对现实经济环境的感知、反思和决策能力。与传统的基于规则或深度学习的智能体相比，这个代理能够更加真实地模仿人类并复制出宏观经济现象。通过这一工作，LLM 的能力为模拟更现实的宏观经济现象提供了一条非常可观的途径。

展望未来，当前工作的前景非常广阔，尤其是在结合快速发展的大语言模型与强化学习范式方面。以下是几个潜在的有前景的发展方向：首先，大语言模型与强化学习的更深层次结合具有极大的潜力。通过进一步整合这两者，我们可以实现主体在宏观经济模拟中的自适应学习和优化。这种方法有望推动经济学领域的发展，为我们提供更准确、实用的模型，有助于更好地理解和应对复杂的宏观经济现象。其次，本文方法的适用性可以进一步扩展到其他经济学和交叉学科领域。通过将大语言模型与强化学习结合，我们可以应用这一方法来研究不同领域的经济现象，例如市场行为、资源分配、社会影响等。这将为跨学科研究提供新的方法和工具，拓宽我们对于复杂社会系统的认识。第三，引入多个大语言模型代理可以深入探讨共享环境下的协作、竞争和发展动态。这不仅有助于理解个体行为如何影响整个系统，还能够揭示在复杂互动中产生的新的经济现象。通过研究多智能体系统，我们可以更好地理解群体行为的演变和影响，为社会科学领域提供深刻见解。最后，本实验仍有很多不完善的

地方，例如在宏观经济模拟中缺少公司这一至关重要的主体。未来的工作可以通过使用 LLM agent 模拟多个公司职员或部门，以组建一个更真实的代理公司，从而更全面地考虑公司在整个宏观经济模拟实验中的影响。事实上，原论文中为了方便模拟分析，没有定义代理公司，如果有代理公司的加入，那么实验就可以和完整的 baseline (LEN 和 CAT 这两个 Baseline 的完整版中包含有公司角色，本文在实验中将两个 baseline 的思想抽象出来后重新定义公式并实验的) 进行更深入的对比，从而让实验结果更有说服力。此外，未来的研究还可以将重点扩展到更具体的场景，比如某支公司的股票交易、酒吧问题与少数者博弈等，以深化对具体经济问题的理解。总体而言，未来的发展将继续推动大语言模型与强化学习在经济学领域的应用，为我们提供更深刻、全面的经济学分析工具，助力解决现实世界中的复杂问题。

## 参考文献

- [1] Robert L Axtell and J Doyne Farmer. Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*, 2022.
- [2] Daniil A. Boiko, Robert MacKnight, and Gabe Gomes. Emergent autonomous scientific research capabilities of large language models. *arXiv e-prints*, page arXiv:2304.05332, April 2023.
- [3] Andres M Bran, Sam Cox, Oliver Schilter, Carlo Baldassari, Andrew D White, and Philippe Schwaller. ChemCrow: Augmenting large-language models with chemistry tools. *arXiv e-prints*, page arXiv:2304.05376, April 2023.
- [4] James Brand, Ayelet Israeli, and Donald Ngwe. Using gpt for market research. *Available at SSRN 4395751*, 2023.
- [5] William A Brock and Cars H Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic dynamics and Control*, 22(8-9):1235–1274, 1998.
- [6] Leland Bybee. Surveying Generative AI’s Economic Expectations. *arXiv e-prints*, page arXiv:2305.02823, May 2023.
- [7] Lawrence J Christiano, Martin Eichenbaum, and Charles L Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45, 2005.
- [8] Timothy Cogley and James M Nason. Effects of the hodrick-prescott filter on trend and difference stationary time series implications for business cycle research: *Journal of economic dynamics and control* 19 (1-2), january-february 1995, pp. 253–278. In *Real Business Cycles*, pages 626–651. Routledge, 2013.
- [9] Herbert Dawid and Domenico Delli Gatti. Agent-based macroeconomics. *Handbook of computational economics*, 4:63–156, 2018.



- [10] Herbert Dawid, Simon Gemkow, Philipp Harting, Sander Van der Hoog, and Michael Neugart. The eurace@ unibi model: An agent-based macroeconomic model for economic policy analysis. 2012.
- [11] Christophe Deissenberg, Sander Van Der Hoog, and Herbert Dawid. Eurace: A massively parallel agent-based model of the european economy. *Applied mathematics and computation*, 204(2):541–552, 2008.
- [12] Giovanni Dosi, Giorgio Fagiolo, Mauro Napoletano, and Andrea Roventini. Income distribution, credit and fiscal policies in an agent-based keynesian model. *Journal of Economic Dynamics and Control*, 37(8):1598–1625, 2013.
- [13] J Doyne Farmer and Duncan Foley. The economy needs agent-based modelling. *Nature*, 460(7256):685–686, 2009.
- [14] Giorgio Franceschelli and Mirco Musolesi. On the Creativity of Large Language Models. *arXiv e-prints*, page arXiv:2304.00008, March 2023.
- [15] Chen Gao, Xiaochong Lan, Zhihong Lu, Jinzhu Mao, Jinghua Piao, Huandong Wang, Depeng Jin, and Yong Li. S3: Social-network Simulation System with Large Language Model-Empowered Agents. *arXiv e-prints*, page arXiv:2307.14984, July 2023.
- [16] John Geanakoplos. The leverage cycle. *NBER macroeconomics annual*, 24(1):1–66, 2010.
- [17] Nigel Gilbert and Klaus Troitzsch. *Simulation for the social scientist*. McGraw-Hill Education (UK), 2005.
- [18] David F Hendry and Jean-Francois Richard. On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, 20(1):3–33, 1982.
- [19] John J Horton. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research, 2023.
- [20] Shi Jinxin, Zhao Jiabao, Wang Yilei, Wu Xingjiao, Li Jiawen, and He Liang. CGMI: Configurable General Multi-Agent Interaction Framework. *arXiv e-prints*, page arXiv:2308.12503, August 2023.
- [21] Dale W Jorgenson. Information technology and the us economy. *American Economic Review*, 91(1):1–32, 2001.
- [22] Grgur Kovač, Rémy Portelas, Peter Ford Dominey, and Pierre-Yves Oudeyer. The SocialAI School: Insights from Developmental Psychology Towards Artificial Socio-Cultural Agents. *arXiv e-prints*, page arXiv:2307.07871, July 2023.
- [23] Finn E Kydland and Edward C Prescott. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370, 1982.

- [24] Chao Li, Xing Su, Haoying Han, Cong Xue, Chunmo Zheng, and Chao Fan. Quantifying the Impact of Large Language Models on Collective Opinion Dynamics. *arXiv e-prints*, page arXiv:2308.03313, August 2023.
- [25] Nian Li, Chen Gao, Yong Li, and Qingmin Liao. Large Language Model-Empowered Agents for Simulating Macroeconomic Activities. *arXiv e-prints*, page arXiv:2310.10436, October 2023.
- [26] Siyu Li, Jin Yang, and Kui Zhao. Are you in a Masquerade? Exploring the Behavior and Impact of Large Language Model Driven Social Bots in Online Social Networks. *arXiv e-prints*, page arXiv:2307.10337, July 2023.
- [27] Jiaju Lin, Haoran Zhao, Aochi Zhang, Yiting Wu, Huqiuyue Ping, and Qin Chen. AgentSims: An Open-Source Sandbox for Large Language Model Evaluation. *arXiv e-prints*, page arXiv:2308.04026, August 2023.
- [28] Arthur M Okun. *Potential GNP: its measurement and significance*. Cowles Foundation for Research in Economics at Yale University, 1963.
- [29] OpenAI. Introducing chatgpt. <https://openai.com/blog/chatgpt>, 2022.
- [30] Joon Sung Park, Joseph C. O’Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative Agents: Interactive Simulacra of Human Behavior. *arXiv e-prints*, page arXiv:2304.03442, April 2023.
- [31] Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Social simulacra: Creating populated prototypes for social computing systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–18, 2022.
- [32] Edmund S Phelps. Phillips curves, expectations of inflation and optimal unemployment over time. *Economica*, pages 254–281, 1967.
- [33] Chen Qian, Xin Cong, Wei Liu, Cheng Yang, Weize Chen, Yusheng Su, Yufan Dang, Jiahao Li, Juyuan Xu, Dahai Li, Zhiyuan Liu, and Maosong Sun. Communicative Agents for Software Development. *arXiv e-prints*, page arXiv:2307.07924, July 2023.
- [34] Kambiz Rasoulkhani, Ali Mostafavi, Maria Presa Reyes, and Mostafa Batouli. Resilience planning in hazards-humans-infrastructure nexus: A multi-agent simulation for exploratory assessment of coastal water supply infrastructure adaptation to sea-level rise. *Environmental Modelling & Software*, 125:104636, 2020.
- [35] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language Models Can Teach Themselves to Use Tools. *arXiv e-prints*, page arXiv:2302.04761, February 2023.

- [36] Frank Schorfheide and Dongho Song. Real-time forecasting with a mixed-frequency var. *Journal of Business & Economic Statistics*, 33(3):366–380, 2015.
- [37] Alexander Trott, Sunil Srinivasa, Douwe van der Wal, Sebastien Haneuse, and Stephan Zheng. Building a Foundation for Data-Driven, Interpretable, and Robust Policy Design using the AI Economist. *arXiv e-prints*, page arXiv:2108.02904, August 2021.
- [38] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Ji-Rong Wen. A Survey on Large Language Model based Autonomous Agents. *arXiv e-prints*, page arXiv:2308.11432, August 2023.
- [39] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wensen Cheng, Qi Zhang, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, and Tao Gui. The Rise and Potential of Large Language Model Based Agents: A Survey. *arXiv e-prints*, page arXiv:2309.07864, September 2023.
- [40] Babyagi Yoheinakajima. Babyagi. <https://github.com/yoheinakajima/babyagi>, October 2023.
- [41] Stephan Zheng, Alexander Trott, Sunil Srinivasa, David C Parkes, and Richard Socher. The ai economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science advances*, 8(18):eabk2607, 2022.
- [42] Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. MiniGPT-4: Enhancing Vision-Language Understanding with Advanced Large Language Models. *arXiv e-prints*, page arXiv:2304.10592, April 2023.