Expectation formation in Large Language Models

Nikoleta Anesti* Edward Hill Andreas Joseph

Bank of England

Macroeconomic Modelling Frontiers for Research and Policy in Central Banks
Frankfurt
27-28 November 2024

The views expressed are our own and do not necessarily reflect those of the Bank of England.

Introduction

- The role of agents' expectations has been central in economics both theoretically and empirically.

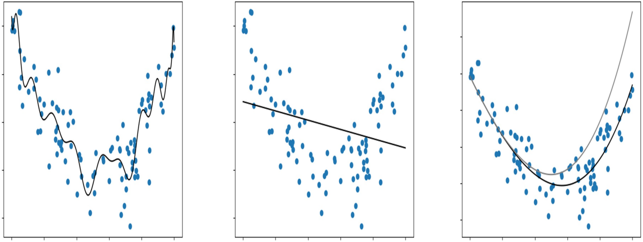
- Despite being widely studied, the details of expectations formation of households (HH) and firms remain not well understood.

Introduction

- The role of agents' expectations has been central in economics both theoretically and empirically.
- Despite being widely studied, the details of expectations formation of households (HH) and firms remain not well understood.
- Can innovations in the Large Language Models (LLMs) literature bring any new insights?
- Can LLMs be used in an economic policy context?

From the Generative AI side. . .

- LLMs must have a model of human behaviour to produce sensible outputs. But could be not generalised enough, over-generalised and lacking detail, or inaccurate but subtly so.



- We investigate if this model is qualitatively and quantitatively accurate in a quantifiable and well explored area with multiple axes and levels of detail
- Are LLMs a new tool for economics or a passing novelty?

Our contribution

- We test the sensitivity of LLMs to the wording of various questions and whether they could used to generate data that would 'look like' official statistics in a cost effective way.

Our contribution

- We test the sensitivity of LLMs to the wording of various questions and whether they could used to generate data that would 'look like' official statistics in a cost effective way.
- We construct prompts to test the ability of ChatGPT to form inflation perceptions and expectations and compare them to household survey data and official statistics in a *quasi-experimental* setup.

Our contribution

- We test the sensitivity of LLMs to the wording of various questions and whether they could used to generate data that would 'look like' official statistics in a cost effective way.
- We construct prompts to test the ability of ChatGPT to form inflation perceptions and expectations and compare them to household survey data and official statistics in a *quasi-experimental* setup.
- We provide evidence of ChatGPT's ability to track surveys and official releases quite well and replicate some empirical regularities of HH inflation expectations data.

Methodology and related literature

- LLM responses to multiple choice questions are ordering dependent (Zheng et al. 2024, Pezeshkpour and Hruschka 2023).

- We do not use chain-of-thought or other techniques, so as to probe the internal state of the model (Wei et al. 2023).

Methodology and related literature

- LLM responses to multiple choice questions are ordering dependent (Zheng et al. 2024, Pezeshkpour and Hruschka 2023).

- We do not use chain-of-thought or other techniques, so as to probe the internal state of the model (Wei et al. 2023).

- Examples of LLMs exhibiting human behaviours across a range of fields (Aher et al. 2022, Brookins and DeBacker 2023, Bybee 2023, Faria-e Castro and Leibovici 2023, Horton 2023, Griffin et al. 2023, Argyle et al. 2023, Bisbee et al. 2023)

- The importance of perceptions and expectations formation: **Frequency of purchase** (Mankiw and Reis 2002, Mackowiak and Wiederholt 2009, Coibion and Gorodnichenko 2015) **Bias in representative bundle** (Van der Klaauw et al. 2012, De Bruin et al., 2011, D'Acunto et al. 2021) **Demographic biases** (D'Acunto et al. 2021), **Upward movement bias**(Mankiw 2003, D'Acunto and Weber (2024).

General setup: Households survey data

- We exploit the Bank of England Inflation Attitude quarterly survey of about 4.5k households,

General setup: Households survey data

- We exploit the Bank of England Inflation Attitude quarterly survey of about 4.5k households,
- Multiple choice questions about inflation perceptions and expectations (1, 2, and 5 years ahead), including other issues e.g. key worries, anticipated response to changes (February waves).
- Rich information on demographics: age, sex, geographic region, housing tenure employment status, income, education, class.

General setup: Survey reproduction

We instruct the model, condition on demographics and economic variables and provide the options:

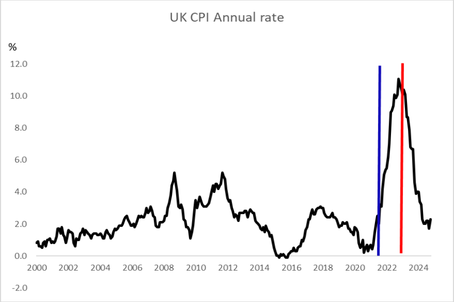
System prompt	You are pretending to be the person described given your best guess as to their personal, social and economic situation.
Demographic conditioning	You are male, aged 16-24, live in the North of England or Northern Ireland, are upper-middle class and are not working with an income of >£45000. You got your A-levels but not a degree and live in a house you rent.
Economic conditioning	In the last few months, food inflation has been 17% (10% in restaurants and cafes), energy price inflation was about 50%. On average the rate of inflation on other goods was about 6%.
Rubric	You are going to be asked questions about your perception of current and future inflation., In five years time, how much would you expect prices in the shops generally to change over a year?
Options	go down by less than 1% go down by 1-2% go down by 2-3% ... rise by more than 15% Please choose one option, no explanation.

General setup: Economic Conditioning

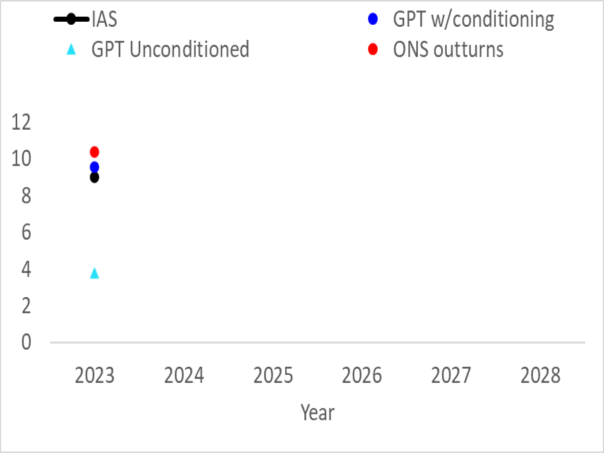
- When conditioning, we condition on information from **February 2023** data: Food inflation 17%, 10% in restaurants, Energy inflation: 50%, Other components: 6%.

General setup: Economic Conditioning

- When conditioning, we condition on information from **February 2023** data: Food inflation 17%, 10% in restaurants, Energy inflation: 50%, Other components: 6%.
- The training data for the model comes from before **September 2021** so the conditions are *out of time* and *out-of-experience*.

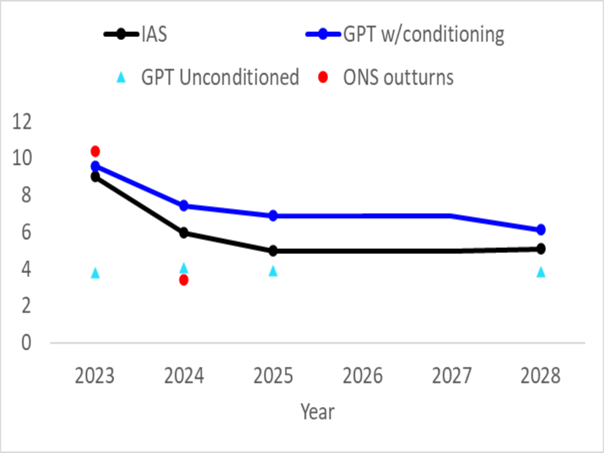


Aggregate Results over Time



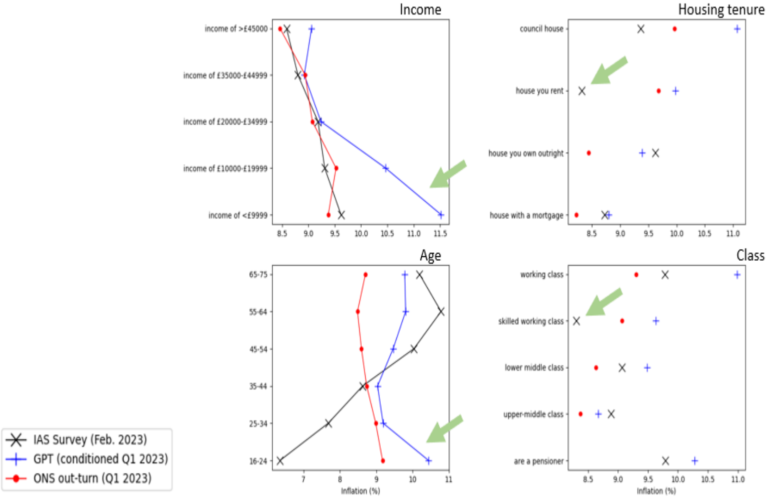
- GPT perceptions about current inflation align almost perfectly with IAS and only a bit below official statistics.

Aggregate Results over Time

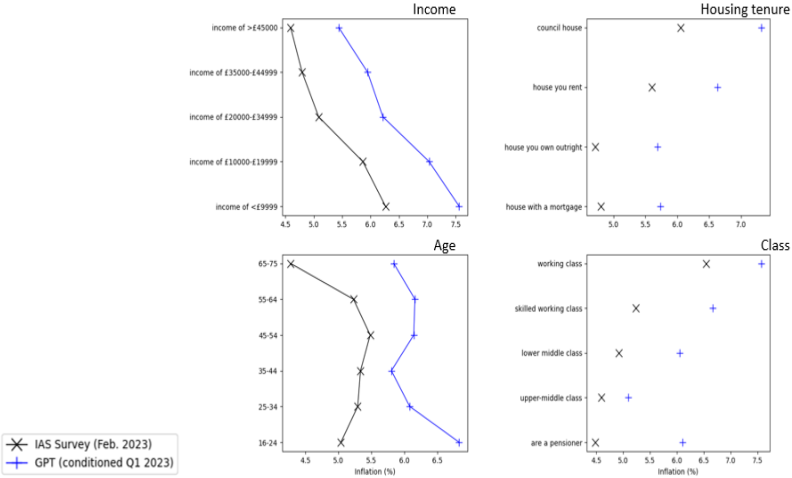


- GPT expectations are higher compared to the IAS although they converge over the 5-year horizon.

Conditioned inflation perceptions at demographic level



Conditioned inflation expectations 5 years ahead



Robustness: Is conditioning information relevant?

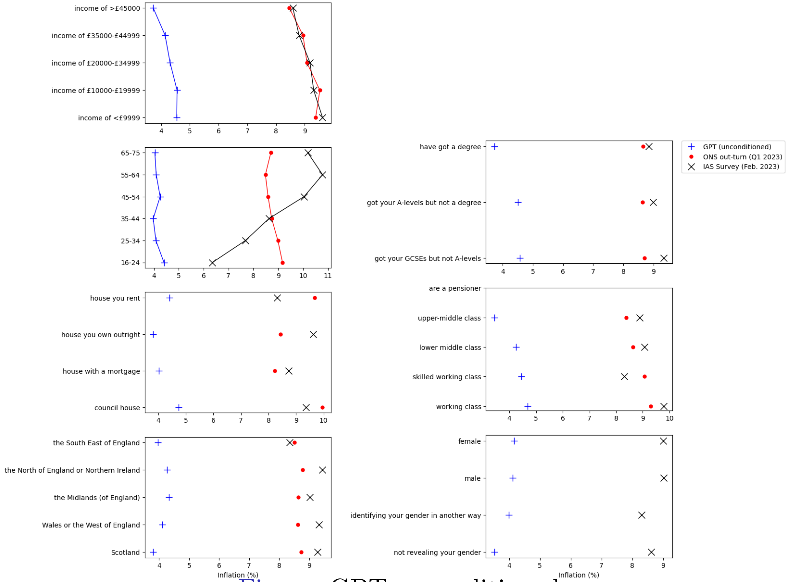
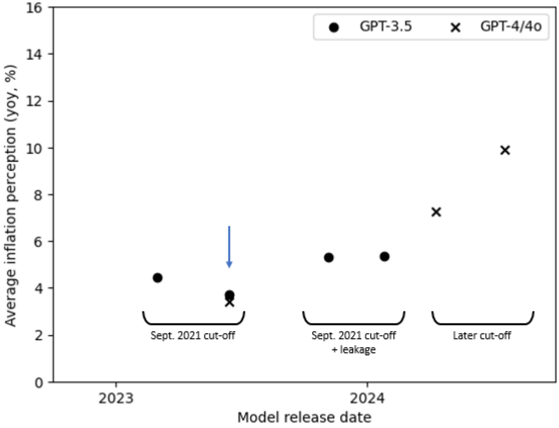


Figure: GPT unconditioned

Robustness: Model choice



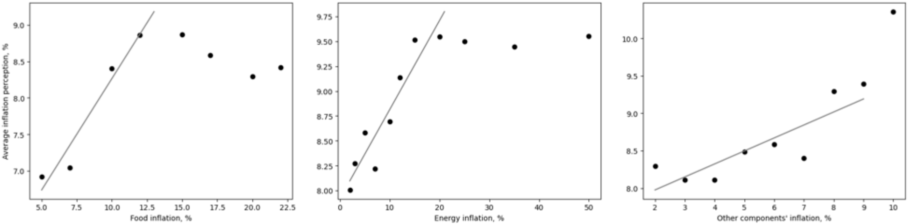
- The results are *not* consistent: further support for the importance of the experimental setting.
- Information leakage: later GPT-3.5 models know about events after their nominal cut-off.

Robustness: Correlation and Model complexity

Number of demographics	Maximum R^2	Chosen demographics
All	0.78	(all)
1	0.31	emp. status
2	0.50	emp. status, tenure
3	0.62	emp. status, tenure, income

- Are the results simply a response to key variables projected onto correlated population demographics?
- Method: OLS on demographic categories as dummies

Robustness: Other economic conditioning/sensitivity checks



Conditioning varied	Sensitivity	Fraction of basket (ONS)
Food	0.31	0.14
Energy	0.09	0.04
Other components	0.17	0.82

Key takeaways

- We elicited inflation perceptions and expectations from an LLM within a quasi-experimental setup.
- While the behaviour in time has features of people's perceptions the demographic breakdown has more in common with the inflation outturns.
- While the LLM shows 'sensible' behaviour the variation over models and the mixture of desired and undesired behaviour suggests:
  1. A need for caution when extracting substantial information from the LLM,
  2. The development of base models validated for economics.