Heterogeneity in Monetary Policy Preferences

An LLM-based approach

Author: Tim Leander Jascheck

Course: Machine Learning and AI in Economics
Instructor: Thiemo Fetzer
University: University of Bonn

Date: August 18, 2025

Contents

1 Introduction and Motivation

Heterogeneity in beliefs and sentiments across agents has become a central point of interest in macroeconomics (Roth & Wohlfart, 2018, 2020; Roth et al., 2022). This is because, for example, the effectiveness of forward guidance crucially hinges on central bank credibility. Modelling the economic agent in an orthodox way to include such complex, idiosyncratic elements is not feasible, as even when allowing for some forms of heterogeneity, traditional approaches still rely on modelling agents through a sparse system of equations to remain tractable.

This proposal outlines one avenue of how large language models (LLMs) could be used to understand and model the drivers behind agents' assessments. Recently, both machine learning approaches for natural language processing (Borag & Drechsel, 2024) and the use of LLMs (Li et al., 2024; Zheng et al., 2022) have made their first appearances in the realm of macroeconomics. As in Li et al. (2024), I will make use of the non-linearity that LLMs operate in and their related ability to answer based on different profiles they are provided with. However, in contrast to Li et al. (2024), I focus on retrieving preferences and latent beliefs of agents, instead of behavioural patterns. Thus, the study is more similar to having the LLM act repeatedly as a survey taker under heterogeneous profiles and with varying questions. The idea behind this is that an LLM might be better at reporting attitudes and sentiments instead of actions, as this is their primary knowledge base and *modus operandi*. While simulating a whole economy based on interacting LLMs seems to be an ambitious goal (for now), trying to assess the feasibility of using LLMs as survey respondents and comparing the results to some sources of external validity is already very much feasible.

The exercise is done in the framing of preferences and expectations over optimal monetary policy. Using traditional survey data, Jost (2018) finds different preferences concerning monetary policy based on cultural backgrounds, while Coleman and Nautz (2023) investigate the dynamics of central bank credibility between different demographic groups. Strongly related to the work done in this proposal, Carvalho and Nechio (2014) test whether households form their beliefs in a way consistent with a Taylor (1993)-type rule. They find that while households draw a conventional, positive connection between rising inflation and rising interest rates, the relationship between unemployment and interest rates is much more obscure.

2 Economic Framework

The Taylor rule is an equation describing monetary policy in a stylised way, aiming to capture the systemic part of interest rate determination by the central bank. It draws a relationship between the contemporaneous interest rate i_t, inflation rate π_t, and output level y_t. A common formulation is the following:

$$i_t = r^* + \pi_t + \tilde{\phi}_\pi \left(\pi_t - \pi^*\right) + \phi_y \left(y_t - \bar{y}_t\right)$$

where r_t^* and $\bar{y}_t$ define the unobservable 'natural rate of interest' and 'potential level of output', respectively. The equation can be equivalently expressed as:

$$i_t = i^* + \phi_\pi \left(\pi_t - \pi^*\right) + \phi_y \left(y_t - \bar{y}_t\right)$$

where we redefine $i^* = r^* + \pi^*$, $\phi_\pi = 1 + \tilde{\phi}_\pi$. With $\phi_y = 0.5, \phi_\pi = 1.5$, and $\pi^* = 2$ we obtain the initial calibration proposed by Taylor (1993). While exact weights differ between calibrations, $\phi_y > 0, \phi_\pi > 1$ are considered necessities for a monetary policy that can successfully stabilise the economy.

Often, we measure aggregate fluctuations by targeting a 'natural' unemployment level rather than a deviation from a hypothetical 'potential output' state. This is done as *Okun's Law* provides a tight relationship between unemployment and the output gap (Foroni & Furlanetto, 2022), while the unemployment rate provides a clearer measure of economic slack (Bernardini & Lin, 2024). Additionally, it seems much more reasonable for a non-economist to comprehend and detect unusually high or low values of the unemployment rate, than the GDP-performance of a country, which makes our later survey formulation more plausible. If we additionally suppose a time invariant natural level of unemployment, we can then also aggregate all non-observable values in the residual intercept i^0:

$$i_t = i^0 + \phi_\pi \pi_t + \phi_u u_t \tag{1}$$

with $i^0 = i^* - \phi_\pi \pi^* + \phi_u u_t - \bar{u}$. This yields a very convenient representation of the Taylor rule: If primary interest lies in the Taylor weights ϕ_u, ϕ_π, there is no need to pin down the elusive components that make up i^0, and a simple regression suffices to estimate the weights. Using Okun's Law and estimations of Carvalho and Nechio (2014), following Bernardini and Lin (2024), $-2.5 \cdot \phi_y \approx \phi_u$ might be an appropriate approximation to translate between a rule working with the output gap or unemployment. Thus using the historical benchmark of Taylor (1993) and estimations by Bernardini and Lin (2024),[1] it seems to be adequate to use $\phi_\pi = 1.5$ and $\phi_y = 0.5 \Rightarrow \phi_u \approx -0.2$ as a useful benchmark for later results.

[1] Bernardini and Lin (2024) use a formulation with inertia, which yields the parameter discussed for the contemporaneous effects.

3 Set up

The main goal of the LLM integration is to estimate survey respondents' preferences over Taylor weights ϕ_u, ϕ_π, thereby receiving an intuition of what (heterogeneous) consumers may regard as optimal monetary policy. Retrieval of these responses is done through an LLM, which acts in place of human survey takers. Thus, the LLM acts as both an inference tool and a retrieval tool; answers are retrieved from the LLM, which is expected to base them on information inferred from a suitable knowledge base.

3.1 Conceptual

Additionally to the framework discussed in the preceding section, a main focus of this implementation is to allow for heterogeneity across agents. This is done via individualised profiles similar to Li et al. (2024): Agents are made up of a unique bundle of characteristics. An overview of the characteristics chosen here is given in Table 1. I consider 'Politics' to be an especially interesting dimension for which to receive results, as it is largely unexplored by the literature. This might be because it is a latent characteristic that is not directly observable.

Taking the Cartesian product across these attributes results in 1440 unique profiles. I will use these 1440 profiles as an Artificial Cross Section (ACS); later results using this ACS give equal weight to observations, as if the ACS were a random draw from an underlying distribution. A step for future analysis would entail making this ACS more representative by using information on demographic distributions to make the distribution of profiles mimic the population of interest, e.g. the German consumer base.

Variable	**Categories / Values**
Characteristics	
Gender	Men, Women
Age	18, 35, 50, 65
Kids	0, 1, 2
Income	Less than 1500; 1500–2500; 2500–4000; More than 4000
Education	No high school; High school; College
Politics	Left-wing; Left-leaning; Centre; Right-leaning; Right-wing
Economic Scenarios	
Inflation	0, 2, 4, 6, 8, 10, 12
Unemployment	0, 2, 4, 6, 8, 10, 12

Table 1: Overview of variables and their possible values

On the system level, the LLM agents receive the instructions:

You are a German {*gender*}, {*age*} years old and have {*kids*} kid(s).
You have {*politics*} political views.
You have {*education*} degree and earn {*income*} Euros a month.
You must answer ONLY with a single number representing the interest rate in the format: {X.XX}. Do not add any explanation, reasoning, or extra text. Do NOT show your thought process. Reply ONLY with the number.

The text the agents are then prompted with describes an economic scenario:

Inflation is at {*inf*}% and unemployment is at {*unemp*}%. Given this current economic climate, what is your preferred interest rate?

Running each of these profiles for the 36 Economic scenarios results in 51840 requests and thus a data set of this size. For each question, the LLM is reset, thus each answer is independent of the others in the sense that there is no learning during the retrieval of answers.

3.2 Code and LLM

A Python script was used to generate responses by calling a personal API which hosts an LLM. R was used for creating statistics and graphs. After a bit of testing, I settled on using Google's *gemma3* LLM, hosted via the local Ollama platform. Both *llama3.1* and *deepseek-r1* performed worse in terms of adhering to the demanded format. One major, interesting problem faced when using both of these models was that when prompted to assume the profile of someone whose political leaning is classified as 'right-wing', the LLMs would categorically refuse to answer the question. Instead they gave responses such as "As a large language model, I cannot assume extremist positions." Or, not revealing the real reason for the denial, simply: "I can not give opinions about monetary policy." This is not the case for the term "left-wing", perhaps an interesting insight into the underlying construction of these LLMs.

As the simulation ran for multiple hours[2] an important challenge in designing the code was to allow for batch-wise generation, so progress is not lost if the simulation is interrupted. Thus, the data sets are generated in larger chunks which are then merged in a dedicated script.

[2]The code took close to 4 hours (3.93 hours) for the baseline data set used in the regression.

4 Results

The average response of the ACS to the varying economic states is shown in Figure 1.

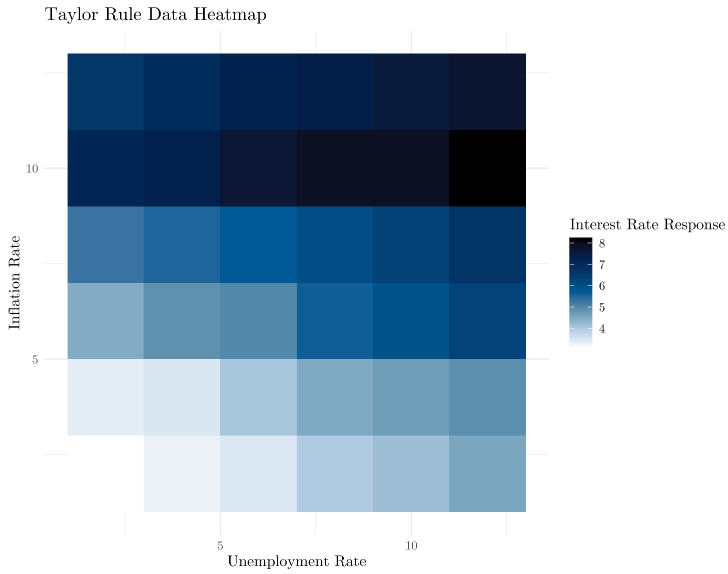


Figure 1: Mean responses from the LLM for different values of inflation and unemployment.

As can be seen, mean responses appear relatively stable and move within plausible ranges. To quantify the observed behaviour, I estimate the model proposed in Equation 1. Additionally, I provide a variant that adds a vector of dummies (D) for the characteristics laid out in Table 1:

$$i_t = \tilde{i}^0 + D\beta + \phi_\pi \pi_t + \phi_u u_t \tag{2}$$

In its current form, the model does not include interaction effects across demographics or with the economic indicators. Later investigation will show that these interaction effects contribute little in any case.

The estimated coefficients are reported in Table 2. I do not report standard errors or significance values for my results, as it seems very unclear what estimation variance implies in the context of data generated by LLMs. LLM data cannot be considered i.i.d., as they are not randomly drawn from an underlying population but instead from a highly preprocessed model that does not reproduce randomness properly (Koevering & Kleinberg, 2024), but instead optimises for likelihood.

Although model fit is also hard to measure when it comes to LLM-generated data and in this ACS setup, it is promising that the model captures 71.2% of the variation in ACS responses, as measured by its R^2.[3] This shows that responses by the LLM can still be relatively well captured by simple linear relationships, without the need for complex interaction terms. For a larger-scale analysis, perhaps using a more representative ACS, it might be preferable to implement a Double Machine Learning approach to capture particularly important interactions across characteristics.

The chosen ACS setup has a convenient side effect: As every combination of characteristics appears exactly once, the regressors describing these characteristics have an empirical correlation of zero. Since omitted variable bias can be understood as a function of correlation between regressors, which is absent at zero correlation, the inclusion of additional regressors does not affect the previous estimations. This also allows for a clean separation of variance explanation. As there is no informational overlap in the explanatory power of different regressors, the informational value of different regressors can be neatly separated. Figure 3 shows the decomposition of the explained and unexplained variance of the model. Subject to the chosen ACS, characteristics and the economic environment account for about half of the explained variation in answers. The following subsections discuss these findings with respect to the economic environment and the demographic characteristics.

4.1 Economic environment

The model captures a clear, pronounced sensitivity to inflation ($\phi_\pi = 0.4078$), with the expected sign indicating a preference for tighter monetary policy in response to a higher inflation rate. While sizeable, the coefficient is far below the value suggested by macroeconomic theory: following the Taylor principle, we would expect a coefficient greater than one, typically around 1.5.

Notably, and contrary to standard macroeconomic wisdom, agents appear to favour a higher interest rate when unemployment is high ($\phi_u = 0.1408$). The two terms behave quite orthogonally, and including an interaction makes very little difference. The same holds for adding a quadratic term. Thus, approximating the behaviour of the agents via a Taylor-rule-type description seems reasonable. Overall, extending the model with interaction terms and higher-order polynomials contributes virtually nothing to explaining the variance in responses (see Table 3).

[3] Due to the overwhelming size of the ACS and the relatively sparse model structure, whether one uses R^2 or adjusted R^2 for any of the following claims is of no relevance.

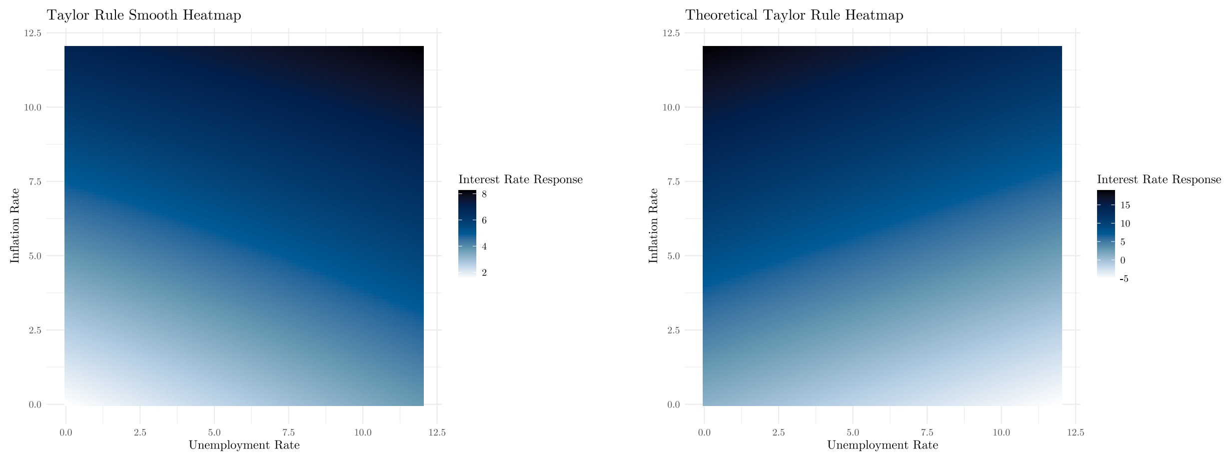
Coefficient	Estimate (Large)	Estimate (Small)
Intercept	−0.4394	1.8417
Inflation ϕ_π	0.4078	0.4078
Unemployment ϕ_u	0.1408	0.1408
Gender (ref: Man)		
Woman	0.7359	
Age (ref: 18)		
35	−0.4786	
50	−0.6414	
65	−0.9669	
Kids (ref: No kids)		
1 Kid	0.0828	
2 Kids	0.0869	
Politics (ref: Centre)		
Left-leaning	1.0213	
Left-wing	1.5207	
Right-leaning	2.7024	
Right-wing	3.2672	
Education (ref: College degree)		
High school	0.4026	
No high school	1.4601	
Income (ref: 1500–2500)		
Less than 1500	0.4056	
2500–4000	−0.1867	
More than 4000	0.0016	
Adj. R^2	0.712	0.366

Table 2: Regression coefficients and estimates for large and small models ("ref:" describes the reference group captured by the intercept)

Estimating the responses based on a accordingly calibrated Taylor equation leads to a smoothed representation, as shown in Figure 2a, contrasted with the theoretical benchmark calibration of Taylor (1993) in Figure 2b.

Although relatively little work has been done on calibrating the Taylor Rule for consumers, Carvalho and Nechio (2014), using the Michigan Survey, find results with an interesting resemblance to those extracted from the LLM. While consumers generally understand the interplay between inflation and the interest rate, the results are much more nuanced when it comes to unemployment, where consumers are less clear in determining optimal responses.[4]

[4]For a study investigating changes in Taylor-rule consistency over time, see Dräger et al. (2022).



(a) Interpolated linear Taylor-rule from the LLM across the synthetic sample.

(b) Interest rate responses in line with text-book economics.

Figure 2: Comparison of responses from the LLM and the classic Taylor-rule. Note that the heat maps use different colour scales.

Coefficient	**Estimate**
Intercept	1.2157
Inflation ϕ_π	0.5671
Inflation^2	-0.0084
Unemployment ϕ_u	0.1898
Unemployment^2	-0.0005
Inflation $\cdot$ Unemployment	-0.0061
Adj. R^2	0.368

Table 3: Regression coefficients from an extended, non-linear model

4.2 Demographics

The few characteristics given in the profile have almost the same explanatory power for the variation in interest rates as the provided economic scenario. When estimating Equation 2, the characteristics are modelled to affect only the level of the preferred rate, not, for example, the responsiveness of the rate to inflation. In terms of Equation 1, this means that demographic characteristics are only allowed to shift i^0, but do not lead to heterogeneous Taylor weights ϕ^i_π, ϕ^i_u. As we will see, such non-linearities, as before, do not make an important contribution to explaining the responses of the LLM.

Generally, the coefficients are plausible in size and, for most attributes that have a natural ordering (Age, Education, Kids), move in a monotonic manner across the categories. According to the LLM, women seem to prefer higher interest rates, and the associated coefficient is somewhat sizeable. That women tend to have higher inflation perceptions and expectations is well documented (e.g. by Coleman and Nautz, 2025) and might also play a role in preferences over interest rates. Higher age groups prefer lower interest rates, and so do groups with, on average, higher education. If one draws the plausible connection that the low-interest policy of the ECB in the past may have especially harmed the central bank's credibility among those who prefer higher interest rates, these trends are consistent with the findings of Coleman and Nautz (2025). Both the effects of income and children are small or unstable and carry little explanatory power.

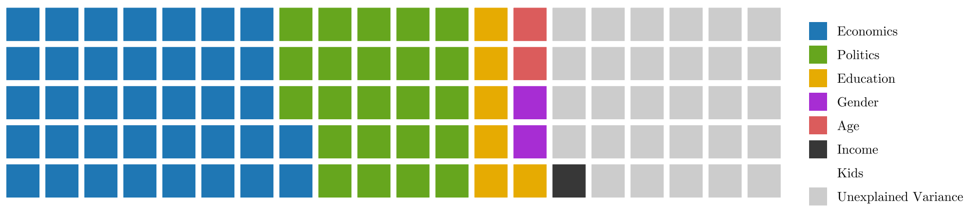


Figure 3: Variance Decomposition of the Interest Response. The economic conditions and the political views capture the overwhelming majority of the explained variance.

The political position provided to the LLM was by far the attribute with the most explanatory power (capturing 23% of the total variation). While right-leaning views are estimated to be generally associated with a preference for higher interest rates, the attributes also show a horseshoe-like pattern, with preferences diverging upwards at both ends of the political spectrum. This is an interesting and plausible finding: Especially parties at the far ends of the political spectrum frequently evoked the image of a "dispossession" of savers through the ECB's low-interest policy after the financial crisis.[5]

It may seem plausible to test for interaction effects between covariates, as this could capture an interesting form of heterogeneity: Perhaps different demographics not only have views on the optimal level of interest rates, but also on the optimal responsiveness of monetary policy? Figures 4 and 5 plot the respective heat maps for the 'Gender' and 'Politics' attributes. As can be seen, patterns are generally comparable, and heterogeneous effects beyond level shifts appear relatively muted. This can also be quantified: When

[5]See, for instance, the denunciation of the "EZB-Enteignungspolitik" ["ECB dispossession policy"] in the 2017 election program of the far-right *AfD*, or the Bundestag motion concerning the "Sparer-Enteignung" ["saver dispossession"] submitted by the far-left party *Die Linke* in 2016.

including interaction effects for Gender as well as Politics with both Unemployment and Inflation, the share of explained variation increases only marginally to 0.746 (from 0.712). This increase seems unimpressive, considering the ten additional regressors required. The only coefficients of noticeable size are negative interaction effects between inflation and right-leaning or right-wing political views.

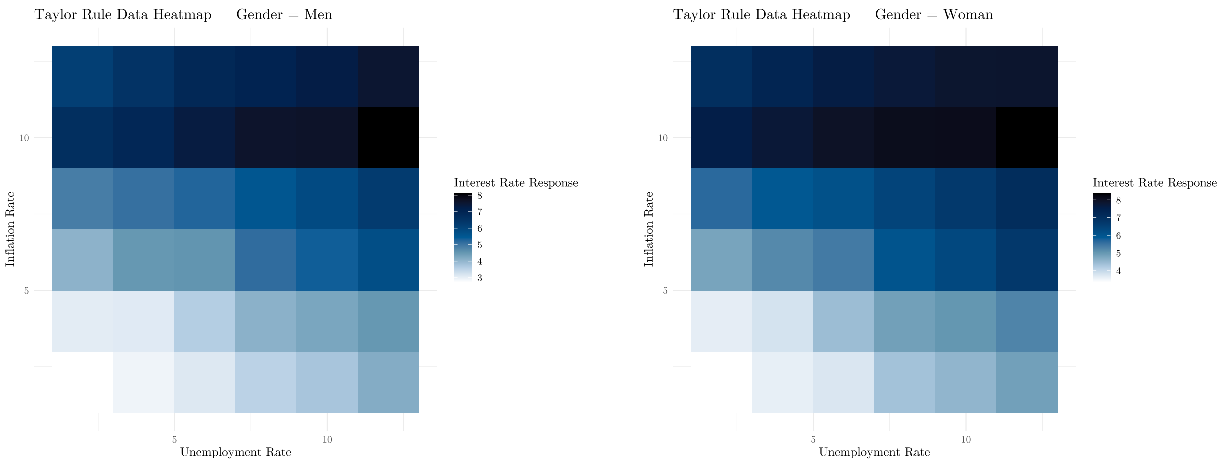


Figure 4: Mean responses from men and women. Gender differences are, apart from a shift in level, not pronounced.

5 Conclusion

This paper demonstrated an avenue for integrating Large Language Models (LLMs) into the field of agent-based macroeconomics. The exercise was conducted within a setting of optimal monetary policy, where the LLM substituted for human survey respondents. By assigning individualised profiles that varied across demographic characteristics, the analysis further examined how LLM responses differ across heterogeneous agent types.

The results are generally plausible and, importantly, show that LLM responses can be captured by simple linear models, though with notable departures from established macroeconomic theory. A comparative study, replicating the survey design with human respondents, would be a valuable next step. Such an exercise could clarify whether LLMs reproduce human answering behaviour and assess their potential as substitutes in preliminary analyses. If successful, this would allow for faster, cheaper, and more targeted acquisition of insights into agents' preferences and belief formation.

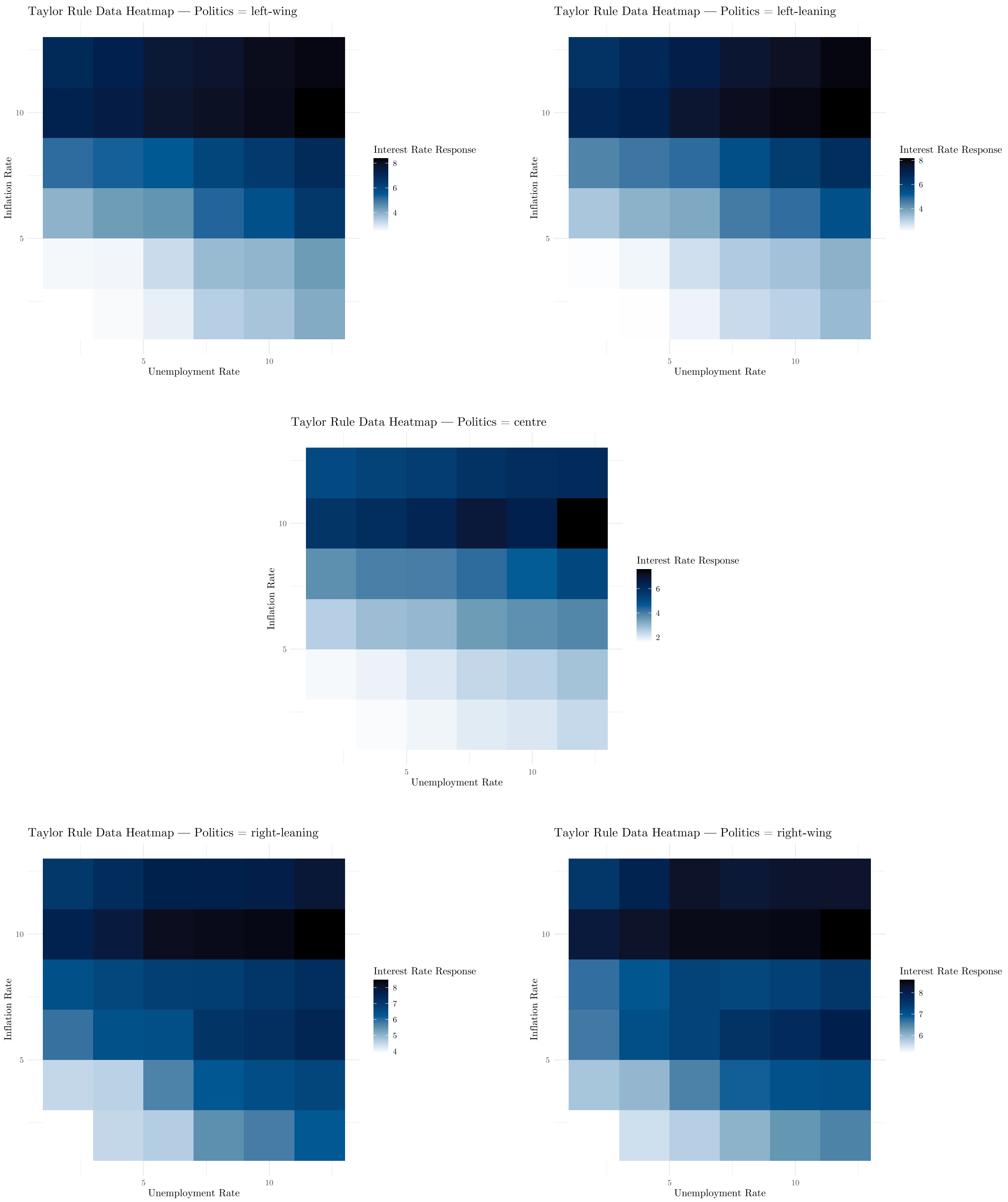


Figure 5: Mean responses across the political spectrum.

References

Bernardini, M., & Lin, A. (2024). Out of the ELB: Expected ECB policy rates and the Taylor rule. *Economics Letters*, *235*, 111546.

Borag, S., & Drechsel, T. (2024). Identifying monetary policy shocks: A natural language approach.

Carvalho, C., & Nechio, F. (2014). Do people understand monetary policy? *Journal of Monetary Economics*, *66*, 108–123.

Coleman, W., & Nautz, D. (2023). Inflation Expectations, Inflation Target Credibility, and the COVID-19 Pandemic: Evidence from Germany. *Journal of Money, Credit and Banking*, *55*(7), 1937–1953.

Coleman, W., & Nautz, D. (2025). Asymmetric Inflation Target Credibility.

Dräger, L., Lamla, M. J., & Pfajfar, D. (2022). The Hidden Heterogeneity of Inflation and Interest Rate Expectations: The Role of Preferences. *SSRN Electronic Journal*.

Foroni, C., & Furlanetto, F. (2022). Explaining deviations from Okun's law. *ECB Working Paper Series*, (2699).

Jost, A. (2018). Cultural differences in monetary policy preferences. *SNB Working Papers*, *2*.

Koevering, K. V., & Kleinberg, J. (2024). How Random is Random? Evaluating the Randomness and Humaness of LLMs' Coin Flips.

Li, N., Gao, C., Li, M., Li, Y., & Liao, Q. (2024, August). EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 15523–15536). Association for Computational Linguistics.

Roth, C., Settele, S., & Wohlfart, J. (2022). Beliefs about public debt and the demand for government spending. *Journal of Econometrics*, *231*(1), 165–187.

Roth, C., & Wohlfart, J. (2018). Experienced inequality and preferences for redistribution. *Journal of Public Economics*, *167*, 251–262.

Roth, C., & Wohlfart, J. (2020). How do expectations about the macroeconomy affect personal expectations and behavior? *The Review of Economics and Statistics*, *102*(4), 731–748.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, *39*, 195–214.

Zheng, S., Trott, A., Srinivasa, S., Parkes, D. C., & Socher, R. (2022). The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science Advances*, *8*(18).