

The Inflation Attention Threshold and Inflation Surges

Oliver Pfäuti*

First Version: August 2023

This version: September 2023

[Link to most recent version](#)

Abstract

At the outbreak of the recent inflation surge, the public's attention to inflation was low but increased rapidly once inflation started to rise. In this paper, I develop a general equilibrium monetary model where it is optimal for agents to pay little attention to inflation when inflation is low and stable, but in which they increase their attention once inflation exceeds a certain threshold. Using survey inflation expectations, I estimate the attention threshold to be at an inflation rate of about 4%, with attention in the high-attention regime being twice as high as in the low-attention regime. When calibrated to match these findings, the model generates inflation and inflation expectation dynamics consistent with the recent inflation surge in the US. The attention threshold induces a state dependency: cost-push shocks become more inflationary in times of loose monetary policy. These state-dependent effects are absent in the model with constant attention or under rational expectations. Following simple Taylor rules triggers frequent and prolonged episodes of heightened attention, thereby increasing the volatility of inflation, and—due to the asymmetry of the attention threshold—also the average level of inflation, which leads to substantial welfare losses.

JEL Codes: E3, E4, E5, E7

Keywords: Inattention, Inflation, Inflation Expectations, Monetary Policy

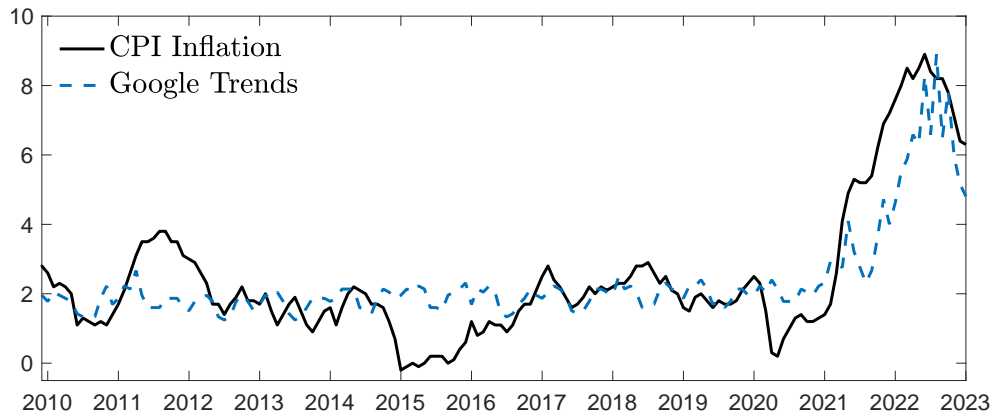
*University of Bern, Department of Economics, Schanzeneckstrasse 1, 3012 Bern, Switzerland, pfaeuti.oliver@gmail.com.

I thank Luca Benati, Laura Gáti, Niklas Kroner, David Munro, Fabian Seyrich and Hannes Twieling for helpful comments and suggestions.

1 Introduction

Inflation is back. After decades of low and stable inflation, inflation surged in many advanced economies during the recovery phase of the pandemic. Inflation turned out to be higher and more persistent than many expected.¹ With inflation rising, the public’s attention to inflation increased as well (Figure 1). But is this increased attention to inflation merely a byproduct or a driver of high and persistent inflation?

Figure 1: Inflation is back... on people’s minds



Notes: The black-solid line shows the monthly year-on-year CPI inflation rate for the US, and the blue-dashed line shows the number of Google searches in the US of the word inflation (normalized such that the two series have the same maximum).

To shed light on this question, I develop a framework in which agents optimally pay little attention to inflation when inflation is low and stable, but in which attention suddenly increases once inflation exceeds a certain threshold. I estimate the varying degrees of attention and the attention threshold using household inflation expectations data spanning the period 1978-2023. I find the attention threshold to be at an annualized quarter-on-quarter inflation rate of about 4.0%, implying that the US economy recently entered the *high-attention regime* in early 2021. Since 1978, the US economy spent about 30% of the time in the high-attention regime in which attention is twice as high compared to the low-attention regime. Once I control for the attention threshold, I do not find any evidence that attention changes within regime.

I develop a New Keynesian model with shifting degrees of attention to understand the monetary policy implications of the inflation attention threshold. Following a sufficiently large inflationary shock, attention increases and this heightened attention then renders inflation expectations more sensitive to the initial increase in inflation which fuels back into even higher and more persistent inflation. When calibrating the model to match the empirical findings, it generates inflation and inflation expectation dynamics mirroring the ones recently observed in the US, and therefore, provides a potential explanation for the observed inflation dynamics during the recent inflation

¹For example, Federal Reserve Chair Jerome H. Powell said in his 2021 Jackson Hole speech that inflation concerns are “likely to prove temporary” (see <https://www.federalreserve.gov/newsevents/speech/powell20210827a.htm>).

surge. Absent the increase in attention, inflation would peak at a lower level and come down more quickly. Accounting for varying attention levels also matters for the model’s normative predictions. When monetary policy follows simple Taylor rules, the associated welfare losses are substantially larger than under optimal policy rules. The reason is that simple Taylor rules induce frequent and long-lasting episodes in which attention is heightened which leads to large welfare losses, not only because these episodes increase the overall volatility of inflation, but—due to the asymmetry of the attention threshold—also the average level of inflation.

After the Great Inflation period in the 1970s and 1980s, as inflation decreased, attention to inflation decreased with it and was at a very low level before the outbreak of the pandemic (Pfäuti, 2023).² When inflation increases, however, attention to inflation increases as well (Bracha and Tang, 2023). Korenok et al. (2022) find, for a large number of countries, that people’s attention to inflation increases with inflation only after inflation exceeds a certain threshold. Using Google search behavior for the period from 2004 to 2022, they estimate this attention threshold to be at an annualized inflation rate of 3.55% for the US. Cavallo et al. (2017) show that survey respondents in high-inflation environments (Argentina) respond less to information about inflation than in low-inflation environments (United States), which is consistent with higher attention to inflation in high-inflation environments. Weber et al. (2023) confirm, using a range of randomized control trials spanning over several years and different countries, that attention of households and firms is indeed higher in times of high inflation. Kroner (2023) focuses on financial markets and shows that attention—measured as asset price responses to inflation news—is higher in times of high inflation. My key innovation relative to these papers is that I provide estimates of the attention threshold and the attention levels in the two regimes in a way that directly maps into otherwise standard macroeconomic models.

To do so, I extend the method I proposed in Pfäuti (2023) by allowing for an attention threshold. Attention captures how strongly agents update their inflation expectations in response to forecast errors. I allow for different degrees of attention depending on whether inflation is above or below a certain attention threshold. I then jointly estimate the attention threshold as well as the attention levels in the two regimes using survey data from the University of Michigan’s Survey of Consumers for the period 1978 to 2023. I estimate an inflation attention threshold of 4.0%, and that attention is about 0.18 when inflation is below the threshold and more than doubles to 0.36 when inflation is above the threshold. An attention level of 0.36

²For the years before the recent inflation surge, Candia et al. (2021, 2023) and Coibion et al. (2020) show that U.S. firms as well as households are usually poorly informed about and inattentive to inflation and monetary policy (see also Weber et al. (2022) for a recent review). Goldstein (2023) also finds that inattention varies over time. Different forms of changing attention are considered, e.g., in Kim and Binder (2023) who examine how repeat participation in surveys that ask about inflation expectations may lead to higher attention to inflation, or in Flynn and Sastry (2022) who show—by using a textual proxy for firms’ attention toward macroeconomic conditions—that attention is counter-cyclical, or Gallegos (2023) who shows that firms’ less sluggish inflation expectations after the Great inflation period offer a potential explanation for the decrease in the persistence of inflation as found, e.g., in Benati (2008).

means that following a 1pp forecast error, agents increase their inflation expectations by 0.36pp. Using the Survey of Consumer Expectations from the New York Fed for the period 2013-2023, I obtain similar results: attention in times when inflation was below 4% is about 0.21 and increases to 0.4 when inflation is above 4%. Once I control for the attention threshold, I do not find any evidence that attention increases with inflation within regime.

To understand the implications of these empirical findings for inflation, inflation expectations and monetary policy, I extend the standard New Keynesian model by introducing an attention threshold and changing levels of attention. Firms and households pay limited attention to inflation, as acquiring and processing information about inflation is costly. The main novelty is that I introduce the aforementioned attention threshold: once inflation exceeds this threshold, it becomes optimal for households and firms to pay more attention to inflation. The attention shift occurs because the cost of information decreases, for example, because news media start reporting more prominently about inflation when inflation exceeds the threshold. I show that news coverage of inflation in the US indeed doubles when inflation is higher than 4% from periods in which inflation is below this threshold. For the most recent inflationary period, I find an even larger increase in news coverage of inflation from the end of 2020 to the peak in late 2022. Consistent with my findings, [Bracha and Tang \(2023\)](#) show that higher inflation rates lead to more media reporting about inflation, [Larsen et al. \(2021\)](#) find that news media coverage predicts households' inflation expectations, and [Lamla and Lein \(2014\)](#) show that more intensive news reporting about inflation improves the accuracy of consumers' inflation expectations.

I then calibrate the model to match the estimated attention levels in the two regimes, the attention threshold, and the frequency of staying in the high-attention regime. Accounting for the attention threshold substantially alters the dynamics of inflation and inflation expectations. Following an inflationary supply shock that pushes inflation just above the attention threshold, inflation keeps on increasing through the endogenous attention increase. As attention increases, agents now update their inflation expectations more strongly in response to the increase in inflation. These higher inflation expectations then fuel further inflation increases, leading to higher inflation expectations, and so on. As the shock dies out, inflation starts to decline after some time. However, inflation may remain elevated for a substantial period of time, because once it falls back below the threshold, people pay less attention to inflation, and hence, only slowly revise their inflation expectations downward. As their prior expectations are now relatively high due to the experienced high inflation period, expectations remain persistently high, leading to a slow decline of actual inflation. I show that these patterns of inflation and inflation expectations are consistent with the recent inflation surge in the US: inflation followed a hump-shaped pattern and inflation expectations initially fell short of the actual inflation surge, but then surpassed it once inflation started to decline. The inflation dynamics my model produces are further consistent with the empirical findings in [Blanco et al. \(2022\)](#) who document, for a number of countries and different periods of inflation surges, that (i) inflation

stays persistently high after the initial surge and (ii) that short-run inflation expectations initially fall short of actual inflation.

The attention threshold further induces a state dependency. The inflationary effects of cost-push shocks are larger when the shock hits at a time of loose monetary policy, and the total effect is in fact substantially larger than the sum of the two effects in isolation. This state dependency arises due to the heightened attention which—together with loose monetary policy—renders inflation more responsive to cost-push shocks. These state-dependent effects are absent in the economy with a constant level of attention or with full-information rational expectations.

Thus, large shocks—supply or demand—may push the economy into the high-attention regime and therefore, lead to high and persistent inflation rates. However, also relatively small shocks to demand and supply *simultaneously* can trigger large inflation surges when the combination of the two shocks pushes inflation above the inflation attention threshold. Thus, the model with the attention threshold may predict high-inflation periods at times the model absent the attention threshold or the one under rational expectations would predict a substantially smaller and more transitory increase in inflation.

Thus, accounting for sudden changes in the public’s attention to inflation may help to understand how inflation surges can arise and sustain themselves for a prolonged period, and offers a new perspective on the last two high-inflation episodes in the US. The US economy was in the high-attention regime during the Great Inflation period and recently entered it again in 2021. [Reis \(2022\)](#) shows that inflation expectations have become unanchored in the Great Inflation period, and [Hilscher et al. \(2022\)](#) find that in 2021 the risk of persistently high inflation in the US increased significantly. Both of these findings are consistent with the predictions of my model: in the high-attention regime, inflation expectations are more sensitive to inflation, i.e., they appear to be less anchored, and the regime switch to the high-attention regime increases the risk of persistently high inflation due to the higher sensitivity of inflation expectations.

Taking the attention threshold into account also matters for the model’s normative implications. When monetary policy sets the nominal interest rate following a simple Taylor rule, the associated welfare losses are significantly larger compared to optimal policy rules or a strict inflation targeting rule (similar to the findings in [Gáti \(2022\)](#)). The reason is that under a standard calibration for the Taylor rule, the economy spends a substantial amount of time in the high-attention regime in which inflation is high and volatile. Due to the asymmetry of the attention threshold—attention increases when inflation is particularly high but remains constant when inflation is particularly low—the *average level of inflation* is higher when the economy is in the high-attention regime frequently. This is in stark contrast to the model without the threshold or the one with fully-informed rational agents, where inflation fluctuates symmetrically around zero. This asymmetry of the attention threshold also offers a potential explanation for why we did not observe long lasting deflationary periods, for example, during the Great financial crisis ([Coibion and Gorodnichenko, 2015](#)). In fact, annualized quarter-on-quarter CPI inflation

was negative in only 7% of the time between 1978 and 2023, whereas it exceeded the attention threshold of 4.0% about 30% of the time during the same period.

Taylor rules with a high degree of interest-rate smoothing lead to especially large welfare losses in the model with two attention regimes—in contrast to the model under rational expectations in which interest-rate smoothing increases welfare. Interest-rate smoothing introduces additional persistence, such that the periods in the high-attention regime last longer and thus, the average level of inflation as well as its volatility increase. The monetary authority can avoid these large welfare losses by implementing *some* optimal policy rule: even when monetary policy ignores that private agents pay limited attention to inflation and implements the policy that would be optimal with rational agents, the welfare losses are small and in fact close to the ones in the model with rational agents.

The theoretical insights in this paper contribute to a growing literature on the role of changes in attention and the degree of anchoring of inflation expectations. [Evans and Ramey \(1995\)](#) propose a model in which agents choose how forward-looking they want to be and show that inflation remains stable when agents are not very forward looking.³ [Gáti \(2022\)](#) studies how changes in the degree of anchoring of long-run inflation expectations affect the optimal monetary policy. In her paper, anchoring changes continuously, thus, can also jump in response to large shocks. In contrast, attention in my model only changes across but not within regime, which is consistent with what I find in the data, namely, that attention is constant within regime. [Carvalho et al. \(2022\)](#) focus on discrete changes in anchoring of long-run inflation expectations, but do so in a partial equilibrium setting. In [Pfäuti \(2023\)](#), I study the implications of low attention for optimal monetary policy when the zero lower bound poses a constraint to monetary policy. In contrast to that paper, I allow here for an inflation attention threshold, whereas [Pfäuti \(2023\)](#) compares economies with different—but time-invariant—degrees of attention. [Hazell et al. \(2022\)](#) show that the greater inflation stability after 1990 is mostly due to more firmly-anchored long-run inflation expectations. Related, [Jørgensen and Lansing \(2023\)](#) find that more strongly anchored inflation expectations can help explain changes in the reduced form Phillips curve, and [Afrouzi and Yang \(2020\)](#) show that in a model of dynamic rational inattention the stance of monetary policy affects the slope of the Phillips curve. The main contribution of the present paper to this literature is that I incorporate and estimate an inflation attention threshold into a general equilibrium model and show how accounting for such a threshold can help us better understand how inflation surges may happen and how they can sustain themselves, as well as the monetary policy implications that follow from such attention-fueled inflation surges.

³Focusing on endogenous but constant attention, [Mackowiak and Wiederholt \(2009\)](#) show how firms set prices optimally information is costly, and [Paciello and Wiederholt \(2014\)](#) how the optimal monetary policy changes when firm managers decide optimally how much attention they want to pay to aggregate conditions compared to a setting where their attention is exogenous (see also [Sims \(2003, 2010\)](#)).

Outline. Section 2 explains how I estimate attention and the attention threshold in the data and presents my empirical findings. I introduce the New Keynesian model with limited attention in Section 3, and the model’s positive results in Section 4. In Section 5, I derive the model’s normative implications and Section 6 concludes.

2 Attention and the Inflation Attention Threshold

In this section, I show how I estimate the inflation attention threshold and the different degrees of attention in the two attention regimes, and present the estimation results. To arrive at these results, I extend the method I develop in Pfäuti (2023) by allowing for an inflation-attention threshold.⁴ Details on the optimal attention choice are relegated to Appendix A.

Agents hold a subjective model of how inflation evolves.⁵ In particular, the agent believes that (demeaned) inflation tomorrow, π' , depends on (demeaned) inflation today, π , as follows

$$\pi' = \rho_\pi \pi + \nu,$$

where $\rho_\pi \in [0, 1]$ denotes the perceived persistence of inflation and $\nu \sim i.i.N.(0, \sigma_\nu^2)$. The agent faces a trade off how attentive she wants to be. On the one hand, making forecast errors is costly, but on the other hand, current inflation is unobservable and acquiring and processing information about it is also costly. I assume that the loss arising from making forecasting mistakes is quadratic with a scaling factor χ . The cost of information is linear in mutual information $I(\pi; s)$, i.e., the expected reduction in entropy of π due to observing the signal s , with a scaling factor $\frac{1}{\lambda_r}$:

$$C(f) = \frac{1}{\lambda_r} I(\pi; s) = \frac{1}{\lambda_r} (H(\pi) - E[H(\pi|s)]),$$

where $H(x) = -\int f(x) \log(f(x)) dx$ is the entropy of x and $\frac{1}{\lambda_r}$ measures the cost of information.

I interpret changes in λ_r as coming from changes in how much the news media report about inflation and I allow λ_r to differ across attention regimes $r \in \{L, H\}$ with $\lambda_H > \lambda_L$, assuming that there is more news reporting about inflation in the H regime. In particular, when inflation exceeds a certain threshold, news media report more prominently about inflation in the next period, thus, lowering the cost of agents to acquire information about inflation, $\frac{1}{\lambda_r}$. Agents do not foresee these changes in news reporting and form their expectations within regime r as if λ_r would remain unchanged.

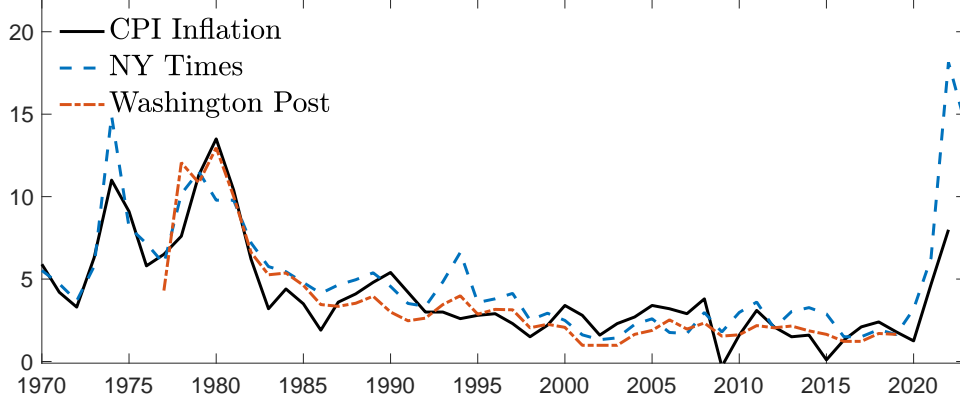
Consistent with the interpretation that changes in λ_r arise due to changes in media reporting, I find that news coverage of inflation is indeed higher in times of high inflation. Figure 2 shows

⁴The method proposed in Pfäuti (2023) builds on Mackowiak et al. (2023) and Vellekoop and Wiederholt (2019).

⁵Agents’ subjective model need not necessarily be consistent with the actual law of motion (see Andre et al. (2022); Macaulay (2022) for empirical evidence that the subjective models agents hold may not necessarily be consistent with the actual behavior of the economy or with experts’ subjective models).

the frequency of the word “inflation” in the New York Times (1970 to July 2023, blue-dashed line) and the Washington Post (1977-2019, red-dashed-dotted line), together with annual CPI inflation (black-solid line). There is a very strong positive correlation between inflation and news coverage of inflation (the correlation between CPI inflation and the two news-coverage series is 0.86 for the New York Times and 0.90 for the Washington Post).

Figure 2: News coverage of inflation is higher in times of high inflation



Notes: The black-solid line shows the annual CPI inflation rate for the US since 1970, the blue-dashed line and the red-dashed-dotted line show the frequency of the word inflation in the New York Times and the Washington post, respectively.

Consistent with this, [Bracha and Tang \(2023\)](#) show that higher inflation rates indeed lead to more media reporting about inflation, and [Lamla and Lein \(2014\)](#) find that more intensive news reporting about inflation improves the accuracy of consumers’ inflation expectations, consistent with agents being more attentive. [Schmidt et al. \(2023\)](#) show that during episodes of intensive newspaper coverage of inflation, news reporting has strong effects on inflation expectations but not during other episodes. [Larsen et al. \(2021\)](#) find that news media coverage predicts households’ inflation expectations, and [Nimark and Pitschner \(2019\)](#) show that major events (such as strong inflation increases) lead to a shift in the news focus towards these events.

Assuming a normal prior, the optimal signal the agent acquires then takes the form

$$s = \pi + \varepsilon,$$

with $\varepsilon \sim i.i.N.(0, \sigma_\varepsilon^2)$ (see [Matějka and McKay \(2015\)](#)). Conditional on the signal, and the perceived law of motion, the optimal expectation about inflation tomorrow is given by $\tilde{E}[\pi'|s] = \rho_\pi E[\pi|s]$. Bayesian updating implies

$$\tilde{E}[\pi'|s] = \rho_\pi (1 - \gamma_{\pi,r}) \hat{\pi} + \rho_\pi \gamma_{\pi,r} s, \quad (1)$$

where $\gamma_{\pi,r} = 1 - \frac{\sigma_{\pi|s,r}^2}{\sigma_\pi^2} \in [0, 1]$ measures attention to inflation in regime r , and $\hat{\pi}$ denotes the prior mean of π .

Solving for the optimal level of attention $\gamma_{\pi,r}$ yields

$$\gamma_{\pi,r} = \max \left\{ 0, 1 - \frac{\frac{1}{\lambda_r}}{2\chi\rho_\pi^2\sigma_\pi^2} \right\}. \quad (2)$$

The expression (2) for the optimal attention level shows that when news reporting λ_r increases, attention increases. Hence, when there is more news about inflation in regime H , reflected in a lower cost of information $\frac{1}{\lambda_H} < \frac{1}{\lambda_L}$, attention is higher: $\gamma_{\pi,H} > \gamma_{\pi,L}$. In the estimation later on, I will test this prediction of the theory rather than impose it.

Dynamics. To estimate the attention threshold and the degrees of attention in the data, I extend the law of motion of inflation expectations, equation (1), to a dynamic setup. The agent believes that inflation π follows

$$\pi_t = (1 - \rho_\pi)\underline{\pi} + \rho_\pi\pi_{t-1} + \nu_t,$$

where $\underline{\pi}$ is the agent's long-run belief about inflation and ρ_π is the perceived persistence of inflation. I assume that the error term ν_t is normally distributed with mean zero and variance σ_ν^2 . The agent receives a signal about inflation of the form

$$s_t = \pi_t + \varepsilon_{r,t},$$

where the noise $\varepsilon_{r,t}$ is assumed to be normally distributed with variance $\sigma_{r,\varepsilon}^2$ which is lower in the H regime, reflecting a lower cost of information and thus, a higher attention.

Given these assumptions, it follows from the (steady state) Kalman filter that optimal updating in regime r is given by

$$\tilde{E}_t\pi_{t+1} = (1 - \rho_\pi)\underline{\pi} + \rho_\pi\tilde{E}_{t-1}\pi_t + \rho_\pi\gamma_{\pi,r} \left(\pi_t - \tilde{E}_{t-1}\pi_t \right) + u_t. \quad (3)$$

From equation (3), we observe that lower attention, i.e., a lower $\gamma_{\pi,r}$, implies that the agent updates her expectations to a given forecast error, $(\pi_t - \tilde{E}_{t-1}\pi_t)$, less strongly. Lower attention is reflected in more noisy signals, and more noise means the agent puts less weight on these signals and more weight on the prior.

Across the two regimes, inflation expectations then follow the law of motion:

$$\tilde{E}_t\pi_{t+1} = \begin{cases} (1 - \rho_\pi)\underline{\pi} + \rho_\pi\tilde{E}_{t-1}\pi_t + \rho_\pi\gamma_{\pi,L} \left(\pi_t - \tilde{E}_{t-1}\pi_t \right) + u_t, & \text{when } \pi_{t-1} < \bar{\pi} \\ (1 - \rho_\pi)\underline{\pi} + \rho_\pi\tilde{E}_{t-1}\pi_t + \rho_\pi\gamma_{\pi,H} \left(\pi_t - \tilde{E}_{t-1}\pi_t \right) + u_t, & \text{when } \pi_{t-1} > \bar{\pi}. \end{cases} \quad (4)$$

The theory thus predicts that whenever inflation in the previous period exceeded the inflation attention threshold, $\pi_{t-1} > \bar{\pi}$, the increase in news coverage of inflation $\lambda_H > \lambda_L$ leads to an

increase in agents' attention to inflation: $\gamma_{\pi,H} > \gamma_{\pi,L}$.

2.1 Estimating attention and the attention threshold

In order to estimate the two attention levels $\gamma_{\pi,r}$ for $r \in \{L, H\}$, as well as the attention threshold $\bar{\pi}$, I first rewrite equation (3) as follows:

$$\tilde{E}_t \pi_{t+1} = \beta_0 + \beta_1 \tilde{E}_{t-1} \pi_t + \beta_2 \left(\pi_t - \tilde{E}_{t-1} \pi_t \right) + \epsilon_t, \quad (5)$$

where $\beta_0 = (1 - \rho_\pi) \underline{\pi}$ denotes the intercept, $\beta_1 = \rho_\pi$ captures the perceived persistence, and $\frac{\beta_2}{\beta_1} = \gamma_\pi$.

As my measure of inflation expectations, I rely on the Survey of Consumers from the University of Michigan. For my baseline specification, I use average and median household inflation expectations for the period 1978-2023. For the period 2013-2023, I also use individual inflation expectations from the Survey of Consumer Expectations of the New York Fed. Even though I focus on inflation and inflation expectations over one quarter, I use monthly data to increase the number of observations. As a robustness check, however, I also consider expectations at quarterly frequency. One advantage of using quarterly observations is that the Survey of Consumers provides that data going back to 1960Q2. Both surveys ask consumers for their price growth expectations one-year ahead: $\tilde{E}_t \pi_{t+12}$. I compute one-quarter-ahead forecasts by dividing the one-year-ahead forecasts by four: $\tilde{E}_t \pi_{t+3} \equiv \frac{\tilde{E}_t \pi_{t+12}}{4}$. Using $\tilde{E}_t \pi_{t+3} \equiv \left[\left(1 + \tilde{E}_t \pi_{t+12} \right)^{\frac{1}{4}} - 1 \right]$ instead does not change the results. Computing one-quarter-ahead expectations allows me to compare the results directly to the model which is calibrated at quarterly frequency. For actual inflation, I use the monthly CPI inflation rate from the FRED database and to be consistent with the model, I focus on quarter-on-quarter inflation: $\pi_t \equiv \frac{P_t - P_{t-3}}{P_{t-3}}$. I estimate equation (5) using a linear regression model but I allow the coefficients to differ across regimes, whereas the threshold-defining variable is the lagged inflation rate. I allow all coefficients to potentially differ across regimes. The threshold regression is then given by

$$\begin{aligned} \tilde{E}_t \pi_{t+3} = & \mathbb{1}_{\pi_{t-1} \leq \bar{\pi}} \left(\beta_{0,L} + \beta_{1,L} \tilde{E}_{t-3} \pi_t + \beta_{2,L} \left(\pi_t - \tilde{E}_{t-3} \pi_t \right) \right) \\ & + (1 - \mathbb{1}_{\pi_{t-1} \leq \bar{\pi}}) \left(\beta_{0,H} + \beta_{1,H} \tilde{E}_{t-3} \pi_t + \beta_{2,H} \left(\pi_t - \tilde{E}_{t-3} \pi_t \right) \right) + \epsilon_t, \end{aligned} \quad (6)$$

where $\mathbb{1}_{\pi_{t-1} \leq \bar{\pi}}$ is the indicator function that equals one when inflation in the previous month π_{t-1} was below the threshold $\bar{\pi}$ and zero otherwise. The threshold value is then estimated by minimizing the sum of squared residuals obtained for all possible thresholds (see, e.g., [Gonzalo and Pitarakis \(2002\)](#) and [Hansen \(2011\)](#) and the references therein). Note, that I do not impose that the attention level in the regime in which inflation is above the threshold needs to be higher than in the regime with inflation below the threshold. Therefore, the results will serve as a test

of the theory outlined above.

Estimation results. Table 1 shows the estimation results. For the baseline specification where I use mean expectations, I estimate an attention threshold of 3.98%.⁶ Attention in the regime in which inflation is below this threshold is $\gamma_{\pi,L} = 0.18$, and attention in the high-inflation regime is equal to $\gamma_{\pi,H} = 0.36$. Thus, attention in the high-inflation regime is higher than in the low-inflation regime, as predicted by the theory. In fact, attention is twice as high in the regime with inflation above the threshold compared to attention in the regime with inflation below the threshold. I therefore refer to the regime with inflation above the threshold as the *high-attention regime*. The threshold value of 3.98% implies that the US economy spent about 31% of the time in the period 1978-2023 in the high-attention regime. Most recently, inflation exceeded this threshold in early 2021. Google searches also started to increase shortly after that time (see Figure 1 in the Introduction). The results for median expectations are similar. The main difference is that attention tends to be somewhat lower in this case and the attention threshold higher. In both cases, the null hypothesis that the two attention levels across regimes are equal is clearly rejected (p -values of 0.000 in both cases, see last column).

The last two rows in Table 1 show the results when using observations at quarterly frequency for the period 1960Q2-2023Q2. We see that the estimated threshold is somewhat lower at 3.21%. The estimated attention levels within regime are very similar to my baseline specification and again, the difference in attention across regimes is highly statistically significant.

The estimated perceived persistence coefficients for my baseline specification, $\beta_{1,L}$ and $\beta_{1,H}$, are $\beta_{1,L} = 0.74$ and $\beta_{1,H} = 0.87$, respectively. These estimates are within the 90% confidence intervals of Benati (2008) who estimates inflation persistence in the Great Inflation period for the US to be equal to 0.77, and 0.49 for the Post-Volcker stabilization period. Thus, inflation persistence is higher in times of higher inflation—or, in my case, in times of higher attention.

Table 1: Estimated attention levels and the attention threshold

	Threshold $\bar{\pi}$	Low Attention $\hat{\gamma}_{\pi,L}$	High Attention $\hat{\gamma}_{\pi,H}$	p -val. $\gamma_{\pi,L} = \gamma_{\pi,H}$
Mean exp.	3.98%	0.18	0.36	0.000
s.e.		(0.013)	(0.037)	
Median exp.	4.41%	0.16	0.23	0.000
s.e.		(0.013)	(0.028)	
Quarterly freq.	3.21%	0.14	0.38	0.000
s.e.		(0.033)	(0.076)	

Notes: This table shows the results from regression (6), where $\bar{\pi}$ denotes the estimated threshold, $\hat{\gamma}_{\pi,L}$ and $\hat{\gamma}_{\pi,H}$ the estimated attention levels when inflation is below or above the threshold, respectively. The last column shows the p -value for the null hypothesis that the two attention levels are equal. The upper half of the table presents the results when using average expectations, and the lower half when using median expectations. Standard errors are robust with respect to heteroskedasticity.

⁶Using the Bayesian Information Criterion or the Hannan–Quinn information criterion to select the numbers of thresholds, the model prefers the specification with only one threshold.

When I use the current inflation rate rather than the lagged inflation rate as the threshold-defining variable, I find very similar results. Attention is 0.18 when inflation is below 4% and increases to 0.36 when inflation is above 4%.

When using individual consumer inflation expectations from the Survey of Consumer Expectations, I obtain similar results even though this survey is only available since 2013. When lagged inflation is below the threshold of 4%, I estimate an attention level of 0.21. When inflation was above the threshold, attention doubles to 0.4, and I reject the null hypothesis that the two estimates are equal (p -value of 0.000).

News coverage of inflation is also substantially higher in times inflation is above the attention threshold. I find that the average frequency of the word “inflation” is 2.7 times as high for the New York Times and 2.9 times as high for the Washington Post when CPI inflation is above 4% in that year compared to years in which CPI inflation is below 4%.

My estimated attention threshold is somewhat higher than the one estimated in [Korenok et al. \(2022\)](#). They estimate a threshold for the US of 3.55% using Google search data. One reason why I find a higher threshold could be that people might start to google for information about inflation at lower levels of inflation already, but only start incorporating that information in their expectations once inflation further increases. Additionally, I focus on the time starting in 1978, whereas their sample is restricted to 2004-2022 because the Google search data is not available for years before 2004. They also consider news coverage of inflation, and in that case, they estimate a threshold between 3.77-3.94%, which is very close to the threshold I estimate. Overall my results align with their findings, and my approach provides a quantification of the threshold and the attention levels in a way that is directly applicable to standard macroeconomic models, as we will see in the next section.

Attention changes within regime. In Table 1 we saw that attention increases strongly when inflation exceeds the threshold of 4%. But what about changes within regime? To look at this, I estimate a whole time series of attention. In particular, I estimate equation (5) using a rolling-window approach, where each window has a length of one year. Let me denote the estimated time series by $\hat{\gamma}_{\pi,t}$. Similarly, I compute the window-specific average of the monthly quarter-on-quarter inflation rate, $\bar{\pi}_t$. To then test whether attention within regime is higher when inflation is higher, I estimate the following regression

$$\hat{\gamma}_{\pi,t} = \delta_0 + \delta_1 \mathbb{1}_{\bar{\pi}_t \geq 4} + \delta_2 \bar{\pi}_t + \delta_3 \mathbb{1}_{\bar{\pi}_t \geq 4} \bar{\pi}_t + \varepsilon_t, \quad (7)$$

where $\mathbb{1}_{\bar{\pi}_t \geq 4}$ is an indicator that equals one when average inflation in period t is above 4% and zero otherwise. Thus, δ_1 tells us the difference of attention across regimes, δ_2 the effect of inflation on attention and δ_3 the additional effect of inflation on attention in the high-inflation regime.⁷

⁷I further impose that attention needs to be between 0 and 1, consistent with the theory.

The first row in Table 2 shows the results. We see that attention is significantly higher in the higher-inflation regime, as indicated by the estimate of δ_1 . Yet, inflation does not have any additional significant effect on attention when accounting for the threshold, as depicted by the last two columns.

As a robustness check, I also estimate

$$\hat{\gamma}_{\pi,t} = \delta_0 + \delta_1 \mathbb{1}_{\pi_t \geq 4} + \delta_2 \pi_{t-1} + \delta_3 \mathbb{1}_{\pi_{t-1} \geq 4} \pi_{t-1} + \varepsilon_t, \quad (8)$$

where I use lagged inflation as the independent variable rather than average inflation. The lower part of Table 2 shows that the results are robust. Once I control for the inflation threshold, I do not find any evidence for a positive relationship between inflation and attention on top of the attention shift across regimes.

Table 2: Attention changes within regime

	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$
Regression (7)	0.393**	0.053	-0.079
s.e.	(0.192)	(0.047)	(0.051)
Regression (8)	0.119*	-0.010	0.010
s.e.	(0.0641)	(0.0141)	(0.0141)

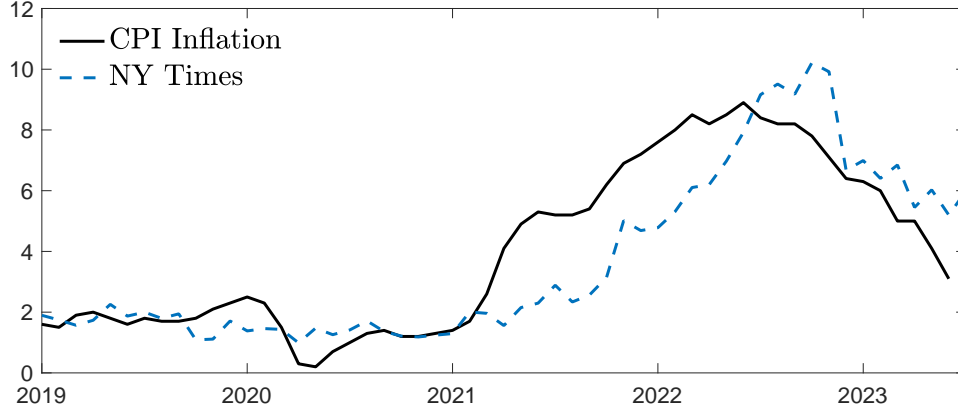
Notes: This table shows the results from regression (7) (upper half of the table) and from regression (8) (lower half of the table). Standard errors are robust with respect to heteroskedasticity and serial correlation (Newey and West (1987) with 12 lags). Significance levels: *: p -value < 0.1, **: p -value < 0.05, ***: p -value < 0.01.

Covid. When focusing on the sample from 2017 until 2023, I estimate an attention threshold of 3.63%, and attention levels of $\gamma_{\pi,L} = 0.03$ and $\gamma_{\pi,H} = 0.25$. The very low attention level before the outbreak of the Covid-19 pandemic confirms the findings in Pfäuti (2023) that attention to inflation was at a historical low at the time. When imposing the threshold at 4% (rather than estimating it), attention in the higher regime equals 0.28 and 0.04 in the low regime.

Figure 3 shows news coverage of inflation in the New York Times for the period 2019 until July 2023 at monthly frequency (blue-dashed line) together with monthly year-on-year CPI inflation. The figure shows the strong positive correlation of 0.85 between inflation and news coverage, and that at the peak, news coverage of inflation quintupled from its pre-pandemic level. Consistent with the theory, news coverage lags inflation slightly (the correlation between news coverage and *lagged* inflation is even higher than the one with current inflation, 0.9 instead of 0.85). The correlation of news coverage with Google searches is 0.94 for the period 2017-2023, further supporting the assumption that news coverage is highly correlated with the public’s attention to inflation.

Heterogeneity. So far, I mainly focused on average (and median) expectations. I now test whether the attention threshold and the different attention levels depend on people’s gender or

Figure 3: News coverage of inflation during Covid



Notes: The black-solid line shows the monthly year-on-year CPI inflation rate for the US and the blue-dashed line the frequency of the word inflation in the New York Times, normalized to have the same standard deviation as CPI inflation.

their age.⁸ As documented in [D’Acunto et al. \(2021\)](#), gender and gender roles (e.g., with respect to grocery shopping) play a big role in explaining differences in how men and women form their inflation expectations. I find that men have a higher attention threshold than women (4.4% vs. 3.9%), indicating that women increase their attention somewhat earlier than men. This might be explained by the fact that women are more likely to go grocery shopping than men ([D’Acunto et al., 2021](#)) and therefore, experience price changes more directly than men. I further find that the threshold for younger people (aged 18-34) is lower than for older people (4.44% vs. 6.8% for people aged between 35 and 54, and 5.7% for people aged older than 55), but that their attention levels tend to be lower overall (0.21 vs. 0.24 below the threshold and 0.41 vs. 0.74 above it).

3 A Monetary Model with Varying Attention

This section presents the monetary model with limited attention, the key equilibrium equations, and how I calibrate the model in order to match the empirical findings from the previous section. The model builds on the textbook New Keynesian model ([Woodford, 2003](#); [Galí, 2015](#)) but households and firms pay only limited attention to inflation and the output gap.

⁸There is a vast literature documenting heterogeneity in inflation expectations, see, e.g., [D’Acunto et al. \(2019\)](#); [Broer et al. \(2021\)](#); [Pfäuti and Seyrich \(2023\)](#); [D’Acunto et al. \(2023\)](#); [Pedemonte et al. \(2023\)](#); [Weber et al. \(2022\)](#); [Roth et al. \(2023\)](#); [Nord \(2022\)](#); [Meichtry \(2022\)](#), for recent contributions.

3.1 Households

There is a representative household obtaining utility from consumption and disutility from working, with lifetime utility

$$\tilde{E}_0 \sum_{t=0}^{\infty} \beta^t Z_t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \Xi \frac{H_t^{1+\nu}}{1+\nu} \right], \quad (9)$$

where C_t is consumption of the final good, H_t is hours worked, β is the household's time discount factor, and \tilde{E}_t denotes the household's subjective expectations operator based on information available in period t . Z_t are exogenous preference shocks. The parameters σ and ν pin down the relative risk aversion and the inverse Frisch labor elasticity, respectively. Ξ is the utility weight on hours worked.

Households maximize their lifetime utility subject to the flow budget constraints

$$C_t + B_t = w_t H_t + \frac{1+i_{t-1}}{1+\pi_t} B_{t-1} + \frac{T_t}{P_t}, \quad \text{for all } t, \quad (10)$$

where B_t is the real value of government bonds, w_t the real wage, π_t is the net inflation rate, and i_t the nominal interest rate. T_t denotes lump-sum taxes and transfers from the government.

Maximizing (9) subject to (10) yields the Euler equation

$$Z_t C_t^{-\sigma} = \beta(1+i_t) \tilde{E}_t \left[Z_{t+1} C_{t+1}^{-\sigma} \frac{1}{1+\pi_{t+1}} \right], \quad (11)$$

and the labor-leisure condition

$$w_t = \Xi H_t^\nu C_t^\sigma. \quad (12)$$

3.2 Firms

The firm sector is held by risk-neutral managers that discount future profits by β and they have a mass of zero, such that their consumption is 0 and all their profits go to the households, as in [Bayer et al. \(2022\)](#). Firm managers hold the same subjective expectations as households, consistent with [McClure et al. \(2022\)](#) who show that managers and non-managers hold similar average inflation and unemployment expectations and respond similarly to information treatments.⁹

Final good producer. There is a representative final good producer that aggregates the intermediate goods $Y_t(j)$ to a final good Y_t , according to

$$Y_t = \left(\int_0^1 Y_t(j)^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}}, \quad (13)$$

⁹[McClure et al. \(2022\)](#) further show that managers' expectations indeed affect their economic decisions.

with $\epsilon > 1$. Nominal profits are given by $P_t \left(\int_0^1 Y_t(j)^{\frac{\epsilon-1}{\epsilon}} dj \right)^{\frac{\epsilon}{\epsilon-1}} - \int_0^1 P_t(j) Y_t(j) dj$, and profit maximization gives rise to the demand for each variety j :

$$Y_t(j) = \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon} Y_t. \quad (14)$$

Thus, demand for variety j is a function of its relative price, the price elasticity of demand ϵ and aggregate output Y_t . The aggregate price level is given by

$$P_t = \left(\int_0^1 P_t(j)^{1-\epsilon} dj \right)^{\frac{1}{1-\epsilon}}. \quad (15)$$

Intermediate producers. Intermediate producer of variety j produces output $Y_t(j)$ using labor $H_t(j)$ as its only input

$$Y_t(j) = H_t(j). \quad (16)$$

All intermediate producers pay the same wage w_t and a sales tax (or subsidy) τ_t , which in steady state is set such that profits in steady state are 0. These taxes are given back to firms in a lump-sum fashion, denoted $t_t^F(j)$. Taxes are assumed to be constant in the efficient economy, i.e., absent price rigidities, but fluctuate around its steady state in the economy with price rigidities in order to give rise to exogenous cost-push shocks.

Firms are subject to a [Rotemberg \(1982\)](#) price-adjustment friction. As in [Hagedorn et al. \(2019\)](#), I assume that these adjustment costs are *as-if* adjustment costs, that is, firms set prices as if they had to pay the adjustment cost, but they do not actually have to pay these costs. This rules out “price-adjustment booms” that might otherwise arise. Their per-period profits (in real terms) are given by

$$(1 - \tau_t) P_t(j) \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon} \frac{Y_t}{P_t} - w_t H_t(j) - \frac{\psi}{2} \left(\frac{P_t(j)}{P_{t-1}(j)} - 1 \right)^2 Y_t + t_t^F(j), \quad (17)$$

where $\psi \geq 0$ captures the price-adjustment cost parameter. They set prices to maximize

$$Profits_0(j) = \tilde{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(1 - \tau_t) P_t(j) \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon} \frac{Y_t}{P_t} - w_t H_t(j) - \frac{\psi}{2} \left(\frac{P_t(j)}{P_{t-1}(j)} - 1 \right)^2 Y_t + t_t^F(j) \right].$$

As I show in [Appendix C](#), this then gives rise to the linearized New Keynesian Phillips Curve

$$\pi_t = \underbrace{\frac{\epsilon - 1}{\psi} (\sigma + \nu)}_{\equiv \kappa} \hat{x}_t + \beta \tilde{E}_t \pi_{t+1} + \underbrace{\frac{(\epsilon - 1) \mathcal{T}}{\psi}}_{\equiv u_t} \hat{\tau}_t, \quad (18)$$

where κ measures the sensitivity of inflation, π_t , to changes in the output gap, \hat{x}_t , which is the log deviation of output from its efficient counterpart that would prevail under flexible prices. u_t are

supply (or cost-push) shocks.

Government. The government imposes a sales tax τ_t on sales of intermediate goods, issues nominal bonds, and pays lump-sum taxes and transfers T_t to households and $t_t^F(j)$ to firms. The real government budget constraint is given by

$$B_t = B_{t-1} \frac{1 + i_{t-1}}{\Pi_t} + \frac{T_t}{P_t} - \tau Y_t + t_t^f.$$

Lump-sum taxes and transfers are set such that they keep real government debt constant at the initial level B_{-1}/P_{-1} , which I set to zero.

The monetary authority sets the nominal interest rate, according to a (linearized) Taylor rule:

$$\tilde{i}_t = \rho_i \tilde{i}_{t-1} + (1 - \rho_i) (\phi_\pi \pi_t + \phi_x \hat{x}_t), \quad (19)$$

where \tilde{i}_t denotes the nominal interest rate in deviations from its steady state, ρ_i captures interest-rate smoothing, ϕ_π and ϕ_x pin down the response coefficients with respect to inflation and the output gap, respectively.

3.3 Subjective expectations under limited attention

Households and firm managers have the same subjective model in mind and face the same attention problem. In particular, households believe that consumption and inflation both follow a random walk:

$$\begin{aligned} \pi_t &= \pi_{t-1} + \nu_{\pi,t} \\ \hat{c}_t &= \hat{c}_{t-1} + \nu_{c,t}, \end{aligned}$$

where $\nu_{\pi,t}$ and $\nu_{c,t}$ are normally distributed with mean zero, time-invariant standard deviations and independent from each other. The random walk assumption ensures that long-run expectations align with the actual long-run variables. They observe all aggregate shocks perfectly.

At the time the household or firm manager forms her expectations about future consumption and inflation, she does not perfectly observe their current realizations. This assumption could capture that the household is not at all times perfectly monitoring all consumption expenditures of the members of the household and that inflation is not perfectly observable in real time. Instead of observing consumption and inflation perfectly, they only receive noisy signals of the form

$$\begin{aligned} s_{\pi,t} &= \pi_t + \varepsilon_{\pi,r,t} \\ s_{c,t} &= \hat{c}_t + \varepsilon_{c,t}, \end{aligned}$$

with normally distributed noise terms $\varepsilon_{\pi,r,t}$ and $\varepsilon_{c,t}$. Following the attention choice problem outlined in Section 2, the noise variance regarding inflation $\varepsilon_{\pi,r,t}$ is smaller in the H regime, reflecting the higher attention in that regime due to the lower cost of information.

As detailed in Appendix C, potential output, i.e., output under flexible prices, is constant and consumption in log-deviations from steady state is equal to the output gap in equilibrium, $\hat{c}_t = \hat{x}_t$. It follows that if we assume initial values $\tilde{E}_{-1}\hat{c}_0 = \tilde{E}_{-1}\hat{x}_0$, we have $\tilde{E}_t\hat{c}_{t+1} = \tilde{E}_t\hat{x}_{t+1}$ for all t . This holds true, even if the household does not know that the two are equal in equilibrium.

Expectations about the output gap, \hat{x}_t , then evolve as follows:

$$\tilde{E}_t\hat{x}_{t+1} = \tilde{E}_{t-1}\hat{x}_t + \gamma_x \left(\hat{x}_t - \tilde{E}_{t-1}\hat{x}_t \right), \quad (20)$$

where $\tilde{E}_t\hat{x}_{t+1}$ denotes the agent's expectations of the one-period-ahead output gap. The parameter γ_x denotes the optimal level of attention to the output gap, based on the agent's subjective model of how the output gap evolves. A higher γ_x denotes a higher attention level. If $\gamma_x = 0$, the agent is completely inattentive and just sticks to her prior belief $\tilde{E}_{t-1}\hat{x}_t$, whereas $\gamma_x = 1$ captures the case of full attention in which case the agent believes $\tilde{E}_t\hat{x}_{t+1} = \hat{x}_t$, which is the full-information belief of someone who believes that the output gap follows a random walk. As I discuss in more detail in the calibration section later on, I do not find any differences in attention to unemployment changes (which I use as a proxy for the output gap). I therefore impose that attention to the output gap, γ_x , does not change across regimes.

Inflation expectations follow the law of motion derived in equation (4) for $\rho_\pi = 1$, so that they are given by

$$\tilde{E}_t\pi_{t+1} = \begin{cases} \tilde{E}_{t-1}\pi_t + \gamma_{\pi,L} \left(\pi_t - \tilde{E}_{t-1}\pi_t \right), & \text{when } \pi_{t-1} < \bar{\pi} \\ \tilde{E}_{t-1}\pi_t + \gamma_{\pi,H} \left(\pi_t - \tilde{E}_{t-1}\pi_t \right), & \text{when } \pi_{t-1} > \bar{\pi}, \end{cases} \quad (21)$$

where $\gamma_{\pi,r}$ captures the optimal level of attention to inflation in regime $r \in \{L, H\}$, depending on the respective cost of information $\frac{1}{\lambda_r}$.

3.4 Equilibrium

The model can be summarized by three equilibrium equations when expressed in log-deviations from the zero-inflation steady state (see Appendix C):

$$\pi_t = \beta \tilde{E}_t\pi_{t+1} + \kappa \hat{x}_t + u_t, \quad (22)$$

$$\hat{x}_t = \tilde{E}_t\hat{x}_{t+1} - \varphi \left(\tilde{i}_t - \tilde{E}_t\pi_{t+1} - r_t^* \right), \quad (23)$$

$$\tilde{i}_t = \rho_i \tilde{i}_{t-1} + (1 - \rho_i) (\phi_\pi \pi_t + \phi_x \hat{x}_t), \quad (24)$$

as well as the two law of motions for output gap expectations and inflation expectations, equations (20) and (21). Equation (22) is the New Keynesian Phillips curve, representing the supply side

of the economy, and equation (23) denotes the aggregate Euler equation. $\varphi > 0$ measures the real rate elasticity of output, and r_t^* is the natural interest rate. The natural interest rate is the real rate that prevails in the economy with fully flexible prices, and solely depends on the exogenous shocks Z_t . It follows an AR(1) process with persistence $\rho_r \in [0, 1]$ and innovations $\varepsilon^r \sim i.i.N.(0, \sigma_r^2)$, independent of ε^u . I will refer to shocks to r_t^* as demand shocks. The nominal interest rate \tilde{i}_t and the natural rate are both expressed in absolute deviations of their respective steady state values, \underline{i} and \underline{r}^* , with $\underline{i} = \underline{r}^*$, as the model is linearized around the zero-inflation steady state. Equation (24) represents the Taylor rule that monetary policy follows when setting the nominal interest rate.

Calibration. Following the empirical findings in Section 2, I set $\gamma_{\pi,L} = 0.18$ and $\gamma_{\pi,H} = 0.36$, and the attention threshold to $\bar{\pi} = 4\%$ (annualized). I calibrate the shock volatilities in order to match the finding that the economy spends 31% of the time in the high-attention regime, and I assume that both shocks have the same volatility and persistence. This results in $\sigma_r = \sigma_u = 0.50\%$ and $\rho_r = 0.8 = \rho_u = 0.8$.

The rest of the calibration is standard. I set the discount factor β to target a steady state natural rate of 1% (annualized), the interest-rate sensitivity φ to 1, and the slope of the Phillips curve κ to 0.057. The Taylor rule coefficients are set to $\rho_i = 0.7$, $\phi_\pi = 2$ and $\phi_x = 0.125$. To calibrate the attention parameter with respect to the output gap γ_x , I follow the same procedure as for inflation but focus on expectations about unemployment changes (see Appendix B.1). This results in $\gamma_x = 0.24$ when inflation is below 4%, and $\gamma_x = 0.25$ when inflation is above it. The two estimates are not statistically significant from each other. I therefore impose that γ_x does not change across regimes and set it to $\gamma_x = 0.25$.

4 Inflation Surges

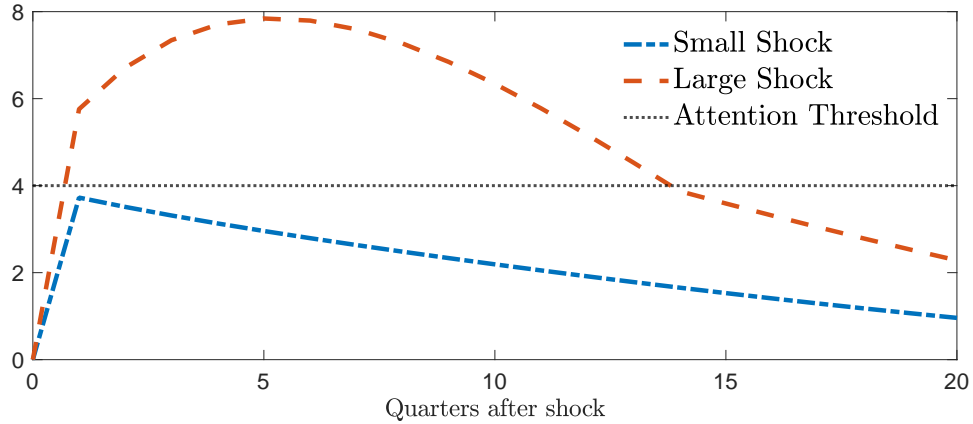
In this section, I show how the model with two inflation-attention regimes can jointly generate persistent heightened inflation periods following large shocks and forecast error dynamics that mirror the empirically-observed patterns during the recent high-inflation period in the United States. Further, I show that cost-push shocks are especially inflationary in times of loose monetary policy and that this state dependency is only present in the economy with the attention threshold. Additionally, I show that even relatively small shocks to demand and supply simultaneously can trigger large inflation surges in the model with two attention regimes, whereas inflation would stay relatively low absent the regime change or when only demand or only supply shocks would hit the economy.

4.1 Large vs. small shocks

Figure 4 shows the inflation dynamics following a cost-push shock that pushes inflation in the first period above the attention threshold $\bar{\pi}$ (the red-dashed line shows the inflation response, and the black-dotted line the attention threshold), as well as following a shock that does not push inflation above the threshold (blue-dashed-dotted line).¹⁰

The two cases exhibit quite different inflation dynamics. In response to the relatively small shock, inflation increases on impact and then gradually returns back to its initial value of zero. After the larger shock, however, inflation keeps on increasing for about five periods before it peaks and then starts to decrease thereafter. While this decrease happens relatively fast at first, it slows down once inflation falls back below the attention threshold, and thus, inflation remains elevated quite persistently. The model under FIRE exhibits similar dynamics than the blue line but inflation returns even more quickly to its initial value (see Figure 11 in Appendix D).

Figure 4: Inflation dynamics



To understand the dynamics in the model with the attention threshold, both regime switches that take place are key. The first regime switch occurs because the shock impulse is large enough to increase inflation above its threshold. Thus, in the second period, agents become more attentive to inflation and hence, increase their inflation expectations more strongly in response to the forecast errors they make. For a given nominal rate, households now perceive the real rate to be lower and thus, increase their consumption in response. The attention regime change also matters for the supply side. Firms increase their inflation expectations more strongly which leads them to increase their prices more strongly. On top of that, the equilibrium inflation response to a cost-push shock for a given state is higher in the high-attention regime:

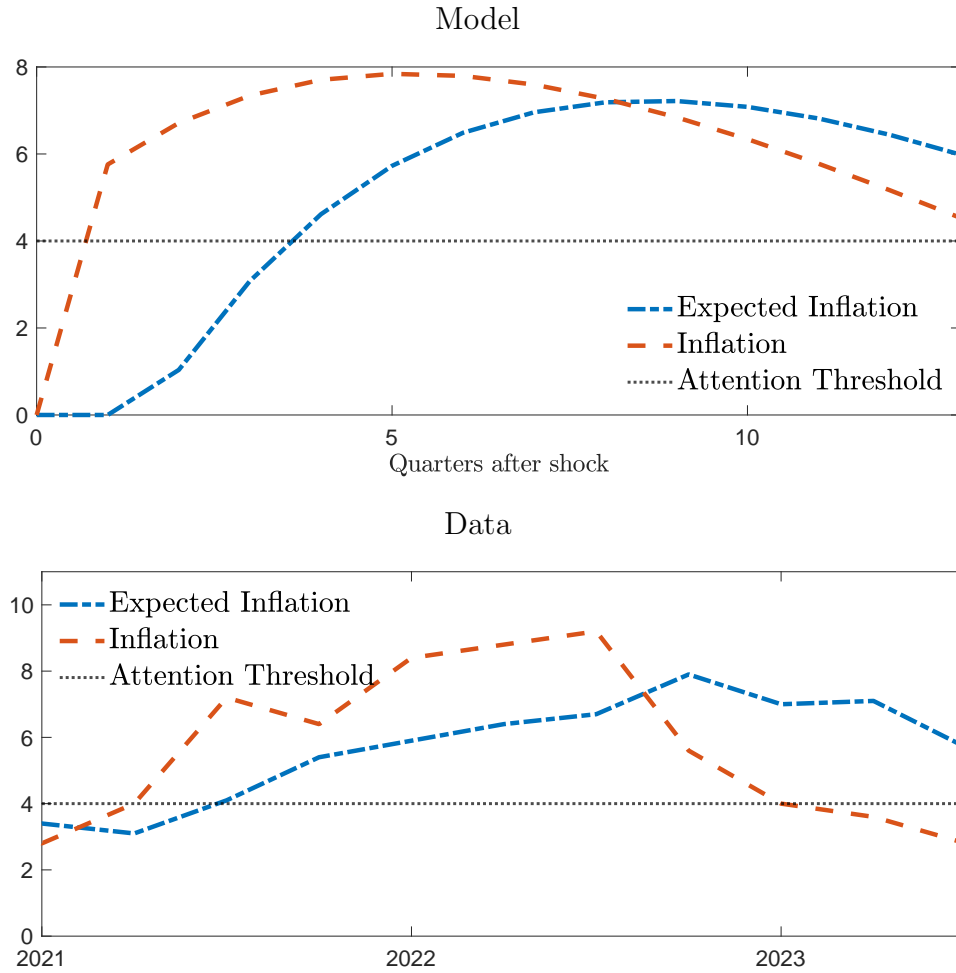
$$\left. \frac{\partial \pi_{t+j}}{\partial u_{t+j}} \right|_{\gamma_{\pi,H}} > \left. \frac{\partial \pi_{t+j}}{\partial u_{t+j}} \right|_{\gamma_{\pi,L}}.$$

¹⁰The dynamics following a demand shock look qualitatively similar and are shown in Figure 10 in Appendix D. Figure 12 shows the inflation response when attention to the output gap decreases when attention to inflation increases. The results are qualitatively similar, but the high-inflation period becomes more persistent.

Hence, inflation keeps on increasing, further fueling higher inflation expectations, leading to additional inflation increases. As the shock slowly dies out, inflation eventually starts to decrease.

The second regime switch—once inflation falls back below the threshold value $\bar{\pi}$ —leads to the slow inflation decrease after the initial inflation surge. When the economy enters the low-attention regime, agents decrease their attention to inflation. Thus, they mainly stick to their prior beliefs and only slowly update their expectations. As their priors are now relatively high due to the high inflation period, inflation expectations remain high persistently which hinders actual inflation from decreasing quickly.

Figure 5: Inflation and inflation expectation dynamics



Notes: The upper panel shows the dynamics of inflation and inflation expectations after a cost-push shock that pushes inflation above the attention threshold. The lower panel plots quarter-on-quarter (annualized) CPI inflation and one-quarter-ahead mean inflation expectations for the period 2021-2023, as well as the estimated attention threshold at 4%. In both panels, the inflation expectations are shifted such that the vertical distance between the two gives the respective forecast errors.

To clearly see these patterns of inflation and inflation expectations, the upper panel in figure 5 plots expected and actual inflation, jointly. Inflation expectations are shifted such that the vertical

distance between the two lines captures the forecast errors. Initially, expected inflation does not quite catch up with actual inflation, leading to positive forecast errors. After some time—around 8 quarters after the shock—when inflation is already decreasing, expected inflation surpasses actual inflation. Hence, forecast errors become negative.¹¹

As the lower panel in figure 5 shows, these patterns are consistent with the recent inflation surge in the US. The figure shows annualized quarter-on-quarter CPI inflation and average inflation expectations from the Michigan Survey for the period from 2021 until 2023Q2. Consistent with the model implications, inflation peaks about a year and a half after its first increase. During this inflation surge, inflation expectations lagged actual inflation. As inflation peaked and began to decline, however, inflation expectations started to surpass actual inflation after about a year and a half after inflation exceeded the attention threshold. Consistent with the model, inflation peaks at a higher value than inflation expectations.

State dependency. What happens when the cost-push shock occurs at a time of loose monetary policy? [Gagliardone and Gertler \(2023\)](#) argue that *easy* monetary policy was a key contributor to the recent inflation surge in the US. To model this, I assume that there is an i.i.d. expansionary monetary policy shock at the same time the cost-push shock occurs. Both shocks are such that they increase inflation on impact by 3 percentage points.¹²

Figure 6 shows the resulting inflation dynamics in response to these shocks. Panel (a) shows the inflation response in the baseline model with the attention threshold, panel (b) shows the case absent the threshold in which attention remains at its low value, panel (c) the model response when attention is permanently at its high value, and panel (d) shows the response in the model under FIRE. In all cases, the black-solid line shows the response to the monetary policy shock in isolation, the orange area shows the response to the cost-push shock, such that the orange-dotted line then shows the sum of the two. The blue-dashed line shows the total response when the two shocks occur simultaneously. Thus, the blue-shaded area captures the interaction effect of the two shocks, or put differently, the state dependency.

We see from panel (a) that this state dependency is substantial in the model with the attention-regime change but is absent in all other models. At peak, inflation increases by five additional percentage points compared to the two shocks in isolation. The reason for this state dependency is the shift in attention that occurs only when both shocks are present.¹³ By entering the high-attention regime, the higher sensitivity of inflation expectations fuel an additional increase in

¹¹[Angeletos et al. \(2020\)](#) show that inflation and unemployment expectations in the data indeed show patterns of initial underreaction (mirrored by positive forecast errors) and delayed overshooting (negative forecast errors). Figures 13 and 14 in Appendix D show that the results look very similar when current rather than lagged inflation is the threshold-defining variable.

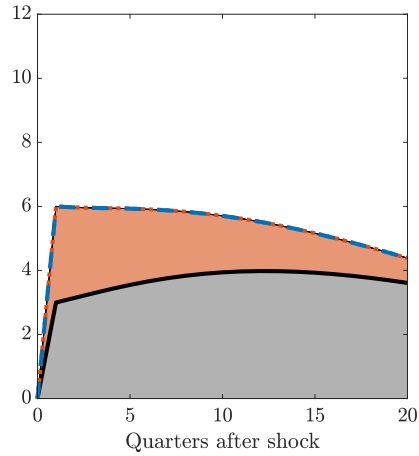
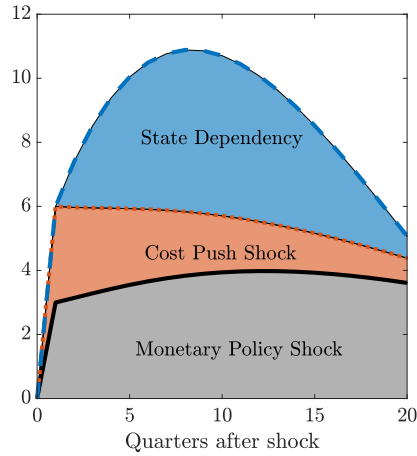
¹²Thus, I re-calibrate the shock sizes when comparing the different models. Only my baseline model with the attention threshold as well as the model with fixed attention at 0.18 have the same shock size.

¹³These state-dependent effects are also present when one shock in isolation would push the economy into the high-attention regime whereas the other shock does not.

Figure 6: Cost-push shocks have larger effects during times of loose monetary policy

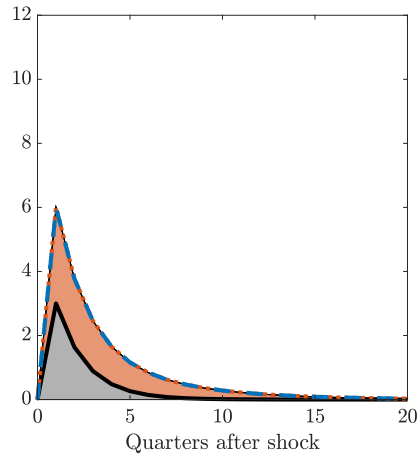
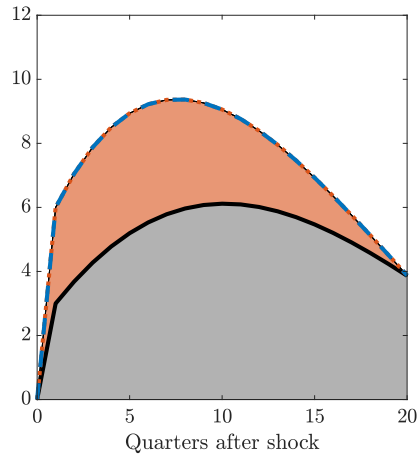
(a) Varying attention: $\gamma_r \in \{0.18, 0.36\}$

(b) Fixed attention: $\gamma = 0.18$



(c) Fixed attention: $\gamma = 0.36$

(d) FIRE



Notes: This figure illustrates the state dependency of cost-push shocks in the economy with the attention threshold. Panel (a) shows the inflation response in the baseline model with the attention threshold, panel (b) shows the case absent the threshold in which attention remains at its low value, panel (c) shows the response in the model with a fixed attention level at the high level of 0.36, and (d) shows the response under FIRE. In all cases, the black-solid line shows the response to a one-time expansionary monetary policy shock in isolation, the orange area shows the response to the cost-push shock, such that the orange-dotted line then shows the sum of the two. The blue-dashed line shows the total response when the two shocks occur simultaneously. The blue-shaded area captures the interaction effect (or the state dependency) of the two shocks.

inflation that is not present when attention remains constant.

These effects are absent in the economies without the attention change or under FIRE, even though the initial inflation surge is the same in all three economies. When attention stays low (panel (b)), the feedback effect through the change in inflation expectations is not strong enough to induce further inflation increases, and hence, there is no state dependency. The same is true for the model under rational expectations. Similarly, the model with a constantly-high level of attention does not generate these interaction effects and the inflation peak is below the one in the model with the attention threshold. Thus, accounting for the attention threshold and the increase in attention offers a potential explanation for why inflation may increase more strongly and faster

when monetary policy is loose at the time of an inflationary shock.

4.2 What drives inflation surges?

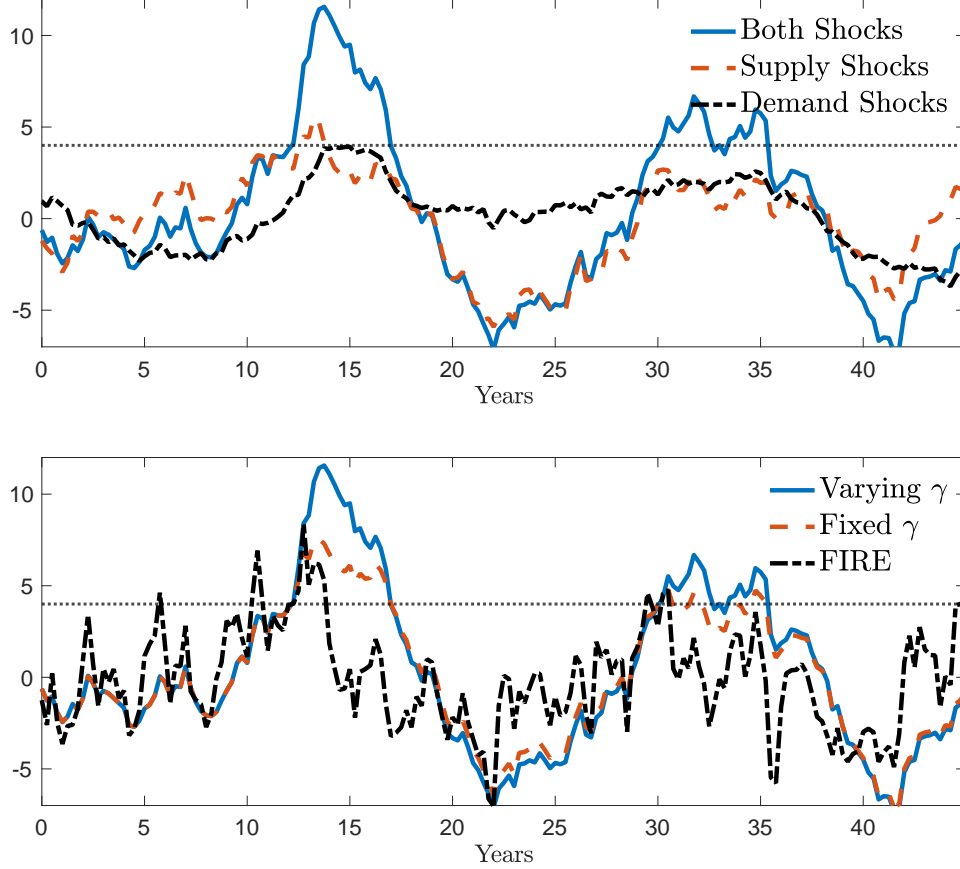
While the debate is still ongoing on what exactly led to the recent inflation surge in the US, I show in this section that high and persistent inflationary episodes may occur even when supply and demand shocks in isolation would only lead to relatively mild inflation rates (see, e.g., [Bernanke and Blanchard \(2023\)](#); [Benigno and Eggertsson \(2023\)](#); [Bianchi and Melosi \(2022\)](#); [Di Giovanni et al. \(2022\)](#); [Pflueger \(2023\)](#); [Gagliardone and Gertler \(2023\)](#) for possible different explanations of the recent inflation dynamics). For this, I simulate the economy for 10,000 periods by randomly drawing demand shocks $\{r_t^*\}$ and supply shocks $\{u_t\}$ and feeding them into the model economy. Figure 7 shows a forty-five year outtake of these simulations. The upper panel shows the inflation dynamics for the case in which the economy is hit by both shocks (blue-solid line), and the cases where the economy is hit only by supply shocks (red-dashed line) or only by demand shocks (black-dashed-dotted line). The lower panel compares the economy with the two attention regimes (blue-solid line) with the model featuring only one attention regime (red-dashed line) and the model under full-information rational expectations (black-dashed-dotted line).

In this specific outtake of the simulation, inflation enters the high-attention twice. From the upper panel of Figure 7, we see that these high-inflation episodes are driven by a combination of supply and demand shocks. If only one shock would be hitting the economy in this specific scenario, inflation would exceed the attention threshold only once and only for a relatively short time. However, as illustrated in Figure 4, high-inflation episodes may also arise due to a single shock if the shock is large enough.

The lower panel illustrates how the regime changes exacerbate the inflationary dynamics in two ways. First, by making these high-inflation episodes more long lasting, especially when compared to the economy under full-information rational expectations. This occurs because the increase in attention triggers self-reinforcing inflation and inflation expectation dynamics, thus, keeping inflation higher for longer, as discussed in section 4.1. Second, by amplifying the increase in inflation. The inflation surges are almost twice as high compared to the economy without the change in attention or as in the model under FIRE.

But while the inflation attention threshold exacerbates high inflationary periods, the periods of low inflation (or even deflation) are actually quite similar across models. Thus, the attention threshold introduces an asymmetry: attention increases when inflation is particularly high but not when it is particularly low. This asymmetry leads to a higher *average* inflation rate. So, the attention threshold not only amplifies inflation fluctuations, but also increases its level. Both of these effects lead to welfare losses, as I discuss next.

Figure 7: Inflation surges



Notes: The upper panel shows the inflation dynamics for the case in which the economy is hit by both shocks (blue-solid line), and the cases where the economy is hit only by supply shocks (red-dashed line) or only by demand shocks (black-dashed-dotted line). The lower panel compares the economy with the two attention regimes (blue-solid line) with the model featuring only one attention regime (red-dashed line) and the model under full-information rational expectations (black-dashed-dotted line).

5 Welfare Implications

I now characterize the normative implications of the inflation attention threshold and the corresponding changes in attention to inflation. To compare different policy rules, I define welfare as

$$-\frac{1}{2}E_0 \sum_{t=0}^{\infty} \beta^t [\pi_t^2 + \Lambda \hat{x}_t^2], \quad (25)$$

where Λ is the relative weight of the output gap, which I set to $\Lambda = 0.007$ as in [Adam and Billi \(2006\)](#). Welfare (25) is obtained by deriving a second-order approximation to the household's utility function (see, e.g., [Woodford \(2003\)](#)). An implicit assumption here is that the policymaker is paternalistic ([Benigno and Paciello, 2014](#)) in the sense that the policymaker evaluates the household's utility under rational expectations. I further assume that the policy maker is not aware of the private agents' limited attention and acts as if they hold rational expectations.

When implementing the optimal policy, the policymaker maximizes welfare (25) subject to the

Table 3: Monetary policy rules

Nr.	Name	Equation
(1)	Taylor rule with smoothing	$\tilde{i}_t = \rho_i \tilde{i}_{t-1} + (1 - \rho_i) (\phi_\pi \pi_t + \phi_x \hat{x}_t)$
(2)	Taylor rule without smoothing	$\tilde{i}_t = \phi_\pi \pi_t + \phi_x \hat{x}_t$
(3)	Optimal commitment policy	$\pi_t + \frac{\Lambda}{\kappa} (\hat{x}_t - \hat{x}_{t-1}) = 0$
(4)	Optimal discretionary policy	$\pi_t + \frac{\Lambda}{\kappa} \hat{x}_t = 0$
(5)	Strict inflation targeting	$\pi_t = 0$

private sector's optimality conditions, given by equations (22)-(23), but as if the private sector's expectations were rational. As Clarida et al. (1999) show, the optimal policy under commitment is given by

$$\pi_t + \frac{\Lambda}{\kappa} (\hat{x}_t - \hat{x}_{t-1}) = 0$$

when agents hold rational expectations. Under discretion, the optimal policy does not take its effects on inflation expectations into account and is then given by

$$\pi_t + \frac{\Lambda}{\kappa} \hat{x}_t = 0.$$

In the following, I compare the welfare implications of these different policy rules and compare them to the Taylor rule (24) as well as to a Taylor rule without interest rate smoothing, i.e., to (24) where $\rho_i = 0$, as well as to a strict-inflation targeting regime in which inflation is kept at zero at all times. Table 3 summarizes these different policy rules.

I then simulate the economy for 10,000 periods and for the different policy rules. Figure 8 plots welfare (25) for the model with the attention threshold (blue-solid line), with limited attention but absent the attention threshold (red-dashed line) and the model under FIRE (black-dashed-dotted line) for the 5 different policy rules. Panel (a) in Figure 9 shows the respective inflation volatilities, panel (b) the average level of inflation, and panel (c) the frequency of how often inflation is above the threshold of 4%.

Figure 8: Welfare

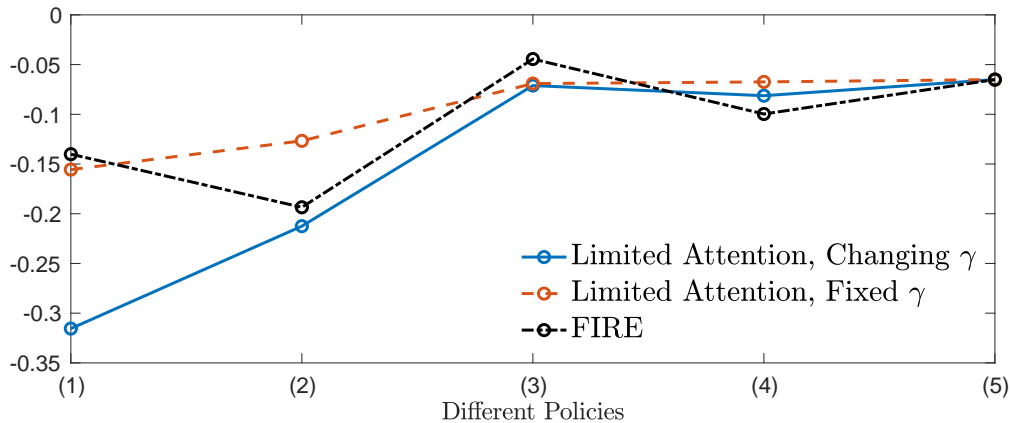
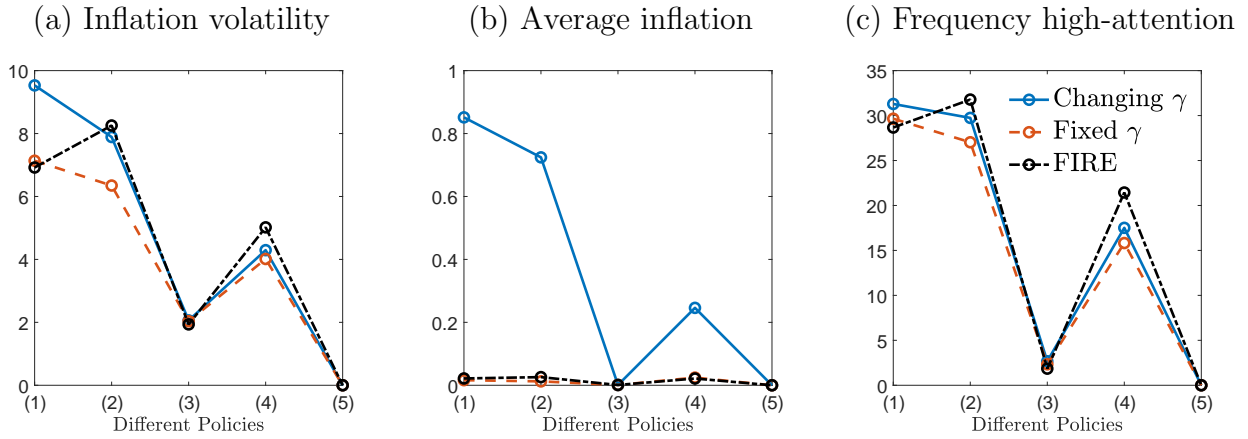


Figure 9: Inflation volatility, average inflation and time spent in the high-attention regime



There are two main takeaways from figures 8 and 9. First, following simple Taylor rules (policy rules (1) and (2)) leads to substantially larger welfare losses compared to optimal policy rules or a strict inflation targeting rule in the model with the attention threshold, especially for the case with interest-rate smoothing.¹⁴ The reason is that the economy spends a substantial amount of time in the high-attention regime in which inflation is high and volatile (see panel (c) in figure 9). Due to the asymmetry of the attention threshold the average level of inflation is higher when the economy is in the high-attention regime frequently (panel (b)). This is in stark contrast to the model without the threshold or the one with fully-informed rational agents, where inflation fluctuates symmetrically around zero. Interest rate smoothing introduces additional persistence, such that the periods in the high-attention regime last longer and thus, the average level of inflation as well as its volatility increase.

The second main take away is that neither limited attention nor the attention threshold have large welfare implications, when monetary policy follows one of the rules (3)-(5). In all cases, inflation is relatively stable and fluctuates almost symmetrically around 0, as the economy very rarely stays in the high-attention regime (an exception is the optimal discretion rule (2)). In the case of the strict-inflation targeting regime (5) the inflation volatility is exactly 0. However, in that case, the output gap is more volatile (not shown) such that the overall welfare losses are very similar to the ones under the other policy rules from (3) and (4).

Overall, figures 8 and 9 illustrate that following simple Taylor rules, especially ones with interest-rate smoothing and relatively low inflation-response coefficients, can lead to large welfare losses when the public's attention to inflation may increase substantially during times of high inflation. Policies that are more hawkish and induce much smaller inflation fluctuations, in contrast, can mitigate the potentially detrimental effects of sudden increases in attention much more effectively, or even prevent these episodes from happening completely.

¹⁴Gáti (2022) shows that in her model with varying degrees of long-run-expectations anchoring, the welfare losses of Taylor rules with a low inflation-response coefficient may go to infinity, as in this case the model becomes unstable and inflation (and the output gap and interest rates) becomes explosive.

6 Conclusion

The recent inflation surge in many advanced economies brought inflation back on people’s minds. In this paper, I quantify the inflation attention threshold after which people start to pay more attention to inflation. I estimate this attention threshold to be at an inflation rate of 4% in the United States and that attention doubles from the low-attention regime to the high-attention regime. I estimate these statistics in a way that allows me to directly map the empirical estimates into an otherwise standard New Keynesian model.

Accounting for the inflation attention threshold substantially alters inflation and inflation expectation dynamics in response to shocks. When calibrating the model to match the empirical facts, the model generates inflation and inflation expectation dynamics that are consistent with the ones recently observed in the US.

Accounting for varying attention levels also matters for the models’ normative predictions. Following simple Taylor rules leads to much larger welfare losses compared to optimal policy rules. Such simple policies induce frequent and long-lasting episodes in which attention is heightened which leads to large welfare losses, not only because these episodes increase the overall volatility of inflation, but—due to the asymmetry of the attention threshold—also the average inflation rates.

References

- ADAM, K. AND R. M. BILLI (2006): “Optimal monetary policy under commitment with a zero bound on nominal interest rates,” *Journal of Money, Credit and Banking*, 1877–1905.
- AFROUZI, H. AND C. YANG (2020): “Dynamic inattention, the phillips curve, and forward guidance,” Tech. rep., Working Paper.
- ANDRE, P., C. PIZZINELLI, C. ROTH, AND J. WOHLFART (2022): “Subjective models of the macroeconomy: Evidence from experts and representative samples,” *The Review of Economic Studies*, 89, 2958–2991.
- ANGELETOS, G.-M., Z. HUO, AND K. SASTRY (2020): “Imperfect expectations: Theory and evidence,” in *NBER Macroeconomics Annual 2020, volume 35*, University of Chicago Press.
- BAYER, C., B. BORN, AND R. LUETTICKE (2022): “Shocks, frictions, and inequality in US business cycles,” .
- BENATI, L. (2008): “Investigating inflation persistence across monetary regimes,” *The Quarterly Journal of Economics*, 123, 1005–1060.
- BENIGNO, P. AND G. B. EGGERTSSON (2023): “It’s baaack: The surge in inflation in the 2020s and the return of the non-linear phillips curve,” *Working paper*.

- BENIGNO, P. AND L. PACIELLO (2014): “Monetary policy, doubts and asset prices,” *Journal of Monetary Economics*, 64, 85–98.
- BERNANKE, B. AND O. BLANCHARD (2023): “What caused the US pandemic-era inflation?” *Hutchins Center Working Papers*.
- BHANDARI, A., J. BOROVÍČKA, AND P. HO (2019): “Survey data and subjective beliefs in business cycle models,” .
- BIANCHI, F. AND L. MELOSI (2022): “Inflation as a fiscal limit,” .
- BLANCO, A., P. OTTONELLO, AND T. RANOSOVA (2022): “The Dynamics of Large Inflation Surges,” *Working paper*.
- BRACHA, A. AND J. TANG (2023): “Inflation levels and (in) attention,” *Working Paper*.
- BROER, T., A. KOHLHAS, K. MITMAN, AND K. SCHLAFMANN (2021): “Information and wealth heterogeneity in the macroeconomy,” *Working paper*.
- CANDIA, B., O. COIBION, AND Y. GORODNICHENKO (2021): “The Inflation Expectations of US Firms: Evidence from a New Survey,” *Working paper*.
- CANDIA, B., M. WEBER, Y. GORODNICHENKO, AND O. COIBION (2023): “Perceived and Expected Rates of Inflation of US Firms,” 113, 52–55.
- CARLSON, J. A. AND M. PARKIN (1975): “Inflation expectations,” *Economica*, 42, 123–138.
- CARVALHO, C., S. EUSEPI, E. MOENCH, AND B. PRESTON (2022): “Anchored inflation expectations,” *American Economic Journal: Macroeconomics* (forthcoming).
- CAVALLO, A., G. CRUCES, AND R. PEREZ-TRUGLIA (2017): “Inflation expectations, learning, and supermarket prices: Evidence from survey experiments,” *American Economic Journal: Macroeconomics*, 9, 1–35.
- CLARIDA, R., J. GALI, AND M. GERTLER (1999): “The science of monetary policy: a new Keynesian perspective,” *Journal of economic literature*, 37, 1661–1707.
- COIBION, O. AND Y. GORODNICHENKO (2015): “Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation,” *American Economic Journal: Macroeconomics*, 7, 197–232.
- COIBION, O., Y. GORODNICHENKO, E. S. I. KNOTEK, AND R. SCHOENLE (2020): “Average Inflation Targeting and Household Expectations,” *Journal of Political Economy* (forthcoming).

- D’ACUNTO, F., U. MALMENDIER, AND M. WEBER (2023): “What do the data tell us about inflation expectations?” in *Handbook of economic expectations*, Elsevier, 133–161.
- DI GIOVANNI, J., S. KALEMLI OZCAN, A. SILVA, AND M. YILDIRIM (2022): “Global supply chain pressures, international trade, and inflation,” *Working paper*.
- D’ACUNTO, F., D. HOANG, M. PALOVHITA, AND M. WEBER (2019): “Cognitive abilities and inflation expectations,” in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 109, 562–566.
- D’ACUNTO, F., U. MALMENDIER, AND M. WEBER (2021): “Gender roles produce divergent economic expectations,” *Proceedings of the National Academy of Sciences*, 118, e2008534118.
- EVANS, G. W. AND G. RAMEY (1995): “Expectation calculation, hyperinflation and currency collapse,” *The new macroeconomics: Imperfect markets and policy effectiveness*, 307.
- FLYNN, J. P. AND K. SASTRY (2022): “Attention cycles,” *Working Paper*.
- GAGLIARDONE, L. AND M. GERTLER (2023): “Oil Prices, Monetary Policy and Inflation Surges,” *Working paper*.
- GALÍ, J. (2015): *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press.
- GALLEGOS, J. E. (2023): “Inflation persistence, noisy information and the Phillips curve,” *Working paper*.
- GÁTI, L. (2022): “Monetary Policy & Anchored Expectations An Endogenous Gain Learning Model,” *Journal of Monetary Economics* (forthcoming).
- GOLDSTEIN, N. (2023): “Tracking inattention,” *Journal of the European Economic Association*, jvad022.
- GONZALO, J. AND J.-Y. PITARAKIS (2002): “Estimation and model selection based inference in single and multiple threshold models,” *Journal of econometrics*, 110, 319–352.
- HAGEDORN, M., J. LUO, I. MANOVSKII, AND K. MITMAN (2019): “Forward guidance,” *Journal of Monetary Economics*, 102, 1–23.
- HANSEN, B. E. (2011): “Threshold autoregression in economics,” *Statistics and its Interface*, 4, 123–127.
- HAZELL, J., J. HERRENO, E. NAKAMURA, AND J. STEINSSON (2022): “The slope of the Phillips Curve: evidence from US states,” *The Quarterly Journal of Economics*, 137, 1299–1344.

- HILSCHER, J., A. RAVIV, AND R. REIS (2022): “How likely is an inflation disaster?” .
- JØRGENSEN, P. L. AND K. J. LANSING (2023): “Anchored inflation expectations and the slope of the phillips curve,” *Working Paper*.
- KIM, G. AND C. BINDER (2023): “Learning-through-survey in inflation expectations,” *American Economic Journal: Macroeconomics*, 15, 254–278.
- KORENOK, O., D. MUNRO, AND J. CHEN (2022): “Inflation and attention thresholds,” *Working paper*.
- KRONER, N. (2023): “Inflation and Attention: Evidence from the Market Reaction to Macro Announcements,” *Available at SSRN 4527424*.
- LAMLA, M. J. AND S. M. LEIN (2014): “The role of media for consumers’ inflation expectation formation,” *Journal of Economic Behavior & Organization*, 106, 62–77.
- LARSEN, V. H., L. A. THORSRUD, AND J. ZHULANOVA (2021): “News-driven inflation expectations and information rigidities,” *Journal of Monetary Economics*, 117, 507–520.
- MACAULAY, A. (2022): “Shock transmission and the sources of heterogeneous expectations,” *Working Paper*.
- MACKOWIAK, B., F. MATEJKA, AND M. WIEDERHOLT (2023): “Rational Inattention: A Review,” *Journal of Economic Literature*.
- MACKOWIAK, B. AND M. WIEDERHOLT (2009): “Optimal sticky prices under rational inattention,” *American Economic Review*, 99, 769–803.
- MATĚJKA, F. AND A. MCKAY (2015): “Rational inattention to discrete choices: A new foundation for the multinomial logit model,” *American Economic Review*, 105, 272–98.
- MCCLURE, E. M., O. COIBION, AND Y. GORODNICHENKO (2022): “The Macroeconomic Expectations of US Managers,” *Working Paper*.
- MEICHTRY, P. (2022): “Sticky Information, Heterogeneity, and Aggregate Demand,” .
- NEWKEY, W. K. AND K. D. WEST (1987): “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation,” *Econometrica*, 55, 703–708.
- NIMARK, K. P. AND S. PITSCHNER (2019): “News media and delegated information choice,” *Journal of Economic Theory*, 181, 160–196.
- NORD, L. (2022): “Who Cares about Inflation? Endogenous Expectation Formation of Heterogeneous Households,” *Working paper*.

- PACIELLO, L. AND M. WIEDERHOLT (2014): “Exogenous information, endogenous information, and optimal monetary policy,” *Review of Economic Studies*, 81, 356–388.
- PEDEMONTE, M., H. TOMA, AND E. VERDUGO (2023): “Aggregate Implications of Heterogeneous Inflation Expectations: The Role of Individual Experience,” .
- PFÄUTI, O. (2023): “Inflation—who cares? Monetary Policy in Times of Low Attention,” *Working paper*.
- PFÄUTI, O. AND F. SEYRICH (2023): “A behavioral heterogeneous agent New Keynesian model,” *Working paper*.
- PFLUEGER, C. (2023): “Back to the 1980s or not? The drivers of inflation and real risks in Treasury bonds,” Tech. rep., National Bureau of Economic Research.
- REIS, R. (2022): “Losing the inflation anchor,” *Brookings Papers on Economic Activity*, 2021, 307–379.
- ROTEMBERG, J. J. (1982): “Sticky prices in the United States,” *Journal of political economy*, 90, 1187–1211.
- ROTH, C., M. WIEDERHOLT, AND J. WOHLFART (2023): “The Effects of Monetary Policy: Theory with Measured Expectations,” *Working paper*.
- SCHMIDT, T., H. MÜLLER, J. RIEGER, T. SCHMIDT, AND C. JENTSCH (2023): “Inflation Perception and the Formation of Inflation Expectations,” *Working paper*.
- SIMS, C. A. (2003): “Implications of rational inattention,” *Journal of monetary Economics*, 50, 665–690.
- (2010): “Rational inattention and monetary economics,” in *Handbook of monetary economics*, Elsevier, vol. 3, 155–181.
- VELLEKOOP, N. AND M. WIEDERHOLT (2019): “Inflation expectations and choices of households,” .
- WEBER, M., F. D’ACUNTO, Y. GORODNICHENKO, AND O. COIBION (2022): “The subjective inflation expectations of households and firms: Measurement, determinants, and implications,” *Journal of Economic Perspectives*, 36, 157–184.
- WEBER, M., S. SHEFLIN, T. ROPELE, R. LLUBERAS, S. FRACHE, B. MEYER, S. KUMAR, Y. GORODNICHENKO, D. GEORGARAKOS, O. COIBION, ET AL. (2023): “Tell Me Something I Don’t Already Know: Learning in Low and High-Inflation Settings,” .
- WOODFORD, M. (2003): *Interest and Prices*, Princeton University Press.

A A Limited-Attention Model of Inflation Expectations

In this section, I derive the expectations-formation process under limited attention sketched in Section 2. The agent believes that (demeaned) inflation tomorrow, π' , depends on (demeaned) inflation today, π , as follows

$$\pi' = \rho_\pi \pi + \nu,$$

where $\rho_\pi \in [0, 1]$ denotes the perceived persistence of inflation and $\nu \sim i.i.N.(0, \sigma_\nu^2)$. Inflation in the current period is unobservable, so before forming an expectation about future inflation, the agent needs to form an expectation about today's inflation. I denote this nowcast $\tilde{\pi}$, and the resulting forecast about next period's inflation $\pi^e = \rho_\pi \tilde{\pi}$. Given her beliefs, the full-information forecast π^{e*} is

$$\pi^{e*} \equiv \rho_\pi \pi.$$

But since π is not perfectly observable, the actual forecast will deviate from the full-information forecast. Deviating, however, is costly, as this causes the agent to make mistakes in her decisions.

The agent's choice is not only about how to form her expectations given certain information, but about how to choose this information optimally, while taking into account how this will later affect her forecast. That is, she chooses the form of the signal s she receives about current inflation. Since acquiring information is costly, it cannot be optimal to acquire different signals that lead to an identical forecast. Due to this one-to-one relation of signal and forecast, we can directly work with the joint distribution of π^e and π , $f(\pi^e, \pi)$, instead of working with the signal.

Let $U(\pi^e, \pi)$ denote the negative of the loss that is incurred when the agent's forecast deviates from the forecast under full information, and $C(f)$ the cost of information. Then, the agent's problem is given by

$$\begin{aligned} & \max_f \int U(\pi^e, \pi) f(\pi^e, \pi) d\pi d\pi^e - C(f) \\ & \text{subject to } \int f(\pi^e, \pi) d\pi^e = g(\pi), \text{ for all } \pi, \end{aligned} \tag{26}$$

where $g(\pi)$ is the agent's prior, which is assumed to be Gaussian; $\pi \sim N(\hat{\pi}, \sigma_\pi^2)$. $C(\cdot)$ is the cost function that captures how costly information acquisition is. It is linear in *mutual information* $I(\pi; \pi^e)$, i.e., the expected reduction in entropy of π due to knowledge of π^e :

$$C(f) = \frac{1}{\lambda} I(\pi; \pi^e) = \frac{1}{\lambda} (H(\pi) - E[H(\pi|\pi^e)]),$$

where $H(x) = -\int f(x) \log(f(x)) dx$ is the entropy of x and $\frac{1}{\lambda}$ measures the cost of information.

The objective function $U(\cdot)$ is assumed to be quadratic:

$$U(\pi^e, \pi) = -\chi (\rho_\pi \pi - \pi^e)^2,$$

where χ measures the stakes of making a mistake.¹⁵

In this setup, Gaussian signals are optimal (and in fact the unique solution, see [Matějka and McKay \(2015\)](#)). The optimal signal thus has the form

$$s = \pi + \varepsilon,$$

with $\varepsilon \sim i.i.N.(0, \sigma_\varepsilon^2)$.¹⁶ The problem (26) now reads

$$\max_{\sigma_{\pi|s}^2 \leq \sigma_\pi^2} E_\pi \left[E_s \left[-\chi \rho_\pi^2 (\pi - E[\pi|s])^2 \right] \right] - \frac{1}{\lambda} I(\pi; \pi^e) = \max_{\sigma_{\pi|s}^2 \leq \sigma_\pi^2} \left(-\chi \rho_\pi^2 \sigma_{\pi|s}^2 - \frac{1}{2} \log \frac{\sigma_\pi^2}{\sigma_{\pi|s}^2} \right). \quad (27)$$

The optimal forecast is given by $\pi^e = \rho_\pi E[\pi|s]$, and Bayesian updating implies

$$\pi^e = \rho_\pi (1 - \gamma) \hat{\pi} + \rho_\pi \gamma s, \quad (28)$$

where $\gamma = 1 - \frac{\sigma_{\pi|s}^2}{\sigma_\pi^2} \in [0, 1]$ measures how much attention the agent pays to inflation, and $\hat{\pi}$ denotes the prior mean of π .

An equivalent way of writing γ is

$$\gamma = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\varepsilon^2}. \quad (29)$$

Now, since the agent *chooses* the level of attention, we can re-formulate (27) as

$$\max_{\gamma \in [0, 1]} \left(-\chi \rho_\pi^2 (1 - \gamma) \sigma_\pi^2 - \frac{1}{2} \log \frac{1}{1 - \gamma} \right). \quad (30)$$

Solving this optimization problem (30) yields the *optimal* level of attention, presented in Section 2.

B Appendix to Empirical Section 2

B.1 Unemployment expectations

In order to calibrate the attention parameter with respect to the output gap, γ_x , I use expectations about unemployment changes from the Survey of Consumers. The survey asks households whether they expect unemployment to increase, decrease or to remain about the same over the next twelve months. I follow [Carlson and Parkin \(1975\)](#), [Bhandari et al. \(2019\)](#) and [Pfäuti and Seyrich \(2023\)](#) to translate these categorical unemployment expectations into numerical expectations.

¹⁵A quadratic loss function is usually derived from a second-order approximation of the household's utility function or the firm's profit function (see, e.g., [Mackowiak and Wiederholt \(2009\)](#)).

¹⁶In this case, the entropy becomes $H(x) = \frac{1}{2} \log(2\pi e \sigma_x^2)$, where σ_x^2 is the variance of x . Note, that here π denotes the number “pi” and not inflation.

Let q_t^D , q_t^S and q_t^U denote the shares reported at time t that think unemployment will go down, stay roughly the same, or go up over the next year, respectively. I assume that these shares are drawn from a cross-sectional distribution of responses that are normally distributed according to $\mathcal{N}(\mu_t, (\sigma_t)^2)$ and a threshold a such that when a household expects unemployment to remain within the range $[-a, a]$ over the next year, she responds that unemployment will remain "about the same". We thus have

$$q_t^D = \Phi\left(\frac{-a - \mu_t}{\sigma_t}\right) \quad q_t^U = 1 - \Phi\left(\frac{a - \mu_t}{\sigma_t}\right),$$

which after some rearranging yields

$$\begin{aligned} \sigma_t &= \frac{2a}{\Phi^{-1}(1 - q_t^U) - \Phi^{-1}(q_t^D)} \\ \mu_t &= a - \sigma_t \Phi^{-1}(1 - q_t^U). \end{aligned}$$

This leaves us with one degree of freedom, namely a . I follow [Pfäuti and Seyrich \(2023\)](#) and set $a = 0.5$ which means that if a household expects the change in unemployment to be less than half a percentage point (in absolute terms), she reports that she expects unemployment to be about the same as it is at the time of the survey. I use data from FRED for the actual unemployment changes and restrict the sample to end in 2019Q4, due to the extreme behavior of unemployment changes with the outbreak of the Covid-19 pandemic.

I then estimate γ_x separately for whether lagged inflation is above or below the estimated threshold of 4%. This results in estimates $\gamma_{x,L} = 0.25$ and $\gamma_{x,H} = 0.25$. Thus, the two are nearly identical and hence, I abstract from changes in γ_x across regimes in the model.

C Model Derivations

Derivation of equation (23). Linearizing the Euler equation (11) yields

$$\hat{c}_t = \tilde{E}_t \hat{c}_{t+1} - \varphi \left(\tilde{i}_t - \tilde{E}_t \pi_{t+1} - (\hat{z}_t - \tilde{E}_t \hat{z}_{t+1}) \right). \quad (31)$$

The representative household therefore needs to form expectations about future consumption $\tilde{E}_t \hat{c}_{t+1}$, inflation $\tilde{E}_t \pi_{t+1}$ and the exogenous preference shock $\tilde{E}_t \hat{z}_{t+1}$. The household holds rational expectations about the shock, so $\tilde{E}_t \hat{z}_{t+1} = E_t \hat{z}_{t+1}$, where $E_t(\cdot)$ denotes the rational expectations operator, and since \hat{z}_t follows an AR(1) process with persistence ρ_z , it follows that $\tilde{E}_t \hat{z}_{t+1} = E_t \hat{z}_{t+1} = \rho_z \hat{z}_t$.

Following the discussion in Section 2, the household holds a subjective model of how

consumption and inflation evolve. She believes that they both follow a random walk:

$$\begin{aligned}\pi_t &= \pi_{t-1} + \nu_{\pi,t} \\ \hat{c}_t &= \hat{c}_{t-1} + \nu_{c,t},\end{aligned}$$

where $\nu_{\pi,t}$ and $\nu_{c,t}$ are normally distributed with mean zero, time-invariant standard deviations and independent from each other. At the time the household needs to form expectations about future consumption and inflation, she does not perfectly observe their current realizations. This assumption could capture that the household is not at all times perfectly monitoring all consumption expenditures of the members of the household. Instead of observing consumption (and inflation) perfectly, the household only receives noisy signals of the form

$$\begin{aligned}s_{\pi,t} &= \pi_t + \varepsilon_{\pi,t} \\ s_{c,t} &= \hat{c}_t + \varepsilon_{c,t},\end{aligned}$$

with normally distributed noise terms $\varepsilon_{\pi,t}$ and $\varepsilon_{c,t}$. As explained in Sections 2 and A, the noise of these signals is endogenous and depends on how attentive the household is. Given these signals, it follows from the (steady state) Kalman filter that optimal updating in regime r is given by

$$\begin{aligned}\tilde{E}_t \pi_{t+1} &= \tilde{E}_{t-1} \pi_t + \gamma_{\pi,r} \left(\pi_t - \tilde{E}_{t-1} \pi_t \right) + u_{\pi,t} \\ \tilde{E}_t \hat{c}_{t+1} &= \tilde{E}_{t-1} \hat{c}_t + \gamma_{c,r} \left(\hat{c}_t - \tilde{E}_{t-1} \hat{c}_t \right) + u_{c,t}.\end{aligned}$$

As I focus on a representative household, I assume that the noise terms $u_{\pi,t}$ and $u_{c,t}$ wash out and are equal to zero.

In order to express equation (31) in terms of the output gap, we need to derive the efficient allocation, i.e., the one that prevails in the economy with fully flexible prices. From the production function, we have $Y_t = H_t$. The real wage is constant $w_t = 1$. From the labor-leisure equation (12), we get that potential output is therefore also constant and equal to

$$Y_t^* = \Xi^{-\frac{1}{\nu+\sigma}}. \quad (32)$$

Thus, potential output in log-deviations is 0. The Euler equation in the flexible-price economy is therefore given by

$$0 = -\varphi \left(r_t - (\hat{z}_t - \tilde{E}_t z_{t+1}) \right).$$

Since the natural rate is defined as the real rate that prevails under flexible prices, r_t , it follows directly that

$$r_t^* = \hat{z}_t - \tilde{E}_t z_{t+1}. \quad (33)$$

Substituting $\hat{z}_t - \tilde{E}_t z_{t+1}$ with r_t^* in (31) and using that $\hat{c}_t = \hat{y}_t = \hat{x}_t$, since potential output is 0,

yields the IS equation (23):

$$\hat{x}_t = \tilde{E}_t \hat{c}_{t+1} - \varphi \left(\tilde{i}_t - \tilde{E}_t \pi_{t+1} - r_t^* \right). \quad (34)$$

Note, however, that if we assume initial values $\tilde{E}_{-1} \hat{c}_0 = \tilde{E}_{-1} \hat{x}_0$, it follows that $\tilde{E}_t \hat{c}_{t+1} = \tilde{E}_t \hat{x}_{t+1}$ because $\hat{c}_t = \hat{x}_t$ in equilibrium. This holds true, even if the household does not know that they are equal in equilibrium. We can thus write equation (34) as

$$\hat{x}_t = \tilde{E}_t \hat{x}_{t+1} - \varphi \left(\tilde{i}_t - \tilde{E}_t \pi_{t+1} - r_t^* \right), \quad (35)$$

which is equation (23) as stated in the main text.

Derivation of equation (22). To derive the New Keynesian Phillips Curve (22), we start by maximizing

$$Profits_0(j) = \tilde{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(1 - \tau_t) P_t(j) \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon} \frac{Y_t}{P_t} - w_t H_t(j) - \frac{\psi}{2} \left(\frac{P_t(j)}{P_{t-1}(j)} - 1 \right)^2 Y_t + t_t^F(j) \right],$$

which yields the first-order condition

$$\begin{aligned} (1 - \tau_t)(\epsilon - 1) P_t(j)^{-\epsilon} \frac{Y_t}{P_t^{1-\epsilon}} &= \epsilon m c_t \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon-1} \frac{Y_t}{P_t} - \psi \left(\frac{P_t(j)}{P_{t-1}(j)} - 1 \right) \frac{Y_t}{P_{t-1}(j)} \\ &+ \beta \psi \tilde{E}_t \left[\left(\frac{P_{t+1}(j)}{P_t(j)} - 1 \right) \frac{P_{t+1}(j)}{P_t(j)} \frac{Y_{t+1}}{P_t(j)} \right]. \end{aligned}$$

I assume that all firm managers receive the same signal about inflation and since in equilibrium all firms are identical, they all then set the same price and produce the same output. Hence, the problem simplifies to

$$(\epsilon - 1)(1 - \tau_t) = \epsilon m c_t - \psi \pi_t (1 + \pi_t) + \psi \beta \tilde{E}_t (1 + \pi_{t+1}) \pi_{t+1} \frac{Y_{t+1}}{Y_t}.$$

Log-linearizing around the zero-inflation steady state yields

$$\psi \pi_t = \epsilon \underline{m c} \hat{m c}_t + (\epsilon - 1) \underline{\tau} \hat{\tau}_t + \psi \beta \tilde{E}_t \pi_{t+1}. \quad (36)$$

The steady-state marginal cost is given by $\underline{m c} = \frac{\epsilon-1}{\epsilon}$, and the marginal cost in log deviations from the steady state by $\hat{m c}_t = (\sigma + \nu) \hat{y}_t$. We can therefore rewrite the equation as

$$\pi_t = \underbrace{\frac{\epsilon - 1}{\psi} (\sigma + \nu) \hat{y}_t}_{\equiv \kappa} + \beta \tilde{E}_t \pi_{t+1} + \underbrace{\frac{(\epsilon - 1) \underline{\tau}}{\psi} \hat{\tau}_t}_{\equiv u_t}. \quad (37)$$

Since $\hat{y}_t = \hat{x}_t$, we obtain the New Keynesian Phillips Curve (22).

D Additional Results to Section 4

Figure 10: Inflation dynamics after a demand shock

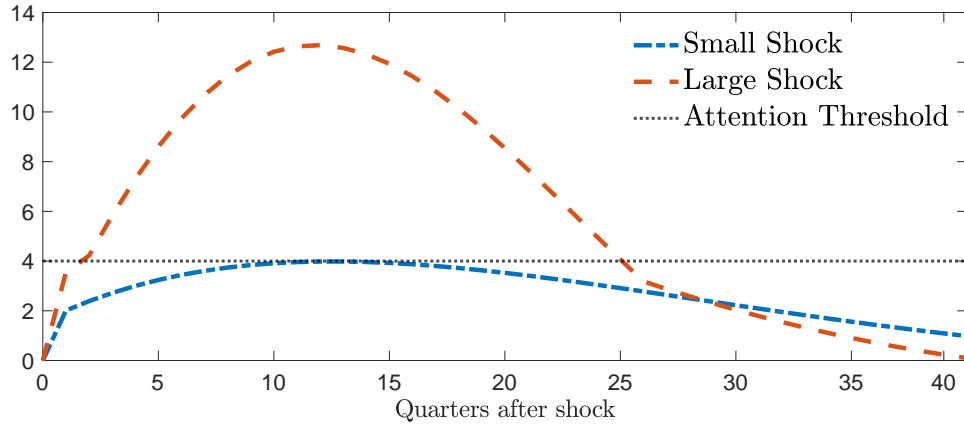


Figure 11: Inflation dynamics under FIRE or absent the attention change

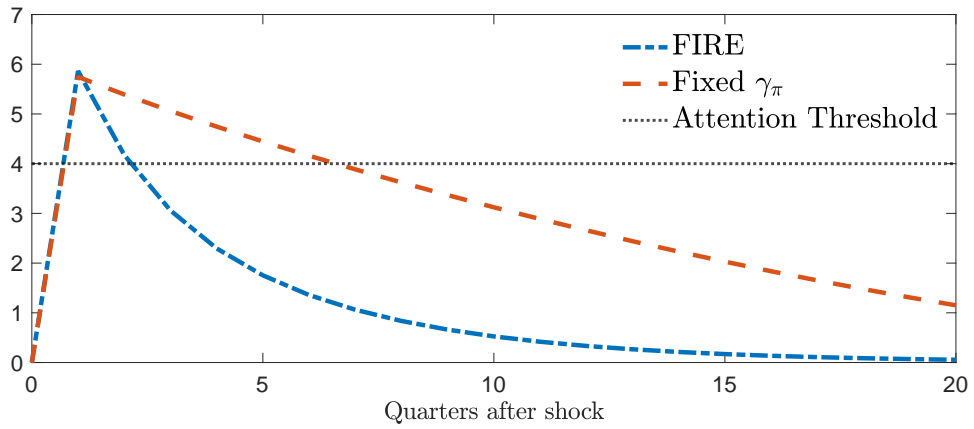


Figure 12: Inflation dynamics with $\gamma_{x,L} = 0.25$ and $\gamma_{x,H} = 0.1$

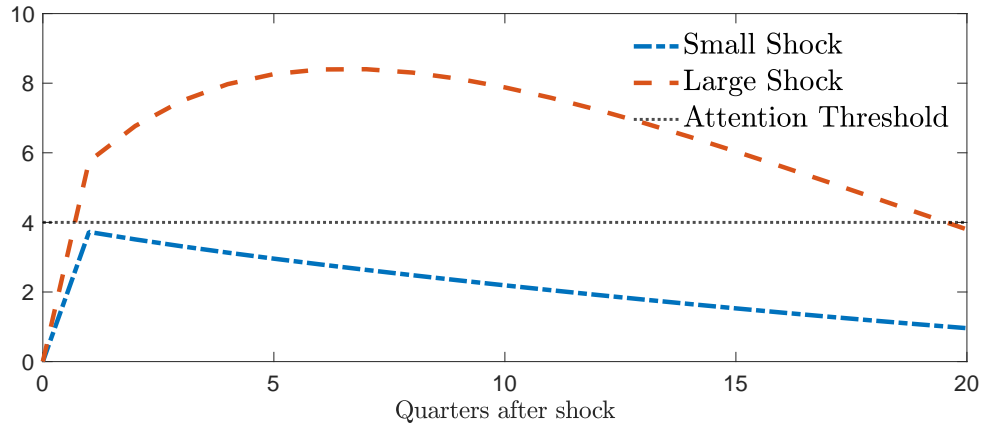


Figure 13: Inflation dynamics with current inflation as the threshold-defining variable

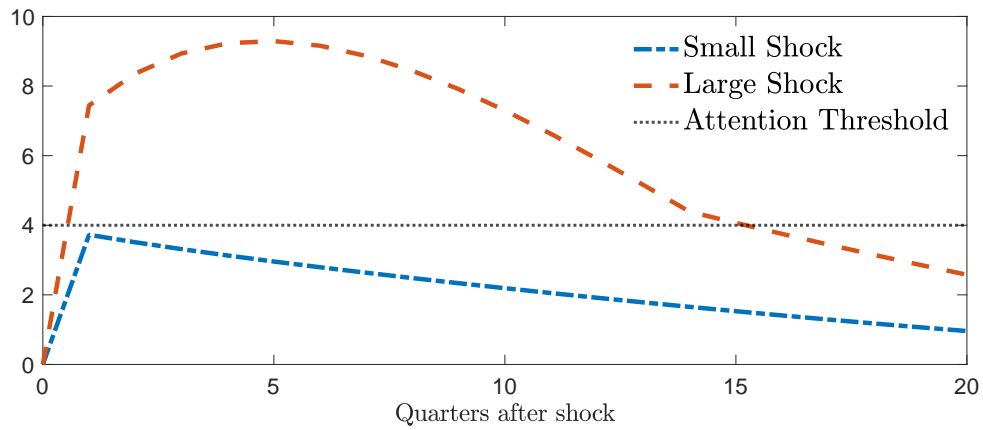


Figure 14: Forecast errors with current inflation as the threshold-defining variable

