The Inflation Attention Threshold and Inflation Surges: Supplemental Appendix

By Oliver Pfäuti*

Appendix to Section I.A

A1. Details and microfoundations

In Section I.A, I show how attention may jump when inflation crosses a certain threshold. In this Section, I provide the mathematical details necessary to arrive at these results. The model builds on Pfäuti (2025).[1] The key novelties are, first, that agents get a certain level of attention for free, which I denote $\gamma_{\pi,L}$, where L denotes that this is a relatively Low level of attention. As discussed in Section I.A, attention γ_π is negatively related to $\sigma^2_{\pi|s}/\sigma^2_\pi$, where σ^2_π denotes the agent's prior uncertainty about inflation π, and $\sigma^2_{\pi|s}$ the posterior uncertainty given the signal s. Second, if the agent wants to increase her attention to inflation above $\gamma_{\pi,L}$, she has two options to do so. First, the agent can choose to pay a fixed cost Λ in which case she can increase her attention to a fixed and known level of attention, $\gamma_{\pi,H}$, where the subscript H indicates that this is a Higher level of attention than the one the agent gets for free. The second option to increase attention above the free level is by paying a standard mutual information cognitive cost and adjusting her attention to inflation smoothly. Conceptually, the fixed cost captures a lumpy one-off action, such as googling "inflation" to obtain the headline result, which delivers a discrete jump in attention to the predetermined level $\gamma_{\pi,H}$. In contrast, the mutual-information cost captures the continuous cognitive effort of processing many small bits of information, such as noticing and processing individual price changes while grocery shopping, which may lead to a smooth increase in attention. When paying the mutual information cost, the agent does not need to pay the fixed cost.

The agent believes that inflation in the next period, π', depends on inflation

* Pfäuti: The University of Texas at Austin, Department of Economics, 2225 Speedway, 78712 Austin TX (pfaeuti.oliver@gmail.com).

[1] The method proposed in Pfäuti (2025) builds on Maćkowiak, Matějka and Wiederholt (2023) and Vellekoop and Wiederholt (2019), see also Weber et al. (2025) for a similar model but with the goal of explaining their RCT findings.

today, π, as follows:[2]

$$\pi' = \tilde{\rho}_\pi \pi + \nu,$$

where $\nu \sim i.i.N.(0, \sigma_\nu^2)$, and $\tilde{\rho}_\pi \in [0, 1]$ denotes the perceived persistence of inflation. Given this perceived law of motion, the full-information forecast π^{e*} is $\pi^{e*} \equiv \tilde{\rho}_\pi \pi$. But because π is not perfectly observable, the actual forecast will deviate from this full-information forecast. These deviations are costly and the agent can pay more attention in order to reduce these deviations. Thus, the agent's choice is not only about how to form her expectations given certain information, but about how to choose and process 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. But processing information is costly, and hence, it cannot be optimal to acquire different signals that lead to identical forecasts. Due to this one-to-one relation of signal and forecast, I will directly let the agent choose the joint distribution of π^e and π, denoted $f(\pi^e, \pi)$, instead of working with the signal. I use $g(\pi)$ to denote the agent's prior, which is assumed to be Gaussian; $\pi \sim N\left(\pi^{prior}, \sigma_\pi^2\right)$. To link the agent's attention choice to the level of inflation, I assume that the agent's prior uncertainty, σ_π^2, increases smoothly (strictly monotonically) with the level of inflation. The agent knows her own subjective prior uncertainty σ_π^2, but she cannot back out the inflation rate π from observing σ_π^2.

Let $U(\pi^e, \pi)$ denote the negative of the utility loss that is incurred when the agent's forecast deviates from the forecast under full information. The agent receives a minimum amount of attention for free, resulting in a fixed ratio of posterior uncertainty to prior uncertainty $\sigma_{\pi|s,L}^2/\sigma_\pi^2$. The agent can reduce her posterior uncertainty further—put differently, the agent can decide to pay costly attention to inflation—in two possible ways. Either by paying the mutual information cost $C(f(\pi^e, \pi))$ and choosing the posterior uncertainty $\sigma_{\pi|s}^2 \in [0, \sigma_{\pi|s,L}^2]$ optimally. The standard assumption in the literature (see, e.g., Maćkowiak, Matějka and Wiederholt (2023)) is that $C(f(\pi^e, \pi))$ is linear in mutual information $I(\pi; \pi^e)$, i.e., the expected reduction in entropy of π due to knowledge of π^e:

$$I(\pi; \pi^e) = \left(H(\pi) - E\left[H(\pi|\pi^e)\right]\right),$$

where $H(\pi) = -\int g(\pi) log(g(\pi)) d\pi$ is the entropy of π, and $H(\pi|\pi^e)$ is the conditional entropy. Given that the agent gets attention $\gamma_{\pi,L}$ for free, the agent only pays the cost when paying additional attention:

$$C(f(\pi^e, \pi)) = \frac{1}{\lambda}\left[I(\pi; \pi^e) - I_{free}\right],$$

[2] Agents' subjective model (or their perceived law of motion) does not necessarily need to be consistent with the actual law of motion (see Andre et al. (2022) and 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). Empirically, AR(1) models for inflation tend to perform quite well (see, e.g., Canova (2007) or Faust and Wright (2013)). Fulton and Hubrich (2021) show that AR(1) models are especially hard to beat in forecasting inflation in real time.

where I_{free} denotes the cost the agent would have to pay to achieve the free level of attention $\gamma_{\pi,L}$, and $\frac{1}{\lambda}$ denotes the parameter that governs the cost of information. Alternatively, the agent can pay the fixed cost Λ in which case she reduces her posterior uncertainty to a fixed ratio of $\sigma^2_{\pi|s,H}/\sigma^2_\pi < \sigma^2_{\pi|s,L}/\sigma^2_\pi$.[3] Taken together, the agent's problem can be written as follows:

$$\text{(A1)} \quad \max_{f(\pi^e,\pi)} \int U(\pi^e,\pi) f(\pi^e,\pi) d\pi d\pi^e - C(f(\pi^e,\pi)) \cdot \mathbf{1}\{\gamma_\pi > \gamma_{\pi,L} \wedge \gamma_\pi \neq \gamma_{\pi,H}\} - \Lambda \cdot \mathbf{1}\{\gamma_\pi = \gamma_{\pi,H}\}$$

$$\text{subject to } \int f(\pi^e,\pi) d\pi^e = g(\pi), \text{ for all } \pi.$$

The indicator functions $\mathbf{1}\{\cdot\}$ capture the alternative attention strategies. If $\mathbf{1}\{\gamma_\pi > \gamma_{\pi,L} \wedge \gamma_\pi \neq \gamma_{\pi,H}\} = \mathbf{1}\{\gamma_\pi = \gamma_{\pi,H}\} = 0$, the agent does not pay any costly attention to inflation. If $\mathbf{1}\{\gamma_\pi = \gamma_{\pi,H}\} = 1$, she pays the fixed cost, and if $\mathbf{1}\{\gamma_\pi > \gamma_{\pi,L} \wedge \gamma_\pi \neq \gamma_{\pi,H}\} = 1$, she pays the mutual information cost. I assume that if the agent decides to pay the mutual information cost, she does not have to pay the fixed cost. This assumption also implies that the agent's optimal attention choice when paying the mutual information cost can be higher or lower than $\gamma_{\pi,H}$. I discuss the case in which the agent always has to pay the fixed cost below. Throughout my analysis, I assume that it is cheaper for the agent to obtain an attention level of $\gamma_{\pi,H}$ by paying the fixed cost rather than the mutual information cost. Otherwise, the fixed cost would be obsolete.

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

$$U(\pi^e,\pi) = -\chi \left(\tilde{\rho}_\pi \pi - \pi^e\right)^2,$$

where χ measures the stakes of making a mistake.

The signals the agent obtains are Gaussian:[4]

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

with noise $\varepsilon \sim i.i.N.(0, \sigma^2_\varepsilon)$. The noise variance σ^2_ε is chosen optimally and reflects the agent's attention.

[3]The fixed cost could also be interpreted through the lens of the System 1/System 2 framework (Kahneman, 2011). In the low-attention regime, agents operate in the fast, heuristic System 1 mode. The fixed cost Λ could then be interpreted as capturing the cognitive cost of switching to the more deliberate System 2 mode—a switch that agents only make when the perceived stakes are high enough. This interpretation is consistent with the experimental evidence discussed in Kahneman (2011) that System 2 engagement is triggered by perceived importance.

[4]In a setup with a quadratic objective function, Gaussian prior and a linear mutual information cost, Gaussian signals are the optimal and unique solution (Matějka and McKay (2015) and Maćkowiak, Matějka and Wiederholt (2023)). This also applies to the interior solution in my case, i.e., the case with $\mathbf{1}\{\gamma_\pi > \gamma_{\pi,L} \& \gamma_\pi \neq \gamma_{\pi,H}\} = 1$. At the corner solution, where attention is equal to the free level of attention $\gamma_{\pi,L}$ or equal to $\gamma_{\pi,H}$, any signal family that yields the required posterior variance is optimal; thus Gaussian signals are still optimal but not uniquely so at that point. To keep the model tractable, I assume that the signals in these cases are Gaussian, too.

Given the Gaussian prior and signal, the optimal forecast is given by $\pi^e = \tilde{\rho}_\pi E[\pi|s]$, where $E[\pi|s]$ is the conditional expectation, and Bayesian updating implies:

$$\pi^e = \tilde{\rho}_\pi (1-\gamma_\pi)\, \pi^{prior} + \tilde{\rho}_\pi \gamma_\pi s, \tag{A2}$$

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

With Gaussian signals, the cost function $C(f(\pi^e,\pi))$ can be expressed as:

$$\begin{aligned} C(f(\pi^e,\pi)) &= \frac{1}{\lambda}\left[I(\pi;\pi^e) - I_{free}\right] = \frac{1}{\lambda}\left[\frac{1}{2}\log\left(\frac{\sigma^2_\pi}{\sigma^2_{\pi|s}}\right) - \frac{1}{2}\log\left(\frac{\sigma^2_\pi}{\sigma^2_{\pi|s,L}}\right)\right] \\ &= \frac{1}{2\lambda}\log\left(\frac{\sigma^2_{\pi|s,L}}{\sigma^2_{\pi|s}}\right), \end{aligned} \tag{A3}$$

for $\sigma^2_{\pi|s} < \sigma^2_{\pi|s,L}$.

For this interior solution, we can express the attention choice problem in terms of γ_π as follows:

$$\max_{\gamma_\pi \in [\gamma_{\pi,L},1] \wedge \gamma_\pi \neq \gamma_{\pi,H}} \left[-\chi\tilde{\rho}^2_\pi(1-\gamma_\pi)\sigma^2_\pi - \frac{1}{2\lambda} log\left(\frac{1-\gamma_{\pi,L}}{1-\gamma_\pi}\right)\right], \tag{A4}$$

where I used $\gamma_\pi = 1 - \frac{\sigma^2_{\pi|s}}{\sigma^2_\pi}$ in the expression for $C(f(\pi^e,\pi))$ in equation (A3) and in $U(\pi^e,\pi)$.

Let γ^{MI}_π denote the solution to the problem in equation (A4), featuring the *M*utual *I*nformation cost. The following Proposition characterizes the optimal attention choice.

PROPOSITION 1: *The optimal attention choice γ^*_π is determined as follows:*

$$\gamma^*_\pi = \begin{cases} \gamma_{\pi,L}, & \textit{if } A \geq B \wedge A \geq C \\ \gamma_{\pi,H}, & \textit{if } B > A \wedge B \geq C \\ \gamma^{MI}_\pi, & \textit{otherwise,} \end{cases}$$

where

$$\gamma^{MI}_\pi = 1 - \frac{1}{2\lambda\chi\tilde{\rho}^2_\pi\sigma^2_\pi} \tag{A5}$$

is the solution to problem (A4)*, and*

$$
\begin{aligned}
A &\equiv -\chi\tilde{\rho}_\pi^2(1-\gamma_{\pi,L})\sigma_\pi^2 \\
B &\equiv -\chi\tilde{\rho}_\pi^2(1-\gamma_{\pi,H})\sigma_\pi^2 - \Lambda \\
C &\equiv -\chi\tilde{\rho}_\pi^2(1-\gamma_\pi^{MI})\sigma_\pi^2 - \frac{1}{2\lambda}\log\left(\frac{1-\gamma_{\pi,L}}{1-\gamma_\pi^{MI}}\right),
\end{aligned}
$$

capture the benefits (minus the attention costs) of the respective attention choices.

Proposition 1 determines the optimal attention choice γ_π^*. To see how the level of inflation affects the agent's optimal attention choice, recall that a higher inflation rate increases the agent's prior uncertainty σ_π^2. Since $\gamma_{\pi,H} > \gamma_{\pi,L}$, we can immediately see that A decreases more strongly with inflation than B because $(1-\gamma_{\pi,L}) > (1-\gamma_{\pi,H})$. Hence, there is an inflation threshold $\bar{\pi}$ at which attention jumps from $\gamma_{\pi,L}$ to $\gamma_{\pi,H}$, unless $C > B$ at $\gamma_\pi^{MI} \in (\gamma_{\pi,L}, \gamma_{\pi,H})$. The sufficient condition for the threshold to emerge is:

$$
\chi\tilde{\rho}_\pi^2\sigma_\pi^2\left(\gamma_{\pi,H} - \gamma_\pi^{MI}\right) > \Lambda - \frac{1}{2\lambda}\log\left(\frac{1-\gamma_{\pi,L}}{1-\gamma_\pi^{MI}}\right), \tag{A6}
$$

for $\gamma_\pi^{MI} \in (\gamma_{\pi,L}, \gamma_{\pi,H})$, which says that increasing one's attention to inflation by paying the fixed cost Λ is relatively cheap compared to paying the mutual information cost. As I will show later, the data favors the threshold-like behavior, indicating that condition (A6) holds in the data. Hence, I calibrate the parameters such that the condition holds. If the agent would always have to pay the fixed cost when increasing her attention above $\gamma_{\pi,L}$, i.e., also when paying the mutual information cost, the threshold-like behavior would always emerge. To see this, note that the sufficient condition in this case would be given by:

$$
\chi\tilde{\rho}_\pi^2\sigma_\pi^2\left(\gamma_{\pi,H} - \gamma_\pi^{MI}\right) > -\frac{1}{2\lambda}\log\left(\frac{1-\gamma_{\pi,L}}{1-\gamma_\pi^{MI}}\right). \tag{A7}
$$

Because the right-hand side of equation (A7) is negative, this condition can only be violated if $\gamma_\pi^{MI} > \gamma_{\pi,H}$. But because σ_π^2 is a smooth function of π, there has to be an inflation level at which attention jumps discretely from $\gamma_{\pi,L}$ to $\gamma_{\pi,H}$, before it may increase to γ_π^{MI} at a higher inflation level.

If condition (A6) holds, we can solve analytically for the threshold $\bar{\pi}$, at which attention jumps from $\gamma_{\pi,L}$ to $\gamma_{\pi,H}$, by setting $A = B$. Solving for the implied prior uncertainty, we obtain the following implicit expression for the threshold:

$$
\bar{\pi} = h^{-1}\left(\frac{\Lambda}{(\gamma_{\pi,H} - \gamma_{\pi,L})\chi\tilde{\rho}_\pi^2}\right), \tag{A8}
$$

where $h(\pi) = \sigma_\pi^2(\pi)$ denotes prior uncertainty as a function of inflation and

$h^{-1}\left(\frac{\Lambda}{(\gamma_{\pi,H}-\gamma_{\pi,L})\chi\tilde{\rho}_\pi^2}\right)$ its inverse evaluated at $\frac{\Lambda}{(\gamma_{\pi,H}-\gamma_{\pi,L})\chi\tilde{\rho}_\pi^2}$. Because the agent cannot invert σ_π^2 to recover π, she does not know the exact inflation rate at which the attention jump occurs; the threshold is defined in her information set only in terms of the observable prior uncertainty (observable for the agent), not in terms of the unobserved π.[5] The expression (A8) shows that the threshold is higher when the fixed cost Λ is higher, as in this case, increasing attention to $\gamma_{\pi,H}$ is costlier. Thus, the benefit of paying the fixed cost has to be higher, which is the case at a higher inflation rate. The threshold is lower when the increase in attention, $(\gamma_{\pi,H}-\gamma_{\pi,L})$, is larger, when the stakes χ are higher, or when the perceived persistence $\tilde{\rho}_\pi$ is higher.

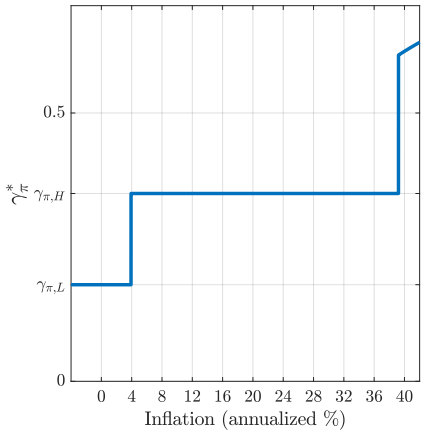
Figure A1 shows γ_π^* graphically as a function of π, for the case in which prior uncertainty is a linear function of inflation, $\sigma_\pi^2 = \underline{\sigma}_\pi^2 + a\cdot\pi$.[6] The Figure illustrates how agents optimally allocate their attention to inflation. At relatively low levels of inflation, agents do not pay more attention than what they obtain for free. The reason is that their prior uncertainty is low in these times and hence, the benefit of increasing their attention is less than the costs of doing so. Once inflation exceeds a threshold, here at $\bar{\pi} = 3.91\%$, the agent pays the fixed cost Λ and her attention jumps to $\gamma_{\pi,H}$. As equation (A8) shows, the threshold is endogenous and depends on the underlying parameters, and the functional form assumption of σ_π^2. Attention then stays constant at $\gamma_{\pi,H}$ even at higher levels of inflation. Up until a second threshold at which attention jumps once more from where on it increases smoothly with inflation. This second jump mirrors the agent's decision to pay the mutual information cost instead of the fixed cost. The reason is that with the fixed cost, the agent's attention is stuck at $\gamma_{\pi,H}$ and hence, the agent's losses from not being fully attentive increase as inflation increases. Once these losses are large enough, the agent decides to pay the mutual information cost which allows her to increase attention even further. Where exactly this threshold lies and to what attention level the agent jumps depends on all of the underlying parameters. My empirical results suggest, however, that this last part has not been relevant in the U.S. since the beginning of my sample. Nevertheless, it illustrates that attention may increase even further if inflation reaches these very high levels.

The following Appendix A.A2 provides an alternative microfoundation in which the fixed cost is a function of inflation and in which the agent needs to pay the fixed cost and the mutual information cost in order to increase her level of attention above the free level of attention. A similar pattern as in Figure A1 emerges.

[5] Note that π is unobservable for the agent, but it is observable for the econometrician. In contrast, the agent observes her prior uncertainty but it is unobservable for the econometrician.

[6] I set $\underline{\sigma}_\pi^2 = 3$, and $a = 0.74$. The rest of the parameters are set as follows: $\gamma_{\pi,L} = 0.18$, $\gamma_{\pi,H} = 0.35$, $\chi = 1$, $\tilde{\rho} = 1$, $\lambda = 0.04$ and $\Lambda = 1$. This results in a threshold $\bar{\pi} = 3.91\%$ at which attention jumps from $\gamma_{\pi,L}$ to $\gamma_{\pi,H}$. These parameters are chosen to be consistent with the empirical results obtained later.

FIGURE A1. OPTIMAL ATTENTION



Note: Notes: This figure illustrates Proposition 1 graphically, i.e., the figure shows optimal attention to inflation, γ^*_π, as a function of the level of inflation π.

A2. An alternative microfoundation

In this section, I provide an alternative microfoundation giving rise to a threshold of attention. Many of the features of this model are the same as the one presented in the previous Section. I will therefore focus on the main differences solely and refer the reader to the previous Section for more details.

As in Section A.A1, agents get a certain level of attention for free, which I denote $\gamma_{\pi,L}$. In addition to the standard mutual information cognitive cost the agent has to pay a fixed cost Λ if she decides to pay more attention than $\gamma_{\pi,L}$. The fixed cost depends negatively on the level of inflation, and I assume that $\Lambda(\pi)$ is strictly decreasing in π. The agent only observes the cost $\Lambda(\pi)$ but she cannot infer π from it, that is, she does not know how π maps into $\Lambda(\pi)$. A key difference from the model in Section A.A1 is that paying the fixed cost itself does not increase the agent's attention (or, equivalently, her obtained signal's precision). A possible interpretation of this fixed cost is that the agent first needs to find out what the reliable information sources regarding inflation are before processing the information and doing so becomes easier when inflation is higher. I assume that all parameters except $\Lambda(\pi)$ (including prior uncertainty) are constant and hence, do not depend on inflation.

The agent's perceived law of motion of inflation, and the loss function $U(\pi^e, \pi)$, are the same as in Section A.A1. Let $C(f(\pi^e, \pi))$ denote the cost of information

function. Then, the agent's problem is given by:

$$\max_{f(\pi^e,\pi)} \int U(\pi^e,\pi) f(\pi^e,\pi) d\pi d\pi^e - C(f(\pi^e,\pi)) - \Lambda(\pi) \cdot \mathbf{1}\{\gamma_\pi > \gamma_{\pi,L}\}$$
$$\text{subject to } \int f(\pi^e,\pi) d\pi^e = g(\pi), \text{ for all } \pi,$$

where $g(\pi)$ is the agent's prior, which is assumed to be Gaussian; $\pi \sim N\left(\pi^{prior}, \sigma_\pi^2\right)$, and σ_π^2 denotes prior uncertainty. The cost function is the same as in Section A.A1. The signals the agent obtains are again Gaussian.[7]

Following the same steps as in Section A.A1, we can express the attention problem in terms of γ_π:

$$\text{(A9)} \quad \max_{\gamma_\pi \in [\gamma_{\pi,L},1]} \left[-\chi \tilde{\rho}_\pi^2 (1-\gamma_\pi)\sigma_\pi^2 - \frac{1}{2\lambda} log\left(\frac{1-\gamma_{\pi,L}}{1-\gamma_\pi}\right) - \Lambda(\pi) \cdot \mathbf{1}\{\gamma_\pi > \gamma_{\pi,L}\} \right],$$

where I used $\gamma_\pi = 1 - \frac{\sigma_{\pi|s}^2}{\sigma_\pi^2}$. Solving the problem in (A9) and defining auxiliary variables A and B:

$$A \equiv -\chi \tilde{\rho}_\pi^2 (1-\gamma_{\pi,L})\sigma_\pi^2$$
$$B \equiv -\chi \tilde{\rho}_\pi^2 (1-\gamma_\pi)\sigma_\pi^2 - \frac{1}{2\lambda} log\left(\frac{1-\gamma_{\pi,L}}{1-\gamma_\pi}\right) - \Lambda(\pi).$$

yields the following Proposition that is akin to Proposition 1, characterizing the *optimal* level of attention.

PROPOSITION 2: *The optimal attention choice γ_π^* is given by:*

$$\text{(A10)} \qquad \gamma_\pi^* = \begin{cases} \gamma_{\pi,L}, & \text{if } A \geq B \\ 1 - \frac{1}{2\lambda\chi\tilde{\rho}_\pi^2\sigma_\pi^2}, & \text{if } A < B. \end{cases}$$

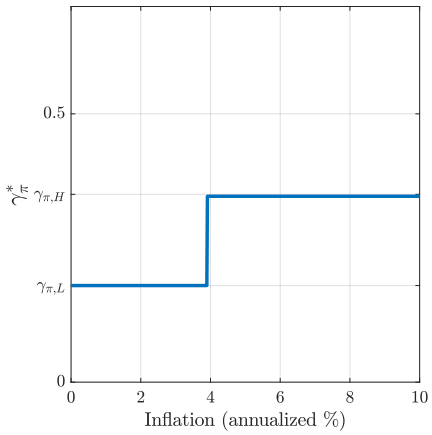
Proposition 2 pins down the optimal attention choice γ_π^*. A higher inflation rate increases B while it leaves A unaffected. Thus, attention jumps discretely from $\gamma_{\pi,L}$ to a higher level given by $\gamma_{\pi,H} = 1 - \frac{1}{2\lambda\chi\tilde{\rho}_\pi^2\sigma_\pi^2}$ when inflation exceeds some threshold $\bar{\pi}$.[8]

[7] The arguments in Matějka and McKay (2015) and Maćkowiak, Matějka and Wiederholt (2023) regarding the optimality and uniqueness of Gaussian signals also apply to the interior solution in my case, i.e., once the fixed cost is paid. At the corner solution, where attention is equal to the free level of attention $\gamma_{\pi,L}$, any signal family that yields the required posterior variance is costless; thus Gaussian signals are still optimal but not uniquely so at that point. To keep the model tractable, I assume that the free signals are Gaussian.

[8] Proposition 2 and the statements following the Proposition require that $\gamma_{\pi,H} \in [\gamma_{\pi,L}, 1]$. If this would not be the case, the agent would never pay more attention than $\gamma_{\pi,L}$.

Figure A2 shows γ_π^* graphically as a function of π.[9] The figure illustrates that optimal attention to inflation takes on two values, depending on the level of inflation. If inflation is relatively low (below 4% in the Figure), $\gamma_\pi^* = \gamma_{\pi,L}$ as the cost of paying attention is too high relative to the benefits of paying more attention. As inflation passes the threshold at 4%, however, attention immediately jumps up to $\gamma_{\pi,H} = 1 - \frac{1}{2\lambda\chi\tilde{\rho}_\pi^2\sigma_\pi^2}$.

Figure A2. Optimal attention



Note: Notes: This figure illustrates Proposition 2 graphically, i.e., the figure shows optimal attention to inflation, γ_π^*, as a function of the level of inflation π.

How high attention is once it jumps, depends on the cost parameter λ (but not $\Lambda(\pi)$), as well as the stakes χ. If the cost of information is lower or the stakes are higher, $\gamma_{\pi,H}$ is higher. Additionally, $\gamma_{\pi,H}$ depends positively on the perceived persistence $\tilde{\rho}_\pi$ and the prior uncertainty σ_π^2. All of these parameters, as well as the fixed cost $\Lambda(\pi)$, also affect the threshold $\bar{\pi}$. Thus, $\gamma_{\pi,H}$ as well as the threshold are endogenous and are the outcome of the attention choice problem.

[9] I assume a linear functional form for $\Lambda(\pi) = \max\{0, a - b\pi\}$ with $a = 0.09$ and $b = 1/112$. The rest of the parameters is set as follows: $\gamma_{\pi,L} = 0.18$, $\chi = 1$, $\tilde{\rho} = 1$, $\lambda = 0.255$ and $\sigma_\pi^2 = 3$. This results in $\gamma_{\pi,H} = 0.35$. These parameters are chosen to be consistent with the empirical results.

Appendix to Empirical Section I

In this section, I discuss a wide range of robustness checks and sensitivity analyses related to my main findings in Section I.

Median expectations.

The results for median expectations are similar, as the second row in Table B1 shows (the first row replicates the baseline results for comparison). Attention in both regimes tends to be somewhat lower when using median expectations and the attention threshold higher. Median expectations tend to be less volatile than average expectations, therefore, it is not surprising that the estimated attention levels—capturing the updating gains—are lower for median expectations. Nevertheless, the null hypothesis that the two attention levels across regimes are equal is marginally rejected (with a p-value of 0.058, see last column). Again, I find that the Bayesian information criterion selects having one threshold only.

Table B1—Estimated attention levels and the attention threshold

	Threshold $\hat{\bar{\pi}}$	Low Att. $\hat{\gamma}_{\pi,L}$	High Att. $\hat{\gamma}_{\pi,H}$	p-val. $H_0 : \gamma_{\pi,L} = \gamma_{\pi,H}$
Mean expectations	3.91%	0.18	0.35	0.000
s.e.		(0.018)	(0.042)	
Median expectations	4.44%	0.16	0.22	0.058
s.e.		(0.018)	(0.030)	
Quarterly frequency	3.21%	0.14	0.38	0.013
s.e.		(0.031)	(0.099)	

Note: Notes: This table shows the results from regression (4), where $\hat{\bar{\pi}}$ 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. Standard errors for the attention estimates are computed using the delta method, based on Newey–West heteroskedasticity- and autocorrelation-robust covariance estimates with 12 lags.

Quarterly observations.

The last two rows in Table B1 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 the baseline monthly specification and again, the difference in attention across regimes is highly statistically significant.

Alternative threshold-defining variables.

When I use average inflation of the last three months as the threshold-defining variable instead of the last month's inflation rate, I estimate a threshold at 3.45% and attention levels of $\gamma_{\pi,L} = 0.18$ and $\gamma_{\pi,H} = 0.33$.

When using the BIC to select whether to use the last-period's inflation rate or the inflation volatility as the threshold-defining variable, the data prefer using the level rather than the volatility. To arrive at this finding, I first estimate a GARCH(1,1) model for quarterly inflation and then use the predicted time-varying volatility as the threshold defining variable. Consistent with my baseline results, I find that the specification with only one threshold is preferred to specifications with multiple thresholds, and the estimated attention parameters are similar: when last-period's volatility was below the threshold, the implied attention parameter is 0.23 and it increases to 0.31 when inflation volatility exceeds the threshold. The threshold is at a volatility of 5.25, indicating that about 17% of the periods fall into the high-attention regime which is lower than the estimated values based on the specification in terms of levels. A key difference of the volatility specification compared to my baseline specification, is that the specification in terms of inflation volatility does not differentiate between periods of pronounced inflation rises and declines. For example, at the outbreak of the Great Financial Crisis, quarterly CPI inflation dropped briefly but quite substantially, leading to a spike in volatility. However, at least households did not seem to pay attention to inflation during that time period when looking at Google Trends data, which may explain why the specification in terms of inflation levels is preferred over the volatility specification.[10]

Alternative surveys.

A drawback of the Survey of Consumers is that it surveys households at most twice. Therefore, it is not possible to examine how individual households update their expectations and how they change their updating behavior with the level of inflation. Focusing on mean (and median) expectations helps to overcome this issue but it also ignores potentially important cross-sectional information. Therefore, I also estimate the attention levels using individual consumer inflation expectations from the Survey of Consumer Expectations. When last month's inflation was below the threshold of 3.91%, I estimate an attention level of 0.08. When inflation was above the threshold, attention increases to 0.33, and I reject the null hypothesis that the two estimates are equal (p-value of 0.000). Thus, I obtain similar, though slightly lower, attention parameters than in my baseline estimates even though this survey is only available since 2013.

When using the firm expectations from the Survey of Firms' Inflation Expectations (SoFIE) from the Cleveland Fed—with the caveat that this data is only available at quarterly frequency and only since 2018Q2, resulting in only 25 observations—I estimate that attention of firms has increased from practically 0.07 up to 0.18 as inflation exceeded the threshold of 3.91%.

[10] Of course, this may look very different for professional forecasters.

A multivariate system of inflation expectations.

As the perceived law of motion of inflation is an AR(1) process, inflation expectations under limited attention only depend on prior beliefs and inflation itself. In principle, however, it could be possible that inflation expectations also depend on other variables such as unemployment.[11] To take this into account, I therefore now estimate the regression:

$$\pi^e_{t+1|t} = \mathbb{1}_{\pi_{t-1}\leq\bar{\pi}}\left[\beta_{0,L} + \beta_{1,L}\pi^e_{t|t-1} + \beta_{2,L}\left(\pi_t - \pi^e_{t|t-1}\right)\right] \tag{B1}$$

$$+ (1-\mathbb{1}_{\pi_{t-1}\leq\bar{\pi}})\left[\beta_{0,H} + \beta_{1,H}\pi^e_{t|t-1} + \beta_{2,H}\left(\pi_t - \pi^e_{t|t-1}\right)\right] \tag{B2}$$

$$+ \beta_3 U^e_{t|t-1} + \beta_4\left(U_t - U^e_{t|t-1}\right) + \tilde{\epsilon}_t, \tag{B3}$$

where $U^e_{t|t-1}$ denotes expectations about unemployment changes and U_t are unemployment changes (see Appendix D.D1 for further details). I find that both, β_3 and β_4, are insignificant ($\hat{\beta}_3 = -0.13$ and $\hat{\beta}_4 = 0.04$ with p-values of 0.332 and 0.337, respectively). The estimated attention-to-inflation parameters are also barely affected. In fact, I estimate attention in the low regime to be equal to 0.18 and 0.34 in the high regime.

Regional variation.

The Survey of Consumers also provides the regions consumers are in, which allows me to leverage regional variation to test for the robustness of my results. In particular, I use the four census regions West, North East, North Central (or Midwest) and South. I control for region fixed effects and estimate that attention is 0.22 with a standard error of 0.023 (standard errors are robust with respect to heteroskedasticity as well as serial and cross-sectional correlation) when last-period's inflation is below the threshold of 3.91%. Attention increases to 0.42 (s.e. of 0.051) when inflation exceeds the threshold. The two estimates are highly statistically significantly different from each other (p-value of 0.000). These results indicate that my baseline results remain robust when considering regional variation.[12]

[11] Kamdar and Ray (2024), for example, show that consumers tend to expect disinflation when expecting an economic expansion.

[12] I do not include time fixed effects in this regression. When I do, I find that if inflation is below the threshold of 3.91%, the estimated attention gain is 0.6 (with a standard error of 0.095), and if inflation is above the threshold, attention is 1.0 (with a s.e. of 0.122). The difference is highly statistically significant with a p-value 0.009.

An alternative attention measure.

In a recent paper, Bracha and Tang (2025) develop an alternative measure of (in)attention to inflation. They use the Michigan Survey of Consumers. In particular, their measure of inattention uses the share of individuals who say that they "Don't know" when asked for an estimate of current inflation. I apply their approach to an extended sample covering 1982-2024. I do this at quarterly and monthly frequency. Additionally, I also do this at the regional level to increase the sample size and to account for possible regional differences.

I then estimate the following threshold regression:

$$\gamma_{BT,t} = \begin{cases} \alpha_L, \text{ if } \pi_{t-1} < \bar{\pi} \\ \alpha_H, \text{ if } \pi_{t-1} \geq \bar{\pi}, \end{cases} \tag{B4}$$

where $\gamma_{BT,t}$ is the Bracha-Tang inattention measure, such that a higher $\gamma_{BT,t}$ implies more *inattention*. I allow for multiple thresholds in the data and use the BIC to select the number of thresholds. As in my baseline specification, I find that the data prefer the specification with exactly one threshold.

Table B2 shows the results. The first row shows that the threshold is with 3.39% somewhat lower than 3.91%. The differences in inattention are significantly different from 0 as the last column indicates. As the second row shows, when I impose a threshold at 3.91%, I find very similar results. The estimated inattention levels are similar as for the specification in row 1 and the levels are again significantly different across regimes. If I additionally include lagged inflation as an independent variable (not shown in the Table), I find that the results are barely affected and that the coefficients related to lagged inflation are insignificant in both regimes (with associated p-values 0.45 and 0.7, respectively). This last finding provides further support that attention moves between two regimes and stays constant within regime.

The middle and lower part in Table B2 show that in all regions and across both monthly and quarterly frequency, inattention is higher when inflation is below the threshold, and in most cases, the differences are significant.

Markov-switching model.

Instead of estimating regression (4), I also estimate a Markov-switching dynamic regression model in which the long-run inflation beliefs $\underline{\pi}$, the perceived persistence $\tilde{\rho}_\pi$ and attention to inflation γ_π are allowed to differ across unobserved states. That implies that I do not need to take a stance on what defines the regimes, in contrast to equation (4) where I assumed that it is last period's inflation rate that determines the regimes (see Chapter 22 of Hamilton (1994) for a textbook treatment).

The implied attention estimates are 0.20 and 0.32 for the two states and thus, close to the estimates from regression (4) of 0.18 and 0.35. When I allow the

TABLE B2—ESTIMATED ATTENTION LEVELS AND THE ATTENTION THRESHOLD

	Threshold $\hat{\bar{\pi}}$	Low Att. $\widehat{\alpha}_L$	High Att. $\widehat{\alpha}_H$	p-val. $H_0 : \alpha_L = \alpha_H$
Estimated threshold	3.39%	0.11	0.08	0.001
s.e.		(0.0038)	(0.0050)	
Imposed threshold	3.91%	0.10	0.08	0.010
s.e.		(0.0059)	(0.0094)	
Monthly frequency	3.91%	0.10	0.08	0.004
s.e.		(0.0056)	(0.0077)	
Regional analysis				
West	3.91%	0.10	0.08	0.277
s.e.		(0.0068)	(0.0129)	
Northcentral	3.91%	0.10	0.07	0.003
s.e.		(0.00624)	(0.0083)	
Northeast	3.91%	0.10	0.07	0.002
s.e.		(0.0071)	(0.0100)	
South	3.91%	0.10	0.08	0.022
s.e.		(0.00683)	(0.0105)	
Regional, monthly				
West	3.91%	0.10	0.08	0.073
s.e.		(0.0065)	(0.0102)	
Northcentral	3.91%	0.10	0.07	0.000
s.e.		(0.0065)	(0.0070)	
Northeast	3.91%	0.10	0.07	0.015
s.e.		(0.0071)	(0.0096)	
South	3.91%	0.10	0.09	0.171
s.e.		(0.0066)	(0.0104)	

Note: Notes: This table shows the results from regression (B4), where $\hat{\bar{\pi}}$ denotes the estimated threshold, $\widehat{\alpha}_L$ and $\widehat{\alpha}_H$ the estimated intercepts when inflation is below or above the threshold, respectively. The last column shows the p-value for the null hypothesis that the two coefficients are equal. Standard errors are robust with respect to heteroskedasticity and serial correlation (Newey and West (1987) with four lags, twelve lags for the monthly specification).

variance to change with the state, I estimate attention parameters of 0.2 and 0.34. The implied probability of being in the low-attention regime from the Markov-switching model is 70% which aligns well with the fact that annualized quarter-on-quarter CPI inflation was below the threshold 70% of the time.

SMOOTH CHANGES IN ATTENTION.

My baseline regression imposes that attention to inflation takes on only two values and jumps from one to the other when inflation exceeds the threshold or falls back below it. Another possibility is that attention changes smoothly with inflation or inflation volatility. To test for this, I estimate the two alternative regressions:

(B5)

$$\pi^e_{t+3|t} = \beta_0 + \beta_1 \pi^e_{t|t-3} + \beta_2 \left(\pi_t - \pi^e_{t|t-3}\right) + \beta_3 \pi_{t-1} \left(\pi_t - \pi^e_{t|t-3}\right) + \varepsilon_t$$

(B6)

$$\pi^e_{t+3|t} = \delta_0 + \delta_1 \pi^e_{t|t-3} + \delta_2 \left(\pi_t - \pi^e_{t|t-3}\right) + \delta_3 \left(\pi_{t-1} - \pi_{t-2}\right)^2 \left(\pi_t - \pi^e_{t|t-3}\right) + \varepsilon_t,$$

where β_3 captures that attention may change with the past level of inflation and δ_3 that it may change with inflation volatility, here measured as the squared change in past inflation.

Again, I use the Bayesian Information Criterion to select which model fits the data best and I find that the data prefer the threshold specification (4) over the two specifications (B5) and (B6). Additionally, even though the estimated β_3 and δ_3 are statistically significantly different from 0, they are close to 0 (0.009 and -0.0002, respectively). In the following, I also test whether a combination of such a 'smooth-attention model' with my threshold model would perform better.

Attention changes within regime.

In Table 1, we saw that attention increases when inflation exceeds the threshold of 3.91%. But what about changes within regime? To look at this, I estimate a time series of attention. In particular, I estimate the regression:

$$\pi^e_{t+3|t} = \beta_0 + \beta_1 \pi^e_{t|t-3} + \beta_2 \left(\pi_t - \pi^e_{t|t-3}\right) + \epsilon_t,$$

using a rolling-window approach, where each window has a length of five years. I denote the estimated time series of attention parameters by $\gamma_{\pi,t}$, and I compute the window-specific average of the monthly quarter-on-quarter inflation rate, π^{avg}_t. To then test whether attention within regime is higher when inflation is higher, I estimate the following regression:

$$\gamma_{\pi,t} = \delta_0 + \delta_1 \mathbb{1}_{\pi^{avg}_t \geq 3.91} + \delta_2 \pi_{t-1} + \delta_3 \mathbb{1}_{\pi^{avg}_t \geq 3.91} \pi_{t-1} + \varepsilon_t, \tag{B7}$$

where $\mathbb{1}_{\pi^{avg}_t \geq 3.91}$ is an indicator that equals one when in period t average inflation over the last twelve months is above 3.91% and zero otherwise. Thus, δ_1 tells us the difference in attention across regimes, δ_2 the effect of last-period's inflation on attention and δ_3 the additional effect of inflation on attention in the high-inflation regime.

Table B3 shows the results. We see that attention is significantly higher in the high-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.

An alternative way to test for changes of attention in addition to the jumps across regimes is to estimate the regression:

$$\begin{aligned}
\pi^e_{t+1|t} =& \mathbb{1}_{\pi_{t-1} \leq \bar{\pi}} \left(\beta_{0,L} + \beta_{1,L} \pi^e_{t|t-1} + \beta_{2,L} \left(\pi_t - \pi^e_{t|t-1}\right)\right) \\
&+ (1 - \mathbb{1}_{\pi_{t-1} \leq \bar{\pi}}) \left(\beta_{0,H} + \beta_{1,H} \pi^e_{t|t-1} + \beta_{2,H} \left(\pi_t - \pi^e_{t|t-1}\right)\right) \\
&+ \beta_3 \left(\pi_t - \pi^e_{t|t-1}\right) \cdot \pi_{t-1} + \beta_4 \cdot \pi^e_{t|t-1} \cdot \pi_{t-1} + \tilde{\epsilon}_t,
\end{aligned}$$

Table B3—Attention changes within regime

	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$
Regression results	0.08**	-0.002	0.001
s.e.	(0.037)	(0.004)	(0.007)

Note: Notes: This table shows the results from regression (B7). 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.

where β_3 and β_4 capture that the perceived persistence and the attention parameter may depend on last-period's inflation in addition to the changes across regimes. I estimate $\hat{\beta}_3 = 0.0037$ (with a standard error of 0.0033 and a p-value of 0.26), and $\hat{\beta}_4 = -0.0031$ (with a standard error of 0.0057 and a p-value of 0.60). When ignoring the effects coming from β_3 and β_4, the attention parameters are $\hat{\gamma}_{\pi,L} = \frac{\beta_{2,L}}{\beta_{1,L}} = 0.20$ and $\hat{\gamma}_{\pi,H} = \frac{\beta_{2,H}}{\beta_{1,H}} = 0.31$. When accounting for the changes in attention through β_3 and β_4 the implied attention parameters are $\hat{\gamma}_{\pi,L} = \frac{\beta_{2,L}+\hat{\beta}_3 \cdot 2}{\beta_{1,L}+\hat{\beta}_4 \cdot 2} = 0.21$ at $\pi_{t-1} = 2\%$ and $\hat{\gamma}_{\pi,H} = \frac{\beta_{2,H}+\hat{\beta}_3 \cdot 5}{\beta_{1,H}+\hat{\beta}_4 \cdot 5} = 0.34$ at $\pi_{t-1} = 5\%$. Overall, these estimated attention parameters are quantitatively close to the ones from my baseline estimation of 0.18 and 0.35. As furthermore the interaction terms are insignificant, these findings suggest that the threshold model of attention captures the data well.

Long-run expectations.

I now turn to long-run expectations. In particular, I consider an exercise that also offers an alternative view and test of my baseline results and the model of short-run expectations. To do so, I rely on the Michigan Survey of Consumers and use the average response to the question "*What about the outlook for prices over the next 5 to 10 years?*" as my measure of long-run expectations. This series is consistently available since 1990. Let me denote this long-run forecast $\pi^e_{LR|t}$. I then run the following threshold regression:

$$\begin{aligned}\pi^e_{LR|t} = {} & \mathbb{1}_{\pi_{t-1}\leq\bar{\pi}}\left(\beta_{0,L} + \beta_{1,L}\pi^e_{t|t-3} + \beta_{2,L}\left(\pi_t - \pi^e_{t|t-3}\right)\right) \\ & + (1 - \mathbb{1}_{\pi_{t-1}\leq\bar{\pi}})\left(\beta_{0,H} + \beta_{1,H}\pi^e_{t|t-3} + \beta_{2,H}\left(\pi_t - \pi^e_{t|t-3}\right)\right) + \epsilon_t,\end{aligned} \tag{B8}$$

which is the exact same regression as my baseline regression (4) except that the dependent variable is now the long-run expectation.

Note that the way I solve the New Keynesian model in Section III is that I assume that agents do not foresee regime changes, and instead act as if they would remain in the current regime forever. Thus, their long-run forecasts (in a

given regime) are given by:

$$\pi^e_{t+h|t} = \underline{\pi}(1-\tilde{\rho}^h_\pi) + \tilde{\rho}^h_\pi \left(\pi^e_{t|t-1} + \gamma_\pi(\pi_t - \pi^e_{t|t-1})\right) + \eta_t. \tag{B9}$$

Therefore, we have $\beta_0 = \underline{\pi}(1-\tilde{\rho}^h_\pi)$, $\beta_1 = \tilde{\rho}^h_\pi$, and $\beta_2 = \tilde{\rho}^h_\pi \gamma_\pi$. Hence, by computing $\frac{\beta_2}{\beta_1}$, I should obtain similar attention estimates as in my baseline regression, and by extension, also a similar threshold.

Table B4 shows the results. The results are indeed quite similar to my baseline results, consistent with the theory. The estimated threshold is at 4.66%, somewhat higher than my baseline (but close to the estimate when using median expectations, see Table B1 above). The estimated attention parameter in times inflation is below the threshold is 0.18, and it increases to 0.33 when inflation exceeds the threshold.[13] These estimates are very close to my baseline estimates of 0.18 and 0.35.

Table B4—Estimated attention levels and the attention threshold for long-run expectations

	Threshold $\hat{\bar{\pi}}$	Low Att. $\hat{\gamma}_{\pi,L}$	High Att. $\hat{\gamma}_{\pi,H}$
Mean expectations	4.66%	0.18	0.33
s.e.		(0.039)	(0.267)

Note: Notes: This table shows the results from regression (B8), where $\hat{\bar{\pi}}$ 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. Standard errors are robust with respect to heteroskedasticity.

Long- and short-run expectations are correlated (see, e.g., Aruoba (2020); Fisher, Melosi and Rast (2025)), and hence, long-run expectations could also affect my attention estimates that are based solely on short-run expectations. To control for this, I re-run the threshold regression (4) while additionally controlling for long-run expectations. Table B5 shows the results. The estimated attention threshold is completely unaffected. The estimated attention parameter in the low-attention regime increases slightly from 0.18 to 0.19, and in the high-attention regime it decreases from 0.35 to 0.3. The coefficients on long-run expectations are 0.08 in the low-attention regime (with a s.e. of 0.034, p-value of 0.02), and 0.06 in the high-attention regime (s.e. 0.06, p-value 0.337). Hence, the coefficient on long-run expectations in the low-attention regime is statistically significant. The magnitude, however, is small. A 1 percentage point higher long-run forecast means that short-run expectations are on average 0.08 percentage points higher (in the low-attention regime).

[13]The standard errors for $\hat{\gamma}_{\pi,H}$ are quite large because the sample here is restricted to after 1990, and hence, there are not that many observations in which inflation was above the threshold.

Table B5—Estimated attention levels and the attention threshold for long-run expectations

	Threshold $\hat{\bar{\pi}}$	Low Att. $\hat{\gamma}_{\pi,L}$	High Att. $\hat{\gamma}_{\pi,H}$
Mean expectations	3.91%	0.19	0.30
s.e.		(0.016)	(0.036)

Note: Notes: This table shows the results from regression (4), while controlling for long-run expectations. Standard errors are robust with respect to heteroskedasticity.

Gasoline prices.

Gasoline prices are a key predictor of inflation expectations (Coibion and Gorodnichenko, 2015). To test whether accounting for gasoline price expectations affects my baseline results, I re-estimate my baseline threshold regression (4) while additionally controlling for the 1-year-ahead gasoline price expectations from the Michigan Survey of Consumers, lagged by one month. This series is available from 1982-1992, and then from 2005 until the end of my sample. The estimated attention parameters are $\hat{\gamma}_{\pi,L} = 0.19$ and $\hat{\gamma}_{\pi,H} = 0.29$ (compared to 0.18 and 0.35 in my baseline estimation). The gasoline-price coefficients are positive in both regimes, but only the one in the low-attention regime is statistically significantly different from 0.[14] The results are very similar when using actual gasoline prices instead of expectations. For this, I use the West Texas Intermediate (WTI) Crude oil price. The implied attention gains are 0.20 and 0.30, and the oil price coefficient is only significant in the low-attention regime. The magnitude of the oil price coefficient is somewhat higher than when considering gasoline-price expectations.

One-year-ahead expectations.

As my baseline in Section I, I focus on one-quarter-ahead expectations and compare them to quarter-on-quarter inflation. When using one-year-ahead expectations and comparing them to year-on-year inflation instead (and using last month's y-o-y inflation), I estimate an attention threshold at 5.03% and that attention increases from 0.54 to 1.22 from the low-attention to the high-attention regime. The difference between the two is highly statistically significant with an associated p-value of 0.000.

Professional forecasters.

Professional forecasters behave quite differently from households and firms. To show this, I run my baseline regression using average 3-month-ahead CPI inflation expectations from the Survey of Professional Forecasters from the Philadelphia

[14]The gas-price-expectations coefficient in times of low attention is 0.015 (with s.e. of 0.007) and in times of high attention, it is 0.008 (with s.e. 0.006). The standard deviation of gas-price expectations over the whole sample is 9.4%. Hence, a coefficient of 0.01 means that a one standard deviation increase in gas price expectations increases inflation expectations by 9.4 basis points.

Fed. According to the Bayesian Information Criterion, the data prefer the specification with only one regime. The estimated attention parameter is 0.13. This very low updating gain is close to the one estimated in Coibion and Gorodnichenko (2012) for professional forecasters, and consistent with evidence of strongly anchored and stable inflation expectations of professional forecasters (see, e.g., Reis (2022*a*,*b*); Coibion and Gorodnichenko (2025)). Through the lens of the model, a likely explanation of these findings is that professional forecasters' prior uncertainty is very low. However, there is also substantive evidence that professional forecasters behave strategically (Gemmi and Valchev, 2023). Such strategic behavior is not captured in my model.

If I enforce a threshold, I estimate this threshold to be at an inflation rate of 3.92%, which is practically identical to the one of 3.91% that I estimate for households. The estimated attention parameters are $\hat{\gamma}_{\pi,L} = 0.07$ and $\hat{\gamma}_{\pi,H} = 0.17$. Thus, professional forecasters also update their short-run inflation expectations more strongly when inflation is above 4%. The differences across regimes and the degrees of updating overall, however, are smaller than for households.

Additional Results and Robustness to Section II

C1. Other shocks

Inflation shock.

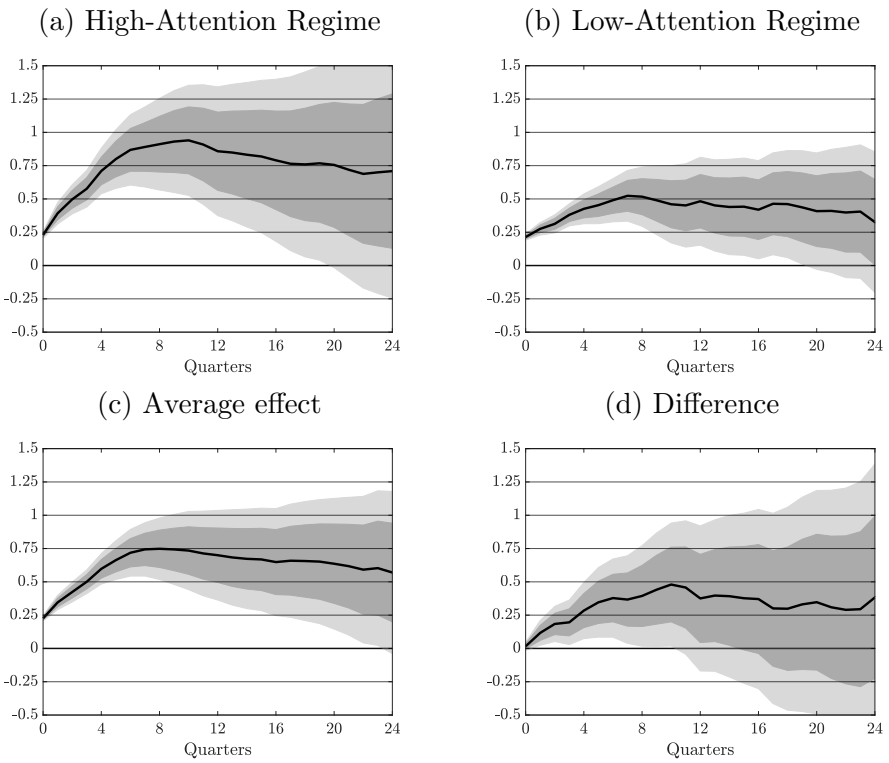
Angeletos, Collard and Dellas (2020) estimate a VAR including 10 key macroeconomic variables and then identify different shocks by maximizing their contribution to the volatility of a given variable over business-cycle frequency (6-32 quarters). I use their shock that contributes most to the volatility of inflation (using the GDP deflator, as in Angeletos, Collard and Dellas (2020)). These shocks are available at quarterly frequency and span the period 1960 until the end of 2017. I use the previous quarter's CPI inflation as an indicator of whether the economy is in the high-attention regime or the low-attention regime (i.e., was the previous annualized CPI inflation rate above or below 4%.)

The dependent variable is the change in the log of the GDP deflator (times 100) from quarter $t-1$ to quarter $t+h$. Figure C1 shows the results. As in the specification with oil news shocks, inflation responds about twice as much to the *inflation shock* when hit in the high-attention regime compared to the low-attention regime. As panel (d) shows, these differences are highly statistically significant and quite persistent: the difference is largest 10 quarters after the shock and still highly statistically significant.

Monetary policy shocks.

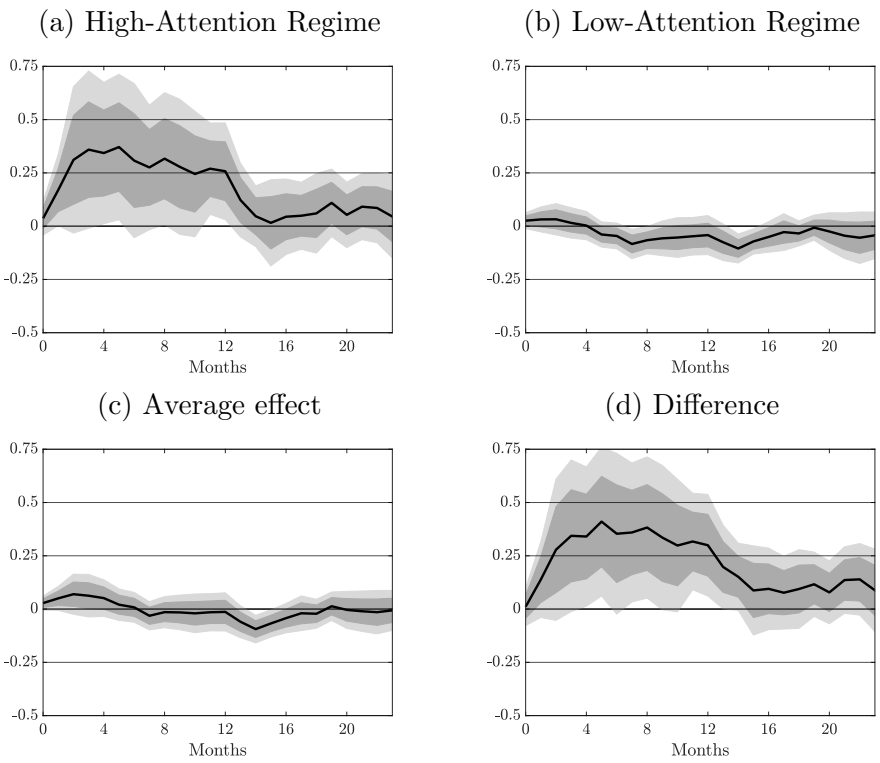
In Figure C2, I show that inflation also responds more strongly in the high-attention regime to monetary policy shocks identified using a high-frequency identification (the shocks are taken from Jarociński and Karadi (2020) and are purged from the information effects of monetary policy statements; I show here the shocks identified using sign restrictions, but the results are practically identical when using the shock series based on the "poor-man approach"). A drawback of using these monetary policy shocks, however, is that they are only available for the period 1990-2019. In particular, there are 49 months out of 353 in which quarter-on-quarter CPI inflation was above 4% during that period, so about 14% of the time, whereas in the full sample the economy spends about 30% of the time in the high-attention regime. Thus, the sample does not include most of the high-inflation periods. Therefore, while the differences across regimes are substantial in magnitude, the differences are less statistically significant.

FIGURE C1. PRICE LEVEL RESPONSE TO THE SHOCK TARGETING INFLATION



Note: Notes: This figure shows the cumulative price level (using the GDP deflator) response to the shock of Angeletos, Collard and Dellas (2020) that targets inflation in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 4 lags). The attention regimes are defined based on the previous quarter's CPI inflation rate.

FIGURE C2. INFLATION RESPONSE TO A MONETARY POLICY SHOCK



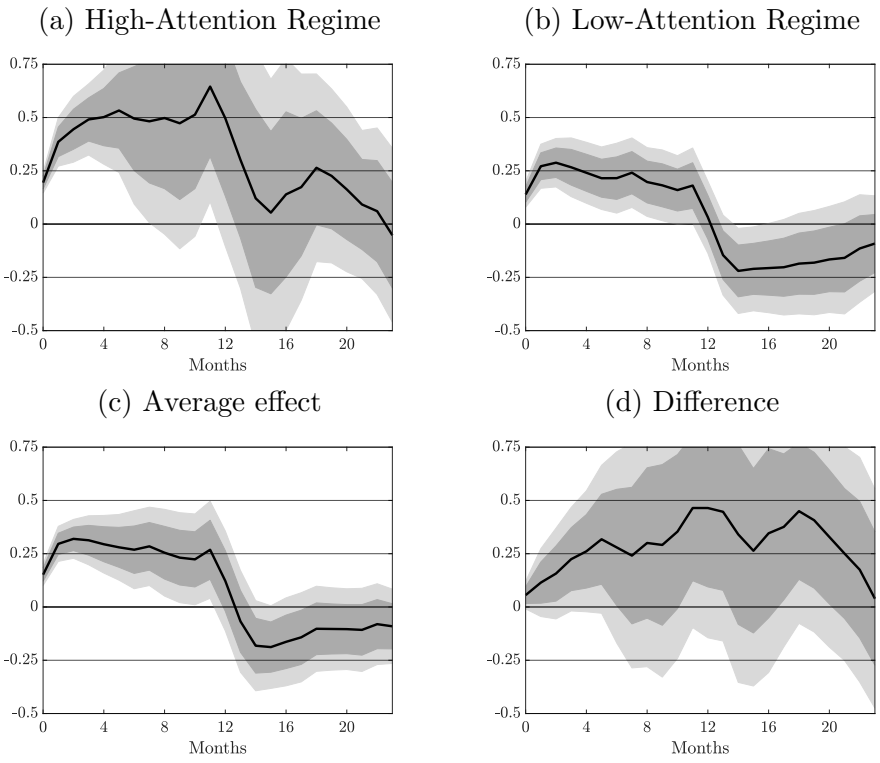
Note: Notes: This figure shows the inflation response to a monetary policy shock in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags). The shocks are from Jarociński and Karadi (2020) and are based on a high-frequency identification and cleaned from information effects of monetary policy.

C2. Robustness analysis

Using Google Trends as the threshold-defining variable.

Instead of using the lagged inflation rate as the threshold-defining variable, I now use Google Trends data to define the regimes. As this data is only available since 2004, the high-attention regime largely coincides with the recent inflation surge. I use the same control variables as in my baseline specification and I further include four lags of the number of Google searches as control variables. Figure C3 shows the results. They confirm the baseline results that inflation responds substantially and significantly more to supply shocks when attention to inflation is high.

Figure C3. Inflation response to an oil supply shock with Google data as threshold-defining variable

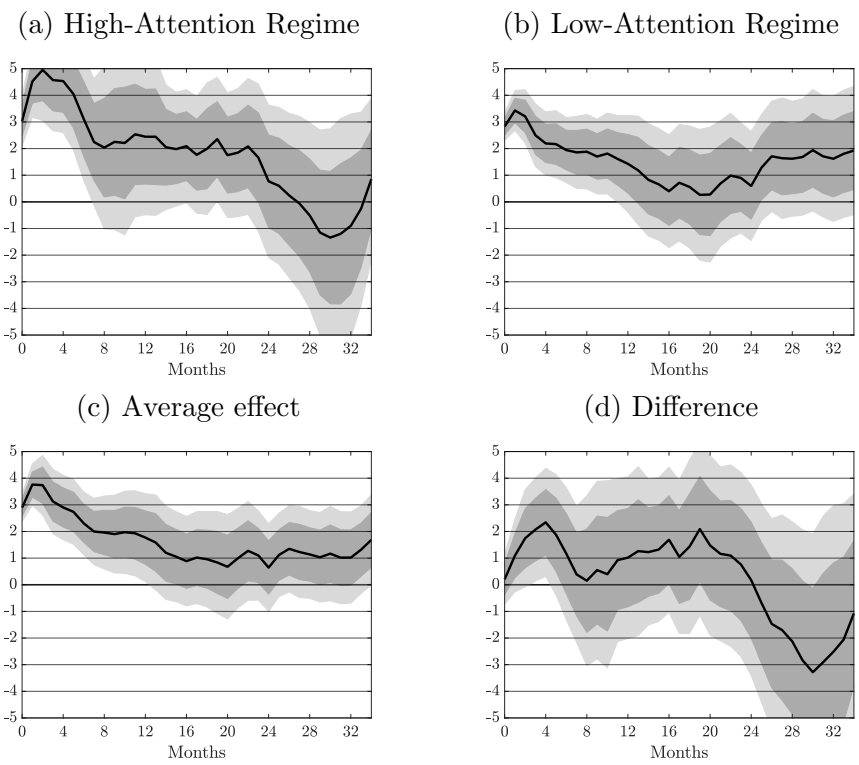


Note: Notes: This figure shows the inflation response to an oil supply news shock in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The attention regimes are defined based on the number of Google searches of the word inflation in the current month. The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags).

Oil price response.

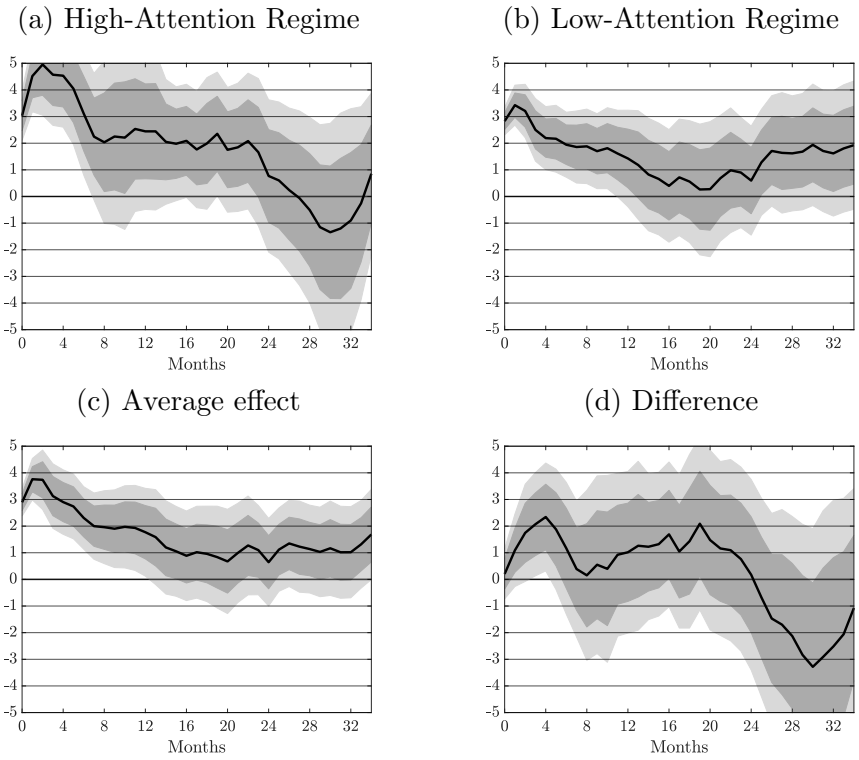
Figures C4 and C5 show the responses of the real and nominal oil price, respectively, to a negative oil supply news shock for the two attention regimes (panels (a) and (b)), the average response in panel (c), and the difference across regimes in (d). We see that the differences across regimes are not significant at the 10% significance level at any horizon, indicating that the differences in the inflation responses discussed in section II are unlikely to be driven by different oil price responses.

Figure C4. Real oil price response to an oil supply news shock



Note: Notes: This figure shows the real oil price response to an oil supply news shock in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags). The attention regimes are defined based on the previous month's inflation rate.

FIGURE C5. NOMINAL OIL PRICE RESPONSE TO AN OIL SUPPLY NEWS SHOCK

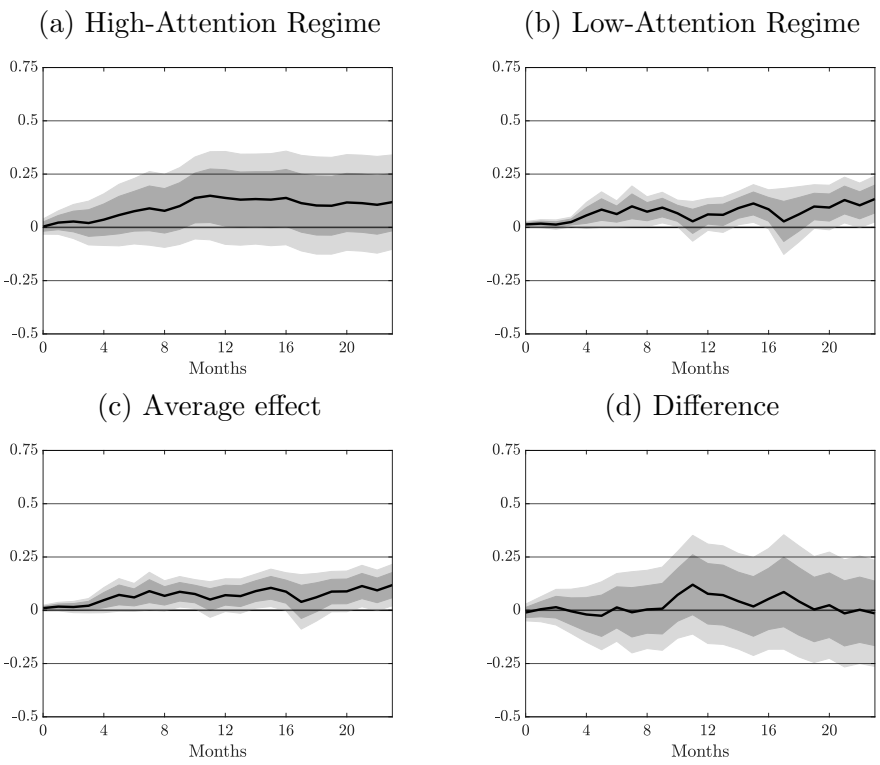


Note: Notes: This figure shows the nominal oil price response to an oil supply news shock in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags). The attention regimes are defined based on the previous month's inflation rate.

Unemployment response.

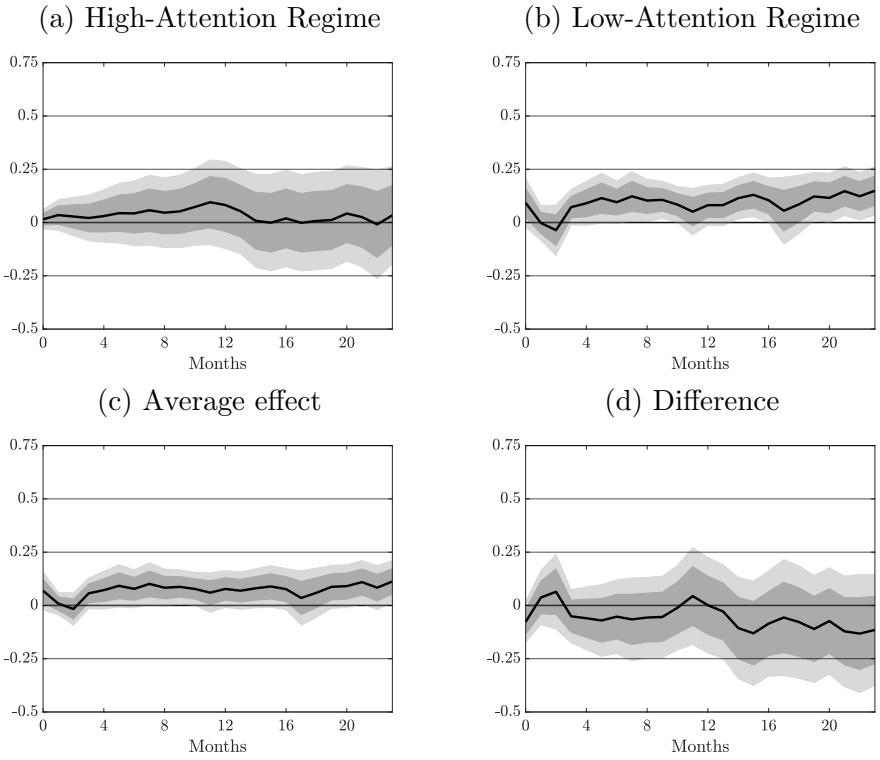
Figures C6 and C7 show the responses of the unemployment rate (excluding the Covid period 2020-2023 and including it, respectively) to a negative oil supply news shock for the two attention regimes (panels (a) and (b)), the average response in panel (c), and the difference across regimes in (d). We see that the differences across regimes are not significant at any conventional significance level at any horizon, indicating that the differences in the inflation responses discussed in section II are unlikely to be driven by generally more responsive macroeconomic variables.

Figure C6. Unemployment response to an oil supply news shock



Note: Notes: This figure shows the unemployment rate response to an oil supply news shock when excluding the Covid-19 period (2020-2023) in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags).

FIGURE C7. UNEMPLOYMENT RESPONSE TO AN OIL SUPPLY NEWS SHOCK INCLUDING COVID

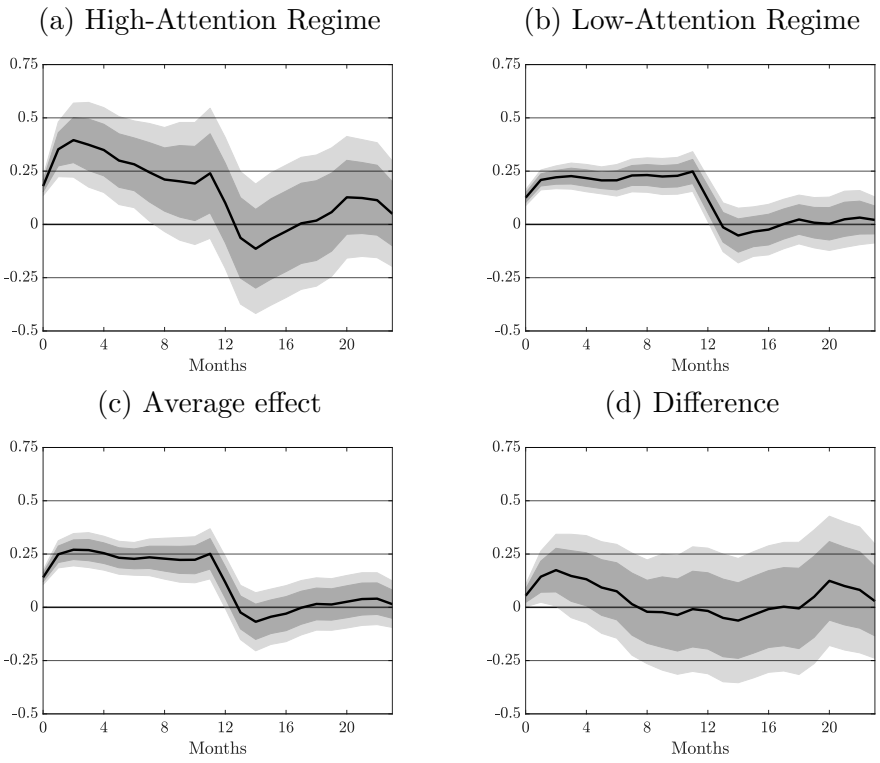


Note: Notes: This figure shows the unemployment rate response to an oil supply news shock when including the Covid-19 period (2020-2023) in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags).

Covid.

Figure C8 shows the inflation response to a negative oil supply news shock for the two attention regimes (panels (a) and (b)), the average response in panel (c), and the difference across regimes in (d) when excluding the Covid period 2020-2023. We see that the differences across regimes are slightly weaker than for the baseline case in Figure 3 but still substantial.

Figure C8. Inflation response to an oil supply news shock excluding Covid

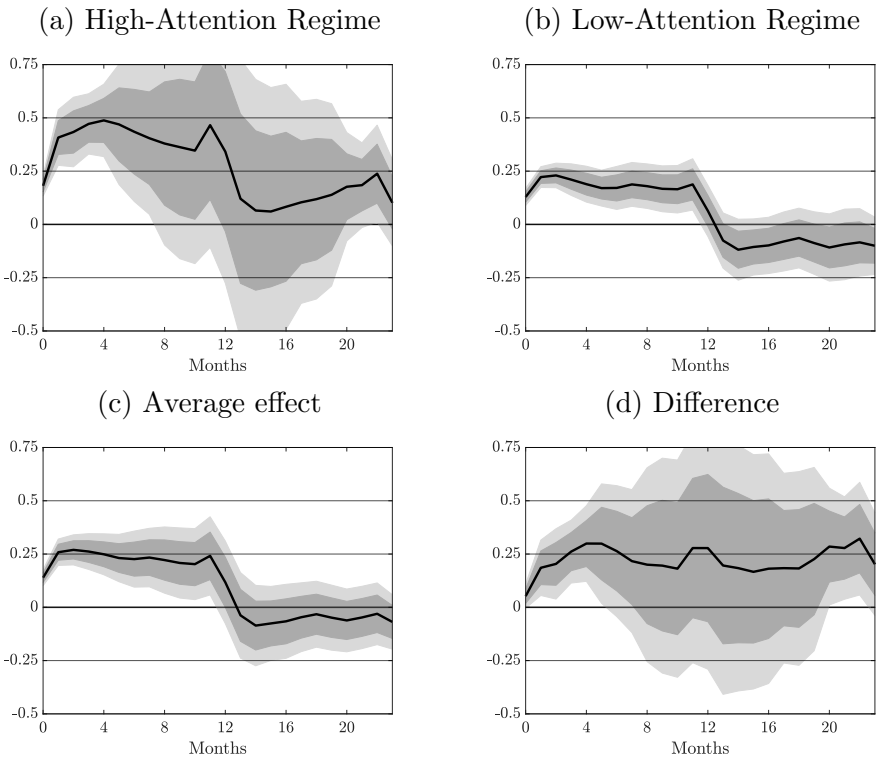


Note: Notes: This figure shows the inflation response to an oil supply news shock when excluding the years 2020, 2021, 2022 and 2023 in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags).

Excluding the Great Inflation period.

Figure C9 shows the inflation response to a negative oil supply news shock for the two attention regimes (panels (a) and (b)), the average response in panel (c), and the difference across regimes in (d) when excluding the Great inflation period, i.e., when excluding observations before 1990. We see that the differences across regimes are even slightly stronger than in the baseline specification.

Figure C9. Inflation response to an oil supply news shock after 1990

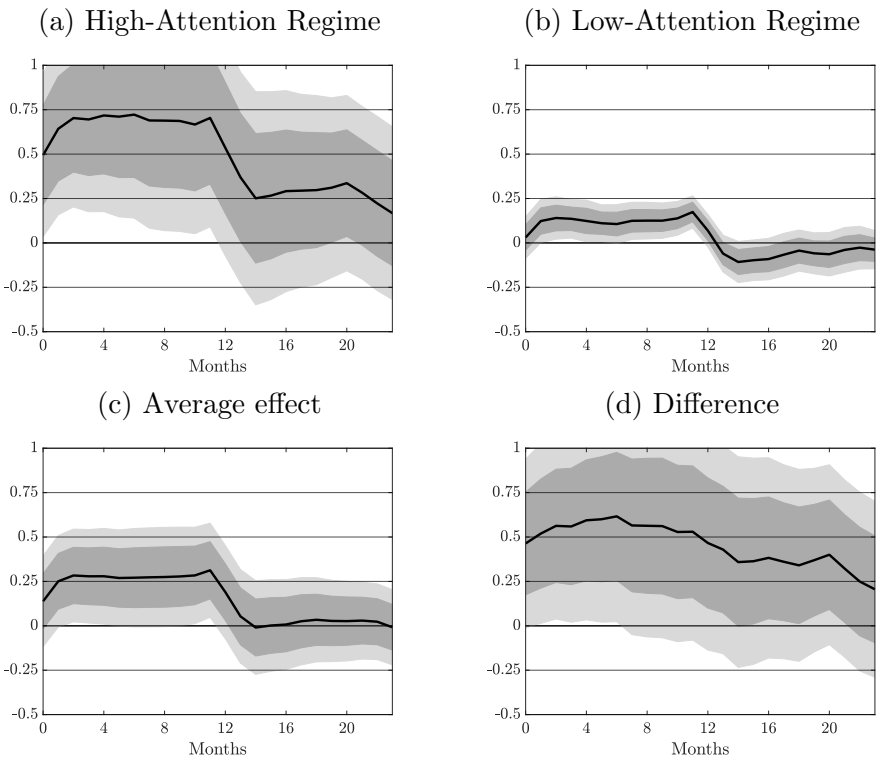


Note: Notes: This figure shows the inflation response to an oil supply news shock when excluding the years prior to 1990, in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags).

Role of control variables.

Figure C10 shows the inflation response to a negative oil supply news shock for the two attention regimes (panels (a) and (b)), the average response in panel (c), and the difference across regimes in (d) when not including any control variables. We see that the differences are still substantial, but the estimates, especially in the high attention regime, are noisier.

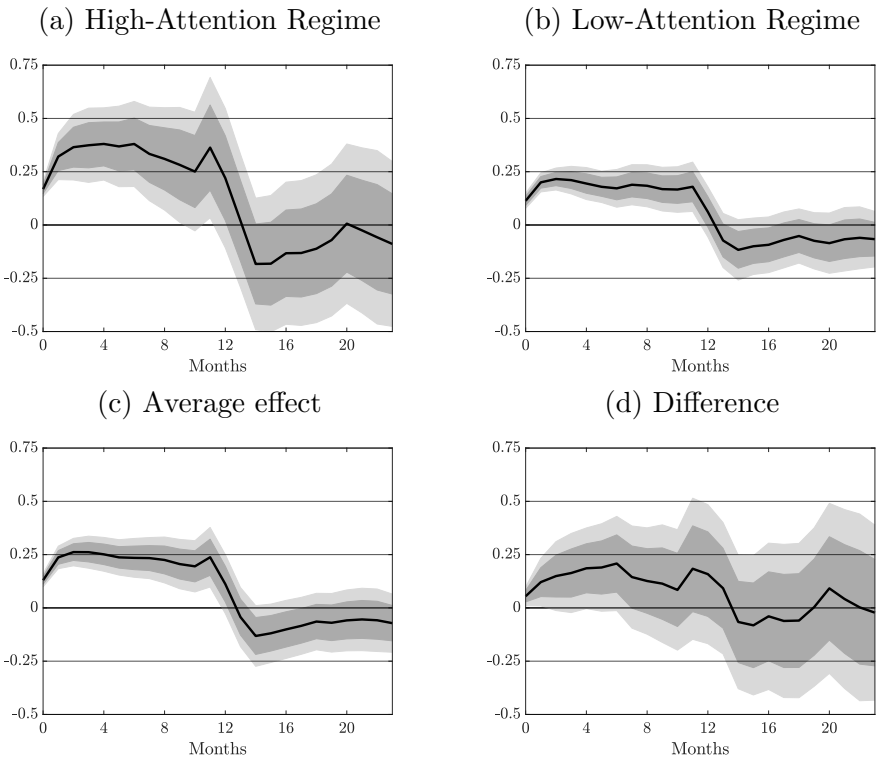
Figure C10. Inflation response to an oil supply news shock (no controls)



Note: Notes: This figure shows the inflation response to an oil supply news shock when using no control variables in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags).

Figure C11 shows the effects when controlling for four lags of the shock, inflation, and the unemployment as well as the interaction of those variables with a regime-dummy (for all four lags). The standard errors for the high-attention regime estimates increase. But the difference across regimes still remains significantly different from 0 at the 10% in the first two months, and the overall results align with the results in Figure 3.

FIGURE C11. CONTROLLING FOR REGIME-DEPENDENT MACROECONOMIC VARIABLES

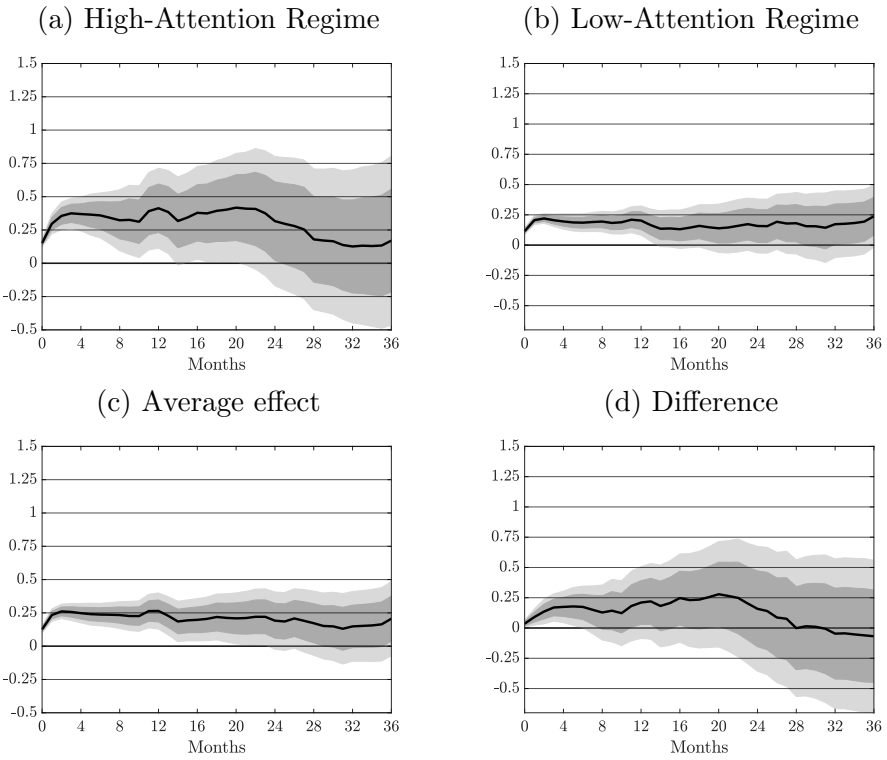


Note: Notes: This figure shows the inflation response to an oil supply news shock (equation (5)) in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The controls include four lags of the shock, inflation, and the unemployment as well as the interaction of those variables with a regime-dummy (for all four lags). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags).

Price-level response.

Figure C12 shows the case where the dependent variable is the cumulative change in the price level: $P_{t+h} - P_{t-1}$, where P_t is the natural logarithm (times 100) of the price level using the CPI. We see from Figure C12 that the results are similar to the baseline case in Figure 3: inflation responds almost twice as strongly in the high-attention regime and the differences are highly statistically significant in the first few months.

Figure C12. Price level response to an oil supply news shock



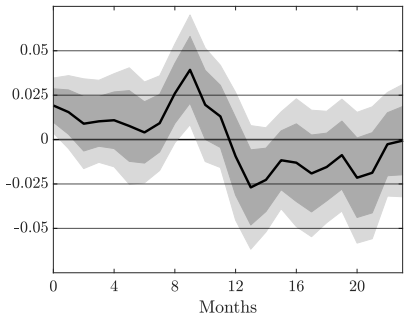
Note: Notes: This figure shows the cumulative price level (using the CPI) response to an oil supply news shock in the high-attention regime (panel (a)), the low-attention regime (panel (b)), on average across regimes (panel (c)), and the difference between the two regimes (panel (d)). The dark shaded areas depict the 68% confidence bands and the light-shaded area the 90% confidence bands. Standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags). The attention regimes are defined based on the previous month's inflation rate.

Alternative attention measures.

Figure C13 shows the interaction effect of the inattention measure based on Bracha and Tang (2025) and the oil supply news shock. The response is normal-

ized to show the effect of a one standard-deviation shock with the inattention measure being one standard deviation above its mean. The coefficients have been normalized such that a higher value indicates a more inflationary effect in times of higher attention. I use regional data to construct the measures and I control for time- and regional fixed effects in the local projections, as well as four lags of the inattention measure and four lags of regional inflation. If inattention is higher, oil supply news shocks are less inflationary and these effects are often significant in the first 10 months. These findings thus support the findings in Section II, using a different measure of (in)attention to inflation.

Figure C13. Alternative measure of inattention and inflation response



Note: Notes: This figure shows the inflation response to an oil supply news shock. In particular, it shows the interaction term of the Bracha-Tang measure of inattention with the shock (for a one standard deviation shock and the Bracha-Tang measure being one standard deviation above its mean) when controlling for time and region fixed effects, as well as four lags of the inattention measure and four lags of regional inflation. The coefficients have been normalized such that a higher value indicates a more inflationary effect in times of higher attention. Standard errors are robust with respect to heteroskedasticity as well as serial and cross-sectional correlation.

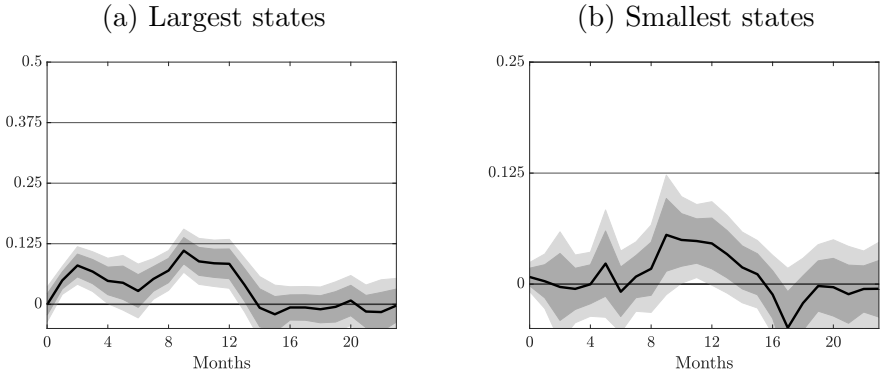
Alternative regional measures of Google Trends.

For my baseline measure of Google Trends at the regional level, I average across all states within the different regions. Within every state, Google Trends are normalized such that the highest value is set to 100. A value of 100 in Texas, however, might not be the same as 100 in Florida. I therefore consider two alternatives. In the first one, for every census region, I use the region's largest state's Google searches as the measure of attention of that region. Second, I use the smallest state's Google searches. Both according to their 2024 population.[15] Thus, the Google Trends data used in these exercises does not feature any within region cross-sectional differences. Differences across regions are absorbed via the

[15]The largest states are New York, Texas, California and Illinois. They account for about 30% of the US population, and about 35% of U.S. GDP according to the BEA in 2023. The smallest states are Vermont, Delaware, North Dakota, and Wyoming.

region fixed effects. Figure C14 shows the results. The results are similar to my baseline estimates, even though the ones using the smallest states (panel (b)) are somewhat noisier.

Figure C14. Interaction with Google Trends using the Largest and Smallest States Only

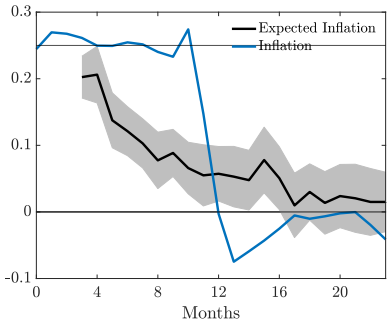


Note: Notes: This figure shows the inflation response to an oil supply news shock. In particular, it shows the interaction term of Google Trends with the shock (for a one standard deviation shock and Google Trends being one standard deviation above its mean) when controlling for time and region fixed effects, as well as four lags of Google Trends and four lags of regional inflation. Instead of using an aggregate measure of Google Trends for every region, I here use Google Trends of the largest (panel (a)) and smallest (panel (b)) state (in terms of population in 2024) within every census region. Standard errors are robust with respect to heteroskedasticity as well as serial and cross-sectional correlation.

Expectations and inflation.

Figure C15 shows the response of inflation forecasts to an adverse oil supply news shock using average expectations from the Survey of Consumers (black-solid line) together with the response of actual inflation (blue-solid line). The response of expectations is shifted such that the vertical distance between the two lines captures forecast errors. Consistent with the model, we see that expectations initially undershoot, followed by a delayed overreaction.

Figure C15. Inflation and expected inflation



Note: Notes: This figure shows the response of inflation forecasts using average expectations from the Survey of Consumers (black-solid line) together with the response of actual inflation (blue-solid line). The response of expectations is shifted such that the vertical distance between the two lines captures forecast errors. The shaded areas depict the 68% confidence bands, standard errors are robust with respect to serial correlation and heteroskedasticity (Newey and West (1987) with 12 lags).

Model Derivations and Calibration of γ_x

Derivation of equation (23).

Linearizing the Euler equation (8) yields:

$$c_t = c^e_{t+1|t} - \left(i_t - \pi^e_{t+1|t} - (z_t - z^e_{t+1|t})\right). \tag{D1}$$

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

In order to express equation (D1) 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 (9), we get that potential output is therefore also constant and equal to:

$$Y_t^* = \Xi^{-1}. \tag{D2}$$

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

$$0 = -\left(r_t - (z_t - z^e_{t+1|t})\right).$$

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

$$r_t^* = z_t - z^e_{t+1|t}. \tag{D3}$$

Substituting $z_t - z^e_{t+1|t}$ with r_t^* in (D1) and using that $c_t = y_t = x_t$, since potential output is 0, yields the IS equation (23):

$$x_t = c^e_{t+1|t} - \left(i_t - \pi^e_{t+1|t} - r_t^*\right). \tag{D4}$$

Note, however, that if we assume initial values $c^e_{0|-1} = x^e_{0|-1}$, it follows that $c^e_{t+1|t} = x^e_{t+1|t}$ because $c_t = x_t$ in equilibrium. This holds true, even if the household does not know that they are equal in equilibrium. We can thus write equation (D4) as:

$$x_t = x^e_{t+1|t} - \left(i_t - \pi^e_{t+1|t} - r_t^*\right), \tag{D5}$$

which is equation (23) in the main text.

Derivation of equation (22).

To derive the New Keynesian Phillips Curve (22), we start by maximizing:

$$\Omega_0(j) = \tilde{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(1-\tau_t) P_t(j) \left(\frac{P_t(j)}{P_t}\right)^{-\epsilon} \frac{Y_t}{P_t} - W_t H_t(j) - \frac{\psi}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - 1\right)^2 Y_t + T_t^F(j)\right],$$

which yields the first-order condition:

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

I define $T_t \equiv 1 - \tau_t$, and set it to eliminate markup distortions in steady state, i.e., $T = \frac{\epsilon}{\epsilon - 1}$. Linearizing around the zero-inflation steady state in which $P(j) = P$ for all j, I obtain:

$$T(\epsilon - 1)\frac{Y}{P}\left[t_t - \epsilon p_t(j) + y_t - (1-\epsilon)p_t\right] \tag{D6}$$

$$= \epsilon MC \frac{Y}{P}\left[(-\epsilon - 1)p_t(j) - (-\epsilon - 1)p_t + mc_t + y_t - p_t\right] \tag{D7}$$

$$-\psi \frac{Y}{P}\left(y_t - p_{t-1}(j) + \pi_t(j) - y_t + p_{t-1}(j)\right) \tag{D8}$$

$$+\beta\psi \frac{Y}{P}\tilde{E}_t^j \pi_{t+1}^j. \tag{D9}$$

Grouping terms, using $MC = 1$, $T = \frac{\epsilon}{\epsilon - 1}$ and dividing by $\frac{Y}{P}$ yields:

$$p_t(j) = \frac{1}{\psi + \epsilon}\left[\psi p_{t-1}(j) + \epsilon\left(mc_t - t_t + p_t\right) + \beta\psi \tilde{E}_t^j \pi_{t+1}^j\right]. \tag{D10}$$

Given the assumptions discussed in Section III, it follows that $\tilde{E}_t^j \pi_{t+1}^j = \pi_{t+1|t}^e$ and $p_{t-1}(j) = p_{t-1}$ such that:

$$p_t(j) = \frac{1}{\psi + \epsilon}\left[\psi p_{t-1} + \epsilon\left(mc_t - t_t + p_t\right) + \beta\psi \pi_{t+1|t}^e\right]. \tag{D11}$$

Therefore, the optimal price $p_t(j)$ is the same for all firms j, $p_t(j) = p_t$, such that we get:

$$\psi \underbrace{\left(p_t - p_{t-1}\right)}_{=\pi_t} = \epsilon\left(mc_t - t_t\right) + \beta\psi \pi_{t+1|t}^e. \tag{D12}$$

Dividing by ψ yields:

$$\pi_t = \frac{\epsilon}{\psi}\left(mc_t - t_t\right) + \beta \pi_{t+1|t}^e. \tag{D13}$$

Using $mc_t = y_t$ and $y_t = x_t$ and defining cost-push shocks as $u_t \equiv -\frac{\epsilon}{\psi} t_t$ and the slope parameter $\kappa \equiv \frac{\epsilon}{\psi}$, I arrive at the linearized New Keynesian Phillips Curve under subjective expectations:

$$\pi_t = \kappa x_t + \beta \pi_{t+1|t}^e + u_t. \tag{D14}$$

D1. 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, de-

crease or to remain about the same over the next twelve months. I follow Carlson and Parkin (1975), Bhandari, Borovička and Ho (2025) and Pfäuti and Seyrich (2023) to translate these categorical unemployment expectations into numerical expectations.

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 $\mathcal{N}\left(\mu_t, (\sigma_t)^2\right)$ 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:

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

which after some rearranging yields:

$$\sigma_t = \frac{2a}{\Phi^{-1}\left(1-q_t^U\right) - \Phi^{-1}\left(q_t^D\right)}$$
$$\mu_t = a - \sigma_t \Phi^{-1}\left(1-q_t^U\right).$$

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 3.91%. This results in estimates $\gamma_{x,L} = 0.254$ and $\gamma_{x,H} = 0.256$. Thus, there are practically no differences in attention to unemployment changes across regimes and hence, I impose γ_x to be the same across regimes in the model.

Additional Figures and Model Results

E1. AS-AD representation

To provide intuition for how the attention threshold can trigger self-reinforcing inflation surges, I use here a slightly stylized version of the model. In particular, I assume that agents are completely inattentive to the output gap, i.e., $\gamma_x = 0$, and that the Taylor rule is given by $i_t = \phi_\pi \pi_t$ with $\phi_\pi > 1$.

The economy starts in the steady state, with $u_0 = 0$, $r_0^* = 0$ and prior expectations at their long-run averages of 0: $\pi^e_{0|-1} = 0$ and $x^e_{0|-1} = 0$. The aggregate supply (AS) curve—derived by plugging in the inflation expectations in the Phillips Curve—in the initial period is given by:

$$AS_0: \quad \pi_0 = \frac{\kappa}{1 - \beta\gamma_{\pi,L}} x_0, \tag{E1}$$

and aggregate demand (AD), which follows from combining the Taylor rule with the aggregate Euler (or IS) equation, is given by:

$$AD_0: \quad \pi_0 = -\frac{1}{\phi_\pi - \gamma_{\pi,L}} x_0. \tag{E2}$$

Panel (a) in Figure E1 depicts this initial situation graphically.[16] Inflation and the output gap are both at their steady state values of 0 and therefore, below the inflation attention threshold $\bar{\pi}$.

In period 1, a positive cost-push shock hits and I assume that it persists for two periods: $u_1 = u_2 > 0$ and returns to zero afterwards, $u_t = 0$ for $t \geq 3$. The AS and AD equations are now given by:

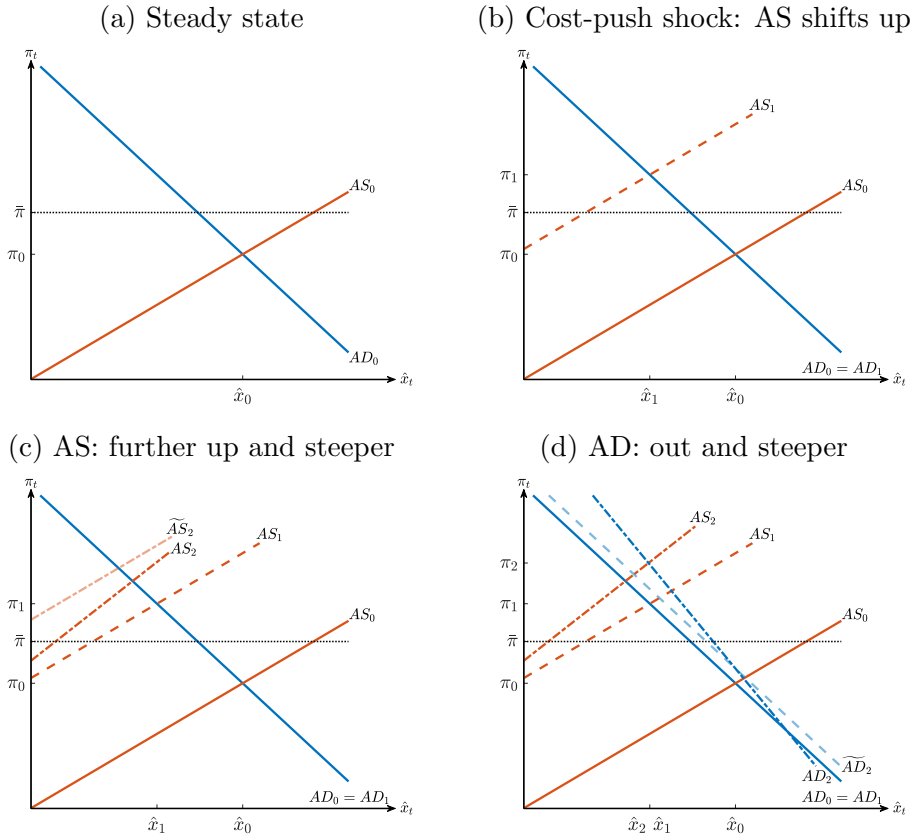
$$\begin{aligned} AS_1: \quad & \pi_1 = \frac{\kappa}{1 - \beta\gamma_{\pi,L}} x_1 + \frac{1}{1 - \beta\gamma_{\pi,L}} u_1 \\ AD_1: \quad & \pi_1 = -\frac{1}{\phi_\pi - \gamma_{\pi,L}} x_1. \end{aligned}$$

This situation is shown in panel (b) of Figure E1. The cost-push shock shifts the AS curve up along the AD curve. The resulting equilibrium is characterized by output below potential, i.e., a negative output gap, and positive inflation. The shock is assumed to be large enough, such that inflation exceeds the threshold.

Due to the increase in inflation in period 1, firms enter period 2 with positive prior inflation expectations: $\pi^e_{2|1} = \gamma_{\pi,L}\pi_1 > 0$. These higher prior expectations together with the still ongoing cost-push shock shift the AS curve further up. This shift is illustrated in panel (c) of Figure E1 by the $\widetilde{AS}_2$ curve, which is given

[16] The values I use in this stylized example are: $\bar{\pi} = 4\%$, $\gamma_{\pi,L} = 0.2$, $\gamma_{\pi,H} = 0.4$, $\beta = 0.99$, $\kappa = 0.6$, $\phi_\pi = 1.05$ and $u_1 = u_2 = 10$.

FIGURE E1. AN ILLUSTRATIVE EXAMPLE



Note: Notes: Panel (a) shows the initial situation in period 0 when the economy is in the steady state. In period 1, a positive cost-push shock hits the economy, leading to an upward shift of the AS curve (panel (b)). In period 2, the AS curve shifts further up (shift from AS_1 to $\widetilde{AS}_2$) and the curve becomes steeper (rotation from $\widetilde{AS}_2$ to AS_2), shown in panel (c). Simultaneously, the AD curve shifts out (shift from AD_1 to $\widetilde{AD}_2$) and becomes steeper (rotation from $\widetilde{AD}_2$ to AD_2), as shown in panel (d). The black-dotted line at $\bar{\pi}$ depicts the inflation attention threshold.

by:

$$\widetilde{AS}_2: \quad \pi_2 = \frac{\kappa}{1-\beta\gamma_{\pi,L}}x_2 + \underbrace{\frac{1}{1-\beta\gamma_{\pi,H}}u_2 + \frac{\beta(1-\gamma_{\pi,H})\gamma_{\pi,L}}{1-\beta\gamma_{\pi,H}}\pi_1}_{\text{Intercept}>0}.$$

The terms denoted "Intercept" capture this shift in the AS curve. Since inflation in the previous period exceeded the attention threshold, attention is now higher. This increase in attention leads to an unambiguously stronger effect of the cost-push shock compared to the case in which attention would have remained

constant:

$$\frac{1}{1-\beta\gamma_{\pi,H}}u_2 > \frac{1}{1-\beta\gamma_{\pi,L}}u_2.$$

The cost-push shock leads to an increase in inflation, and this inflation increase now leads to a larger increase in firms' inflation expectations due to their higher attention. These higher expectations then feed back into higher prices and thus, higher inflation.

The effect of the increase in firm managers' prior expectations on inflation in the second period, captured by $\frac{\beta(1-\gamma_{\pi,H})\gamma_{\pi,L}}{1-\beta\gamma_{\pi,H}}\pi_1$, however, is smaller at the higher attention level. There are two counteracting forces. First, the increase in attention means that firm managers now update their expectations more strongly and put less weight on their prior expectations. This per se leads to a smaller effect. Second, the higher prior increases overall inflation expectations which, ceteris paribus, increases current inflation. But because inflation expectations are discounted by $\beta \leq 1$, the first effect dominates. Therefore, the shift in the AS curve due to the higher prior expectations is smaller at higher attention levels. Quantitatively, however, these differences are very small as $\beta \approx 1$. In fact, if $\beta = 1$, the change in attention has no effect on the shift due to higher prior expectations, and therefore, the total shift of the AS curve is unambiguously higher when attention is higher, due to stronger effect of the cost-push shock.

The shift of the AS curve to $\widetilde{AS}_2$, however, is only part of the story because the increase in attention also steepens the slope of the AS curve. That is, the aggregate supply curve becomes steeper in periods of high attention—a *dynamic* non-linearity. Taking this into account, the AS curve in the second period is given by:

$$AS_2: \quad \pi_2 = \underbrace{\frac{\kappa}{1-\beta\gamma_{\pi,H}}}_{\text{Slope}} x_2 + \frac{1}{1-\beta\gamma_{\pi,H}}u_2 + \frac{\beta(1-\gamma_{\pi,H})\gamma_{\pi,L}}{1-\beta\gamma_{\pi,H}}\pi_1.$$

This steepening of the AS curve is illustrated in panel (c) of Figure E1 by the rotation of the AS curve from $\widetilde{AS}_2$ to AS_2. This steepening of the AS curve eases the inflationary pressures due to the negative output gap. Nevertheless, the steeper AS curve implies that if the AD curve would now shift out, the inflationary effects of this increase in demand would become larger.

It turns out that the higher prior expectations endogenously lead to such a demand increase. This is illustrated by $\widetilde{AD}_2$ in panel (d) of Figure E1, which is given by:

$$\widetilde{AD}_2: \quad \pi_2 = -\frac{1}{\phi_\pi - \gamma_{\pi,L}}x_2 + \frac{(1-\gamma_{\pi,H})\gamma_{\pi,L}}{\phi_\pi - \gamma_{\pi,H}}\pi_1.$$

The higher prior expectations, ceteris paribus, decrease the real rate which leads to the outward shift of the AD curve. These effects, however, are smaller at higher levels of attention as long as the Taylor principle, $\phi_\pi > 1$, is satisfied, because in that case the higher inflation rates due to the higher prior expectations are

counteracted by a more than one-for-one increase in the nominal rate. If monetary policy is relatively dovish, i.e., ϕ_π is close to 1, these differences are small. Furthermore, since the AD curve is now shifted along a steeper AS curve due to the heightened attention, the inflationary effects of a given shift are larger at the higher level of attention. This also implies that additional demand stimulus—for example, due to loose monetary policy or a fiscal stimulus—would have relatively large inflationary effects.

Additionally, the AD curve also becomes steeper, as illustrated by the rotation from $\widetilde{AD}_2$ to AD_2 in panel (d), where AD_2 is given by:

$$AD_2: \quad \pi_2 = -\frac{1}{\phi_\pi - \gamma_{\pi,H}} x_2 + \frac{(1-\gamma_{\pi,H})\gamma_{\pi,L}}{\phi_\pi - \gamma_{\pi,H}} \pi_1.$$

This leads to a further increase in inflation, especially now because the AS curve is steep. In this stylized example, inflation increases substantially from period 1 to period 2 through the change in attention, whereas the output gap remains practically constant. Thus, the attention threshold offers a potential mechanism for how inflation surges may occur and exhibit self-reinforcing dynamics without large changes in output.

After the shock.

In the third period, when the shock has died out, the AS curve shifts back down. The AS curve is given by:

$$AS_3: \quad \pi_3 = \frac{\kappa}{1-\beta\gamma_{\pi,H}} x_3 + \frac{\beta(1-\gamma_{\pi,H})}{1-\beta\gamma_{\pi,H}} \pi^e_{3|2}.$$

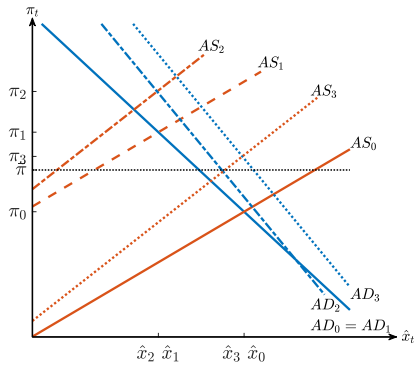
Due to the positive prior expectations, $\pi^e_{3|2} > 0$, the AS curve does not fully shift back to its initial position but remains above it. Due to the increased steepness of the AD curve, however, the shift in the AS curve leads to a stronger reduction in inflation compared to a flatter AD curve.

While the AS curve comes back down, the AD curve shifts further out due to the positive prior expectations that agents have when going into period 3. The AD curve is given by:

$$AD_3: \quad \pi_3 = -\frac{1}{\phi_\pi - \gamma_{\pi,H}} x_3 + \frac{1-\gamma_{\pi,H}}{\phi_\pi - \gamma_{\pi,H}} \pi^e_{3|2}.$$

Thus, output recovers more strongly during this disinflationary period than what would be the case absent this shift in the AD curve, and disinflation occurs more gradually. These results are shown graphically in Figure E2.

FIGURE E2. ILLUSTRATIVE EXAMPLE: AFTER THE SHOCK HAS DIED OUT



Note: Notes: This figure depicts the third period in which the cost-push shock has died out (after being positive for two consecutive periods). The black-dotted line at $\bar{\pi}$ depicts the inflation attention threshold.

TESTING FOR CHANGES IN THE SLOPES OF THE AS AND AD CURVE

I now test the predictions how changes in people's attention to inflation change the slope of the AS and the AD curve. As the output gap is unobserved, I use the unemployment rate and postulate a negative relationship between unemployment and the output gap: $x_t = -\Psi un_t$, where un_t is the unemployment rate and $-\Psi$ is some constant. Abstracting from shocks, in any given period t, the AD and AS curves can then be expressed as:

$$\text{(E3)} \qquad \text{AD:} \quad un_t = \frac{(\phi_\pi - \gamma_{\pi,r})}{\Psi}\pi_t - \frac{(1-\gamma_{\pi,r})}{\Psi}\pi^e_{t|t-1}$$

$$\text{(E4)} \qquad \text{AS:} \quad \pi_t = -\frac{\kappa}{1-\beta\gamma_{\pi,r}}\Psi un_t + \frac{\beta(1-\gamma_{\pi,r})}{1-\beta\gamma_{\pi,r}}\pi^e_{t|t-1}.$$

Note that I represent the AD curve with unemployment on the left hand side of the equation, not inflation. This will help to illustrate the empirical approach I take to estimate the slope of the AD curve. The theory predicts that a given change in inflation should lead to a smaller increase in the unemployment rate when attention to inflation is higher, as can be seen from the AD curve. The reason is that higher inflation leads to a more than one-for-one response of nominal interest rates via the Taylor rule and the resulting higher real rates increase unemployment. When households are more attentive, however, their inflation expectations increase more strongly, and hence, the ex-ante real rate increases by less, so that unemployment increases by less.

For the slope of the AS curve, the theory predicts that a given decrease in the unemployment rate leads to more inflation when firm managers are more attentive. The reason is that their inflation expectations increase more strongly

and hence, they increase their prices by more.

I test those predictions following the approach in Barnichon and Mesters (2021) and Barnichon and Mesters (2020). In all specifications, I use regional data on unemployment and CPI inflation, and I control for region specific fixed effects, as well as for past inflation expectations, as in equations (E3) and (E4).

To test the predictions regarding the AD curve, I regress the current unemployment rate on contemporaneous inflation and instrument inflation with contemporaneous supply shocks. In particular, I again use the oil supply news shocks from Känzig (2021).

Columns (1) and (2) in Table E1 show the results. We see that in both regimes the estimated slope is significantly positive in both regimes and consistent with the theory, it is flatter in times of high attention. The robust F-statistic is larger than 20 in both regimes.

Table E1—Empirical AS and AD curve slopes

	AD slope		AS slope	
	$\pi_{t-1} < 3.91$	$\pi_{t-1} \geq 3.91$	$\pi_{t-1} < 3.91$	$\pi_{t-1} \geq 3.91$
	(1)	(2)	(3)	(4)
Slope estimates	0.25	0.17	−0.38	−0.74
s.e.	(0.147)	(0.103)	(0.113)	(0.091)
p-value	0.084	0.088	0.001	0.000
N	1,316	408	728	148
Control for $\pi^e_{t\|t-1}$	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Robust F-stat.	20.89	28.92	10.92	17.28

Note: Notes: This table shows the estimated slopes of the AD (columns (1) and (2)) and the AS curves (columns (3) and (4)) for the two attention regimes. The last row provides the heteroskedasticity-robust first-stage F–statistic.

To test the predictions regarding the slope of the AS curve, I regress inflation on the unemployment rate and instrument unemployment using monetary policy shocks, as in Barnichon and Mesters (2021). I use the shocks from Jarociński and Karadi (2020) as well as from Aruoba and Drechsel (2024) to get more power. Consistent with the findings in Barnichon and Mesters (2021), I find very weak relations when considering contemporaneous effects as well as concerns regarding weak instruments. This is not surprising given that monetary policy shocks are often claimed to have delayed effects on unemployment (Aruoba and Drechsel, 2024). Barnichon and Mesters (2021) show that the relationship between unemployment and inflation strengthens when considering cumulative changes in those variables over longer horizons. I therefore follow their approach and use the 12-month-ahead cumulative change in the CPI as my dependent variable and the 12-month-ahead cumulative change in the unemployment rate. For the instruments, I include the contemporaneous shocks as well as six lags of the shocks.

Columns (3) and (4) in Table E1 show the results. In both regimes, the slope is significantly negative, and as predicted by the theory, the slope steepens in times of high attention. The robust F-statistics are > 10 in both cases, even though they are weaker than when considering oil shocks. This evidence is consistent with the findings in Barnichon and Mesters (2021), who show that the slope of the AS curve has indeed been flatter after 1990 than before. The time before 1990 largely coincides with the high attention regime (their sample does not include the recent inflation surge).

A simple back of the envelope calculation shows that the very simple model in equations (E3) and (E4) does surprisingly well also quantitatively. The slope of the AD curve is $\frac{\phi_\pi - \gamma_\pi}{\Psi}$. If we fix $\phi_\pi = 2$, we can back out Ψ from the estimate of the slope in times of low attention. Doing so implies $0.25 = \frac{\phi_\pi - \gamma_\pi}{\Psi} = \frac{2-0.18}{\Psi}$, from which we obtain $\Psi = 7.3$. Using this Ψ implies a slope in times of high attention of $\frac{2-0.35}{7.3} = 0.22$ which is not too far from the empirical estimate of 0.17, even though the empirical findings suggest a somewhat more pronounced change in the slope of the AD curve across regimes. The AS curve slope is equal to $-\Psi \frac{\kappa}{1-\beta\gamma_\pi}$. Using my baseline calibration, I obtain a slope in times of low attention of -0.5 and in times of high attention of -0.64. Again, the change is somewhat stronger empirically, but given the simplicity of this model and that those predictions are untargeted, the model does surprisingly well.

E2. Additional figures to Section IV.A

In this section, I discuss alternative model specifications and additional sensitivity analyses related to the model results in Section IV.A.

The long last mile of disinflation.

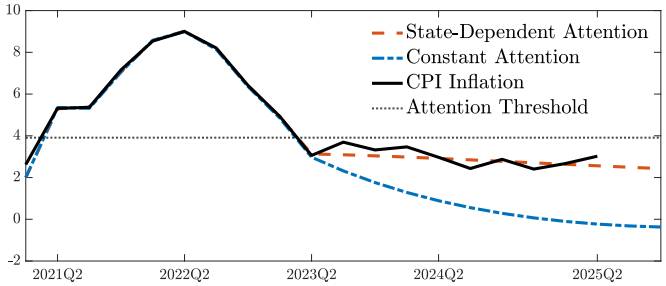
To provide an estimate of the quantitative role of the attention threshold for the recently observed long last mile of disinflation, I consider the following counterfactual. I feed in cost-push shocks to my model to perfectly replicate the recent inflation surge up until inflation fell back below the threshold. After that, I assume that there were no more shocks and examine what inflation dynamics the model produces endogenously. I do this exercise for my model featuring the attention threshold and the counterfactual model in which attention remains high after the inflation surge. Note that replicating the inflation surge before the return to low attention is important for this exercise as the inflation surge shapes people's prior expectations which are an important mechanism for the long last mile of disinflation in my model.

Figure E3 shows the results. The black-solid line shows actual U.S. CPI inflation, the red-dashed line shows my baseline model with the attention threshold, and the blue-dashed-dotted line shows the model where attention is constantly high. Up until 2023Q2, the shocks are set to match the actual inflation perfectly. The rest of the inflation dynamics are endogenously determined within the respective model. We can see that the model with the attention threshold matches the long last mile of disinflation almost perfectly. The model where attention remains high, on the other hand, performs poorly as it would predict a very sharp phase of disinflation and even deflation. Of course, there were exogenous shocks hitting the U.S. economy also after 2023Q2 that are not directly considered in this exercise that may have affected the last mile as well. The results in Figure E3, however, indicate that the net effects of these other forces may be small. If attention remains higher after the inflation surge, this would dampen the role of the attention threshold for the last mile of disinflation. But as long as attention does decrease, it contributes to the longer last mile. I discuss this possibility more extensively below in the discussion of Figure E9.

Alternative model specifications.

Figure E4 shows the inflation response to a cost-push shock for five different models. My model with the attention threshold (labeled "Baseline with threshold" and depicted by the red-dashed line) shows the discussed hump-shaped pattern of the inflation surge and the decrease in the speed of disinflation once inflation falls back below the threshold. As the other four illustrate, these inflation dynamics are unique to the model with the attention threshold. In particular, the models under FIRE (black-solid line) or in which attention remains low (blue-dashed-dotted line) predict that inflation peaks on impact and then comes down

FIGURE E3. THE LAST MILE OF DISINFLATION



Note: Notes: This figure shows the role of the attention threshold for the last mile of disinflation.

relatively quickly. The same is true for the model with a constant attention parameter of 0.24, which corresponds to the estimated attention level in the data if I do not impose a threshold (the inflation response is captured by the blue-dotted line). Additionally, models with FIRE could not account for the observed dynamics of forecast errors. A model variant in which attention to inflation is always high (red-dotted), on the other hand, can produce the initial hump-shaped response but then predicts a very fast decline back to target, i.e., this model would not explain why we see such a long last mile of disinflation. Furthermore, a model in which attention is always high would have predicted a very pronounced deflationary period after the Great Financial Crisis. However, one thing that the model with high-attention illustrates, is that the inflation response is hump shaped when attention is high whereas it peaks on impact when the shock hits in times of low attention, consistent with the empirical findings (panels (a) and (b) in Figure 5).

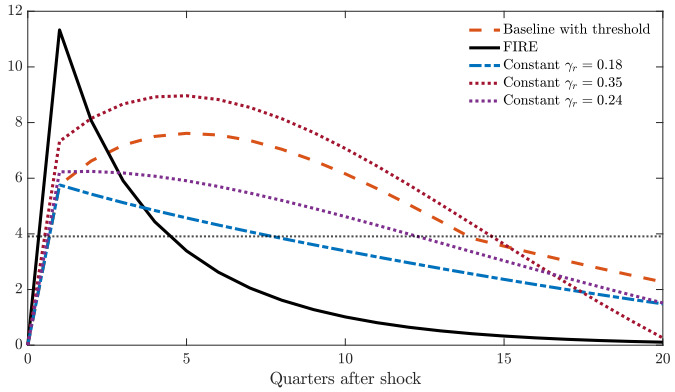
DEMAND SHOCKS.

Figure E5 shows the inflation response to a positive demand shock. We see that the discussed inflation dynamics after a relatively large supply shock are similar to the ones after a demand shock: inflation keeps on increasing, fueled by the attention increase, and the speed of disinflation slows substantially once inflation falls back below the threshold and attention therefore decreases. Note, however, that the inflation surge is somewhat longer lived than in response to the supply shock.

TIME-VARYING ATTENTION TO THE OUTPUT GAP.

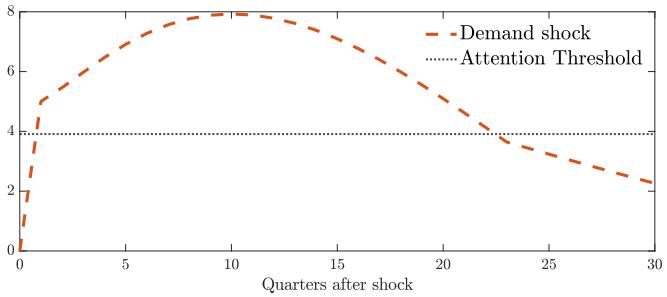
Figure E6 shows the inflation response for the case in which attention to the output gap, γ_x, decreases when attention to inflation increases. This may reflect that agents shift their attention from the output gap to inflation when inflation is high. We see that the inflation surge becomes more persistent in that case

FIGURE E4. INFLATION DYNAMICS UNDER FIRE OR ABSENT THE ATTENTION CHANGE



Note: Notes: This figure shows the inflation response to a cost-push shock for different model specifications.

FIGURE E5. INFLATION DYNAMICS AFTER A DEMAND SHOCK



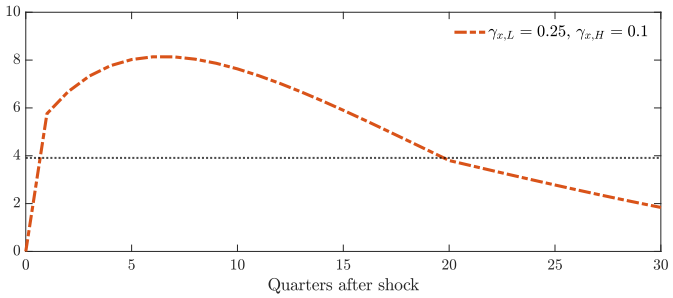
Note: Notes: This figure shows the inflation response in the model with the attention threshold in response to a demand shock r_t^* with persistence $\rho_r = 0.55$.

(note that the x-axis now shows 30 periods rather than 20). The reason is that the negative output gap due to the adverse cost-push shock leads to a smaller downward revision of output gap expectations, thus, agents remain relatively more optimistic about the output gap, and hence, consume relatively more, leading to additional upward pressure on inflation.

RATIONAL CONSUMPTION EXPECTATIONS.

Figure E7 shows the inflation response to a cost-push shock when consumption expectations are formed under full-information rational expectations. We see that the dynamics are qualitatively similar to the baseline model. The endogenous inflation increase is somewhat muted here because the forward-looking consumption expectations do not induce endogenous persistence in the way they

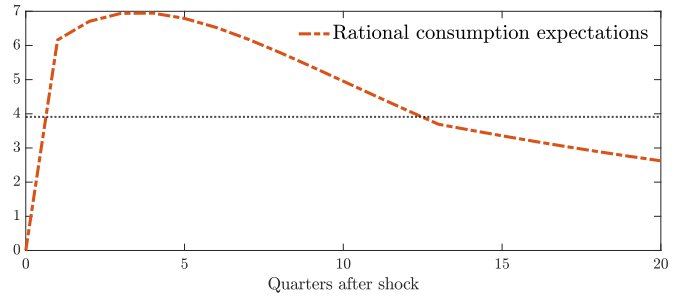
FIGURE E6. INFLATION DYNAMICS WITH $\gamma_{x,L} = 0.25$ AND $\gamma_{x,H} = 0.1$



Note: Notes: This figure shows the inflation response to a cost-push shock when attention to the output gap decreases in times of high attention to inflation.

do in the baseline model. When increasing the exogenous persistence of the shock slightly, however, the models are very similar (not shown).

FIGURE E7. INFLATION DYNAMICS WITH RATIONAL CONSUMPTION EXPECTATIONS



Note: Notes: This figure shows the inflation response to a cost-push shock when consumption expectations are formed under full-information rational expectations.

SMOOTH ATTENTION CHANGES.

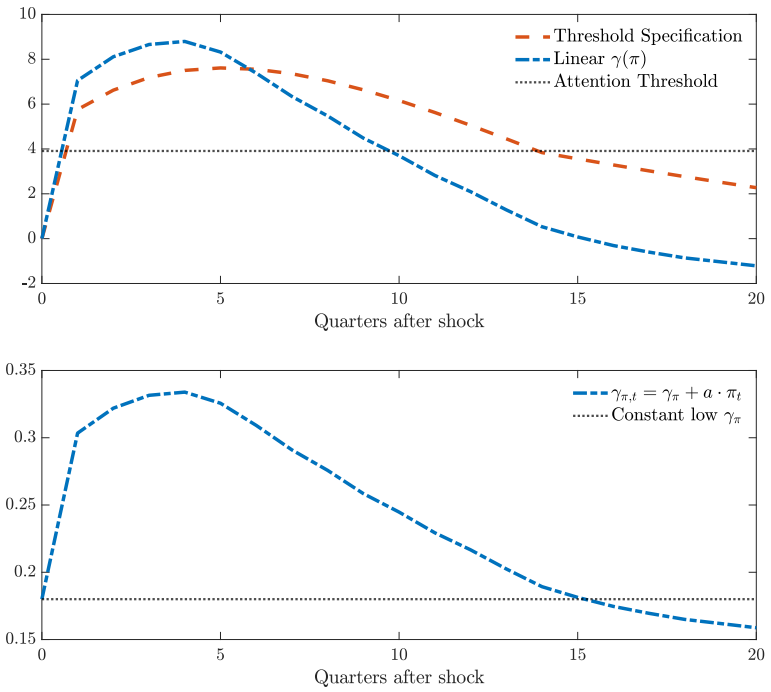
What happens if attention changes smoothly with inflation? The upper panel in Figure E8 shows the inflation response to a cost-push shock when attention is a linear function of inflation (blue-dashed-dotted line) together with the inflation response in my baseline model with a threshold (red-dashed line). The specification I use for attention is $\gamma_{\pi,t} = \gamma_{\pi,L} + a \cdot \pi_t$, where $\gamma_{\pi,L} = 0.18$ is set equal to the attention estimate I obtain in the low-attention regime. I set $a = 7$ in which case $\gamma_{\pi,t}$ peaks at a similar level as my attention estimate in the high-attention regime.[17] The lower panel of Figure E8 shows the time series of $\gamma_{\pi,t}$ in response

[17]Note, here π_t is quarterly inflation in decimals; whereas the figures report annualized percent.

to the cost-push shock.

We see that the two models predict similar though slightly different inflation dynamics. Both predict a hump-shaped inflation response, a relatively fast first phase of disinflation which then slows down. The model with a continuous $\gamma(\pi_t)$, however, predicts a somewhat shorter-lived inflation surge. The reason is that $\gamma_{\pi,t}$ starts falling relatively quickly and hence, the inflationary pressure from inflation expectations becomes muted. Therefore, inflation itself starts falling which decreases $\gamma_{\pi,t}$ further.

FIGURE E8. SMOOTH ATTENTION CHANGES



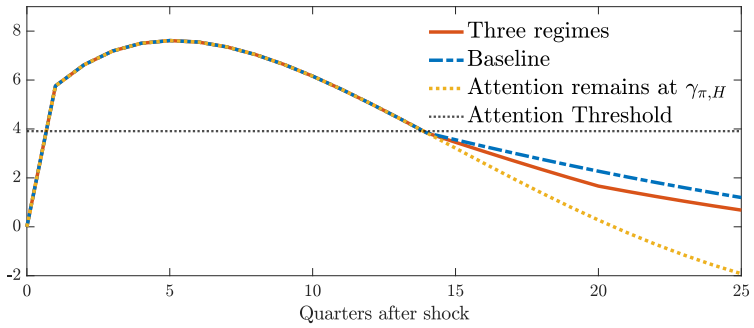
Note: Notes: The upper panel shows the inflation response in my baseline model (red-dashed line) and in a model in which attention changes smoothly with inflation (blue-dashed-dotted line). The lower panel plots the dynamics of attention $\gamma_{\pi,t}$ (blue-dashed-dotted line). In contrast, the black-dotted line in the lower panel plots the constant low level of attention.

THREE ATTENTION REGIMES.

In a recent paper, Korenok and Munro (2024) show that even though attention—measured using Google Trends—has decreased as inflation started to come down, it remained at a higher level than before the inflation surge. To understand how this possibility affects my model's predictions, I consider a third regime. In particular, I allow for a third regime in which attention takes on an intermedi-

ate value of 0.24.[18] The economy reaches that regime once inflation falls below the threshold. The economy is assumed to stay in that regime for six quarters before going back to the low-attention regime. The red-solid line in Figure E9 shows the inflation response to a cost-push shock in that case. For comparison, the blue-dashed-dotted line shows my baseline result with two regimes, and the yellow-dotted line shows the counterfactual scenario in which attention remains at its high level of $\gamma_{\pi,H}$ even after inflation has fallen back below the threshold. We can see that the slowdown in the speed of disinflation is somewhat muted, i.e., inflation comes down somewhat faster compared to my baseline model with two regimes. But nevertheless, there is still a pronounced slowdown in the inflation dynamics, indicating a longer last mile of disinflation compared to an economy absent a decrease in attention. With three regimes, there are two, not one, kinks in the disinflation phase. The first one occurs when inflation falls below the threshold and attention decreases from the high to the medium level. The second one happens in period 20, when attention decreases further to its low level. As long as there is some decrease in attention (which is consistent with the findings in Korenok and Munro (2024)), there is a slowdown in the speed of disinflation through the lens of my model.

Figure E9. Inflation dynamics with three attention regimes



Note: Notes: The blue-dashed-dotted line shows the (annualized) inflation response to a cost-push shock in my baseline model with two attention regimes, the red-solid line shows the inflation response for the model with three regimes, and the yellow-dotted line shows the model response if attention remains at its high level of $\gamma_{\pi,H}$ even after inflation has fallen back below the threshold.

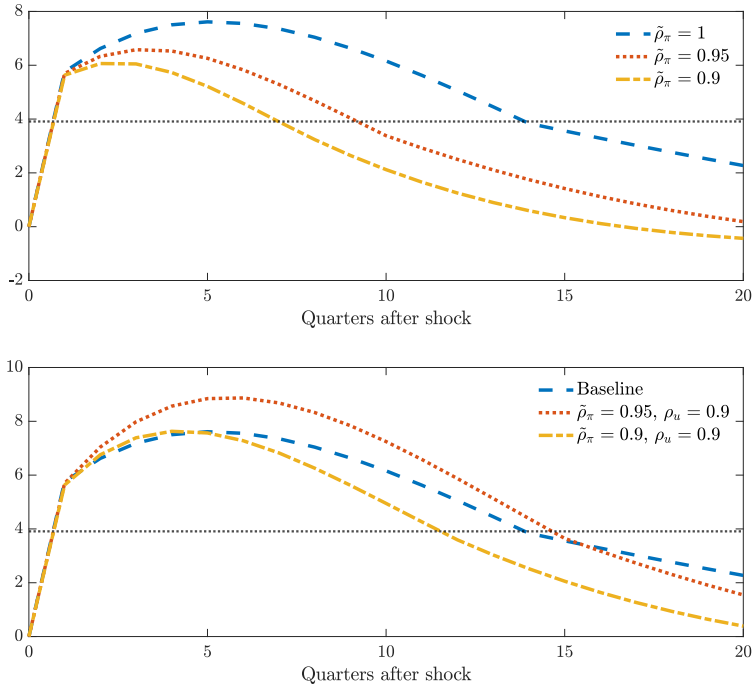
No random walk.

In the baseline calibration, I assume that the perceived persistence is set to $\tilde{\rho}_\pi = 1$. This random walk assumption ensures that inflation expectations and inflation coincide in the long run absent any shocks. Hence, I do not need to specify $\underline{\pi}$. I relax this assumption in the following.

[18] 0.24 is the average attention parameter I estimate when I do not allow for multiple regimes.

Figure E10 shows the inflation response to a cost-push shock for different calibrations. In all of them, I set $\underline{\pi} = 0$. This is a conservative assumption as imposing a positive g$\underline{\pi}$ (as observed in the data) would lead to an upward bias in inflation expectations and hence, *ceteris paribus*, to more inflationary pressure. The blue-dashed line shows the baseline case with $\tilde{\rho}_\pi = 1$. The red-dotted line shows the case with $\tilde{\rho}_\pi = 0.95$ and the yellow-dashed-dotted line shows the case $\tilde{\rho}_\pi = 0.9$. Qualitatively, the three are all quite similar. Not surprisingly, however, the inflation surge is less persistent as the model produces less endogenous persistence when $\tilde{\rho}_\pi$ is lower. I therefore consider in the lower panel of Figure E10 the case in which the exogenous shock persistence ρ_u is increased from 0.8 to 0.9 and compare the results to my baseline calibration. We see that in this case, the three inflation responses are also quantitatively very similar.

FIGURE E10. INFLATION DYNAMICS WITH $\tilde{\rho}_\pi < 1$



Note: Notes: The blue-dashed line shows the (annualized) inflation response to a cost-push shock in the baseline model with $\tilde{\rho}_\pi = 1$ and $\rho_u = 0.8$. The red-dotted line shows the model with $\tilde{\rho}_\pi = 0.95$, and the yellow-dashed-dotted line shows the model with $\tilde{\rho}_\pi = 0.9$.

TIME-VARYING SLOPE OF THE PHILLIPS CURVE.

Figure E11 shows the model predictions if attention remains constant at its low level $\gamma_{\pi,L}$, but in which the price adjustment cost decreases, reflected in a

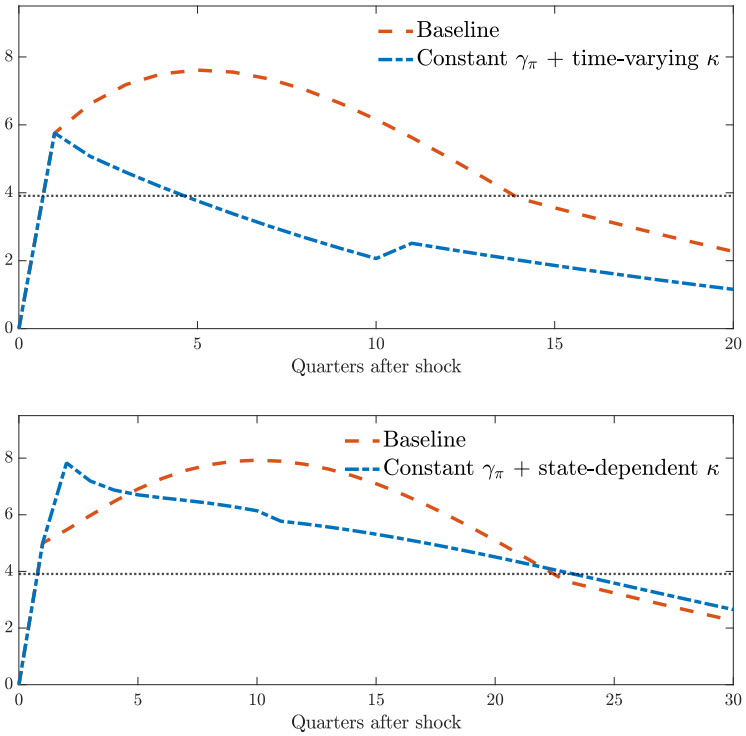
temporarily higher κ. Decreasing the adjustment-cost parameter is a reduced-form way to capture the possibility of greater price flexibility. The blue-dashed-dotted line in the upper panel shows the inflation response to a cost-push shock assuming that κ doubles for 10 quarters after the shock. The lower panel shows the response to a demand shock. In both figures, the red-dashed line shows the responses in my baseline model with regime-dependent attention but constant κ.

The model with fixed attention and time-varying κ cannot explain the hump-shaped inflation response. In the case of the demand shock, disinflation happens very gradually. Thus, that model may explain the long last mile of disinflation, but it cannot simultaneously explain why disinflation happened relatively fast initially after inflation peaked. Nevertheless, the empirical evidence shows that the frequency of price adjustments does fluctuate over time (see, e.g., Blanco et al., 2024).

Figure E12 further provides the inflation dynamics after supply shocks for both small and large shocks for the two model specifications. The differences between the baseline model and the time-varying-κ alternative are smaller for the small shock, since inflation does not cross the attention threshold and hence attention remains fixed.

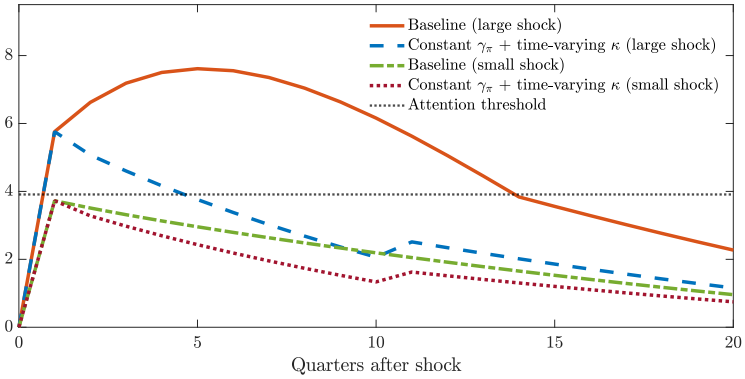
A full analysis of endogenous price adjustment frequency and the attention threshold is left for future work.

FIGURE E11. INFLATION DYNAMICS WITH STATE-DEPENDENT κ



Note: Notes: The red-dashed line in the upper panel shows the (annualized) inflation response to a cost-push shock in the baseline model with constant slope of the Phillips Curve κ but two attention regimes. The blue-dashed-dotted line shows the inflation dynamics in response to a cost-push shock in the model with constant attention but in which κ doubles between periods 1 and 10. The lower panel shows the two model responses to a demand shock.

FIGURE E12. INFLATION DYNAMICS WITH STATE-DEPENDENT κ



Note: Notes: The red-solid line shows the baseline response to a large shock in the model with the attention threshold, and the blue-dashed line shows the response to a large shock in the model with constant attention but in which κ doubles between periods 1 and 10. The green-dashed-dotted line shows the response of the attention-threshold model to a small shock, and the red-dotted line the model response to a small shock when κ doubles between periods 1 and 10.

*

REFERENCES

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. 2022. "Subjective models of the macroeconomy: Evidence from experts and representative samples." *The Review of Economic Studies*, 89(6): 2958–2991.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas. 2020. "Business-cycle anatomy." *American Economic Review*, 110(10): 3030–3070.

Aruoba, Borağan. 2020. "Term structures of inflation expectations and real interest rates." *Journal of Business & Economic Statistics*, 38(3): 542–553.

Aruoba, S Borağan, and Thomas Drechsel. 2024. "Identifying monetary policy shocks: A natural language approach." *Working paper*.

Barnichon, Regis, and Geert Mesters. 2020. "Identifying modern macro equations with old shocks." *The Quarterly Journal of Economics*, 135(4): 2255–2298.

Barnichon, Regis, and Geert Mesters. 2021. "The phillips multiplier." *Journal of Monetary Economics*, 117: 689–705.

Bhandari, Anmol, Jaroslav Borovička, and Paul Ho. 2025. "Survey data and subjective beliefs in business cycle models." *Review of Economic Studies*, 92(3): 1375–1437.

Blanco, Andrés, Corina Boar, Callum J Jones, and Virgiliu Midrigan. 2024. "The Inflation Accelerator." *Working paper*.

Bracha, Anat, and Jenny Tang. 2025. "Inflation levels and (in) attention." *Review of Economic Studies*, 92(3): 1564–1594.

Canova, Fabio. 2007. "G-7 inflation forecasts: Random walk, Phillips curve or what else?" *Macroeconomic Dynamics*, 11(1): 1.

Carlson, John A, and Michael Parkin. 1975. "Inflation expectations." *Economica*, 42(166): 123–138.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy*, 120(1): 116–159.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation." *American Economic Journal: Macroeconomics*, 7(1): 197–232.

Coibion, Olivier, and Yuriy Gorodnichenko. 2025. "Inflation, Expectations and Monetary Policy: What Have We Learned and to What End?"

Faust, Jon,, and Jonathan H Wright. 2013. "Forecasting inflation." In *Handbook of economic forecasting*. Vol. 2, 2–56. Elsevier.

Fisher, Jonas DM, Leonardo Melosi, and Sebastian Rast. 2025. "Long-run inflation expectations." *Working Paper*.

Fulton, Chad,, and Kirstin Hubrich. 2021. "Forecasting US inflation in real time." *Econometrics*, 9(4): 36.

Gemmi, Luca, and Rosen Valchev. 2023. "Biased surveys."

Hamilton, James D. 1994. *Time series analysis.* Princeton university press.

Jarociński, Marek, and Peter Karadi. 2020. "Deconstructing monetary policy surprises—the role of information shocks." *American Economic Journal: Macroeconomics*, 12(2): 1–43.

Kahneman, Daniel. 2011. *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kamdar, Rupal, and Walker Ray. 2024. "Attention-Driven Sentiment and the Business Cycle." *Working Paper*.

Känzig, Diego R. 2021. "The macroeconomic effects of oil supply news: Evidence from OPEC announcements." *American Economic Review*, 111(4): 1092–1125.

Korenok, Oleg, and David Munro. 2024. "The Rockets and Feathers of Inflation Attention." *Working Paper*.

Macaulay, Alistair. 2022. "Shock transmission and the sources of heterogeneous expectations." *Working Paper*.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. "Rational inattention: A review." *Journal of Economic Literature*, 61(1): 226–273.

Matějka, Filip, and Alisdair McKay. 2015. "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105(1): 272–98.

Newey, Whitney K, and Kenneth D West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation." *Econometrica*, 55(3): 703–708.

Pfäuti, Oliver. 2025. "Inflation–who cares? Monetary Policy in Times of Low Attention." *Journal of Money, Credit and Banking*, 57, Issue 5.

Pfäuti, Oliver, and Fabian Seyrich. 2023. "A behavioral heterogeneous agent New Keynesian model." *Working paper*.

Reis, Ricardo. 2022*a*. *The burst of high inflation in 2021-22: How and why did we get here?* Bank for International Settlements, Monetary and Economic Department.

Reis, Ricardo. 2022*b*. "Losing the inflation anchor." *Brookings Papers on Economic Activity*, 2021(2): 307–379.

Vellekoop, Nathanael, and Mirko Wiederholt. 2019. "Inflation expectations and choices of households." *Working Paper*.

Weber, Michael, Bernardo Candia, Hassan Afrouzi, Tiziano Ropele, Rodrigo Lluberas, Serafin Frache, Brent Meyer, Saten Kumar, Yuriy Gorodnichenko, Dimitris Georgarakos, et al. 2025. "Tell Me Something I Don't Already Know: Learning in Low-and High-Inflation Settings." *Econometrica*, 93(1): 229–264.