TI 2024-083/VI
Tinbergen Institute Discussion Paper

Soft Landing and Inflation Scares

Jim Bullard[1]
Alex Grimaud[2]
Isabelle Salle[3]
Gauthier Vermandel[4]

1 Purdue University

2 Bank of Austria, TU Wien

3 University of Amsterdam, University of Ottawa, Tinbergen Institute

4 Ecole Polytechnique Paris, Bank of France

Soft Landing and Inflation Scares *

Jim Bullard[†] Alex Grimaud[‡] Isabelle Salle[§] Gauthier Vermandel[¶]

January 7, 2025

Abstract

We discuss the timing and strength of the Fed's reaction to the recent inflation surge within an estimated macroeconomic model where long-run inflation expectations are heterogeneous and can lose their anchoring to the target. The resulting inflation scare worsens the real cost of disinflation. We derive a closed-form solution that retains the entire time-varying cross-sectional distribution of subjective inflation beliefs. We estimate the model using Bayesian techniques on both US macroeconomic time series and forecast data from the Survey of Professional Forecasters. Counterfactual simulations show that the timing – rather than the strength – of the policy reaction to the inflation surge is critical to contain the development of an inflation scare and prevent the entrenchment of above-target inflation. We show that the Fed fell behind the curve in 2021 since an earlier tightening could have reduced the inflation peak without triggering a recession. However, further delays would have unanchored inflation expectations, aggravated the inflation scare and strengthened the inflation surge, resulting in larger output losses.

1 Introduction

While the Volcker's disinflation over 1979-1987 came with substantial output losses, the recent inflation surge was brought down without any major recession (see Fig. 1 for illustration). What made this soft landing possible? This paper discusses this question with counter-factual monetary policy simulations from a novel estimated framework where long-run inflation expectations can lose their anchoring to the target.

In the earlier rational expectations (RE) literature, credible disinflations i.e., under which agents' expectations align with the policy objective, are not detrimental to output.[1] Hence, the costs of disinflation in the 1980s have been largely attributed to the

*Disclaimer: The views expressed are those of the authors and do not necessarily reflect those of the Oesterreichische Nationalbank (OeNB), the Banque de France (BdF), or the Eurosystem. Funding from the Canadian Research Chair program and the OeNB grant No. 18611 is gratefully acknowledged.

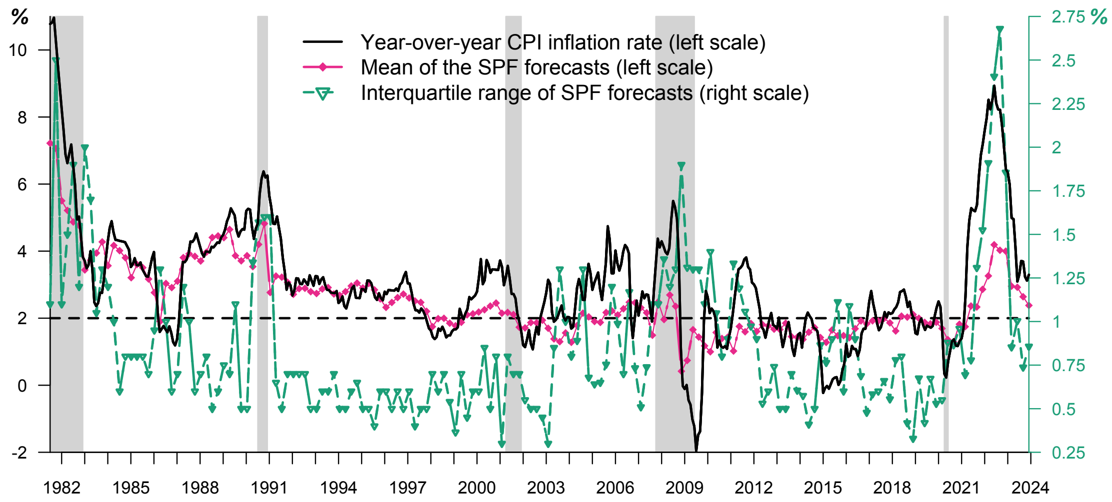
†Dr. Samuel R. Allen Dean of the Daniels School of Business, Purdue University, IN, USA.

‡Monetary Policy Section, Oesterreichische Nationalbank, Vienna, AT and Institute of Statistics and Mathematical Methods in Economics, TU Wien, Vienna, AT.

§Department of Economics, University of Ottawa, Ottawa, ON, CA; Amsterdam School of Economics & University of Amsterdam & Tinbergen Institute, Amsterdam, NL.

¶CMAP, École polytechnique, Palaiseau, FR; Universités PSL & Paris-Dauphine, Paris, FR and Banque de France, Paris, FR.

[1] This view is evident in the New Classical models (Kydland & Prescott 1977, Lucas & Sargent 1981) and the New Keynesian (NK) synthesis (Clarida et al. 1999). See Sargent (2022) for a recent discussion of the 1980s and RE models.



Notes: Authors' computations. Latest observation as of 2023:12, first data point in 1981:Q3 since individual SPF forecasts are unavailable earlier. Data are in percentage points (p.p.). The dashed line represents the Fed's 2% inflation target. The shaded areas depict the NBER recessions. The time series of CPI inflation (left scale) is from the FRED database (CPIAUCSL), the mean SPF inflation forecasts (left scale) and their interquartile range (right scale) are, respectively, CPI3 and CPI_D1(T+1) from the SPF data.

Figure 1: Inflation, inflation expectations and business cycles in the US

public's disbelief in the Fed's commitment to reduce inflation (Goodfriend & King 2005, King & Lu 2022):[2] coupled with past experiences with high inflation, this lack of credibility led to unanchored (long-run) expectations. These 'inflation scares' – in the terminology of Goodfriend (1993) – called for a costly tightening to prevent high inflation from becoming entrenched. Does the recent soft landing imply that the Fed avoided an inflation scare? On the other hand, the inflationary pressures stemming from the pandemic's unique disruptions within the context of the well-established institutional framework of inflation targeting may have been transitory.[3] In other words, would the outcome have been different under alternative policy actions?

Discussing these questions within full-information RE (FIRE) models is challenging because policies are assumed to be credible and expectations mean-reverting or, in other words, anchored to the central bank (CB) target in the long run. These elements leave out the possibility of inflation scares developing in the wake of an inflation surge.[4] In such a framework, Walsh (2022) shows that delaying tightening along the recent inflation surge does not make a sizeable difference in terms of subsequent inflation and output. Using a model where long-run inflation expectations are policy-invariant, Reifschneider (2024) concludes that earlier tightening from the Fed would not have tamed the inflation surge, but would have resulted in substantial costs for the real economy. Hakamada & Walsh (2024) further show that a moderate delay in the policy reaction to

[2]The model of King & Lu (2022), in which private agents learn about the policy maker's ability to commit to low inflation, is a salient example of this view. Within this context, the authors emphasize the lack of credible commitment of the Fed in the 1970s, and interpret the ensuing costly disinflation in the 1980s as an episode of reputation building.

[3]For instance, using a structural vector autoregression (SVAR) model with added exogenous variables, Bernanke & Blanchard (forthcoming) estimate that most of the recent inflation dynamics can be attributed to price shocks (food and, in particular, energy prices), which tend to have transitory effects on inflation. Gagliardone & Gertler (2023) also emphasize oil price and money shocks.

[4]This is true for a well-specified FIRE model, with specific restrictions on parameter values to ensure path determinacy, while exogenous shocks, albeit persistent, are assumed to be transitory.

the recent inflation surge could be compensated by a stronger reaction. Their conclusion holds even when deviating from the FIRE benchmark and instead considering cognitive discounting, under which the expectation channel is weakened but expectations still mean-revert.

FIRE models also overlook the heterogeneity in expectations, even though it is a pervasive feature of real-world forecasts that tends to amplify amid inflation surges (see Fig. 1).[5] Particularly during the post-pandemic era, inflation expectations became saliently more dispersed, while their average rose only modestly. This pattern suggests a distinct role of forecast dispersion in expectation dynamics, the ensuing evolution of CB credibility, and the effectiveness of monetary policy.

Against this background, we examine the recent soft landing using a model where inflation scares may arise as an outcome of the time-varying dispersion in inflation expectations. We develop and estimate a micro-founded heterogeneous-expectation New Keynesian (HENK) model where information about the long-run value of inflation is assumed to be dispersed and sticky. Agents hold idiosyncratic beliefs about inflation which they exchange during social interactions implemented by social learning (SL).[6] While nesting the FIRE benchmark, this information friction allows for persistent departures from it in the form of inflation scares. These scares arise whenever the CB repeatedly misses the target due to shocks and/or monetary policy actions. As a result, off-target inflation beliefs are validated and spread among agents, making them self-fulfilling. Importantly, in our model, the heterogeneity in expectations plays a distinct role in the transmission of monetary policy since a greater dispersion in beliefs amplifies the effects of the shocks on the economy, and complicates the stabilization task of the CB. This mechanism of expectation unanchoring provides the necessary ingredient to assess the real costs of disinflation and the challenge of keeping expectations coordinated at the target despite adverse shocks.[7] In other words, our framework strengthens the role of CBs as managers of expectations.

Our formulation of heterogeneous expectations offers a closed-form solution which can be estimated using full-information methods while preserving the entire time-varying cross-sectional distribution of beliefs. The resulting framework can then be used to perform counterfactual policy simulations with a micro-founded measure of welfare that accounts for the additional costs stemming from belief and price dispersion. We estimate the HENK model using US macroeconomic and survey data from the survey of professional forecasters (SPF).[8]

[5]See, in particular, Meeks & Monti (2023) on the quantitative relevance of expectations heterogeneity in accounting for inflation dynamics.

[6]Our model is inspired by the pioneered work of Arifovic (1995, 1996) and the related contributions of Arifovic et al. (2013, 2018, 2024), Jia & Wu (2023), but we implement substantial differences to retain microfoundations and obtain a closed-form solution to bring the model to the data.

[7]Adaptive learning is another common form of deviation from FIRE; see, for example, Evans & Honkapohja (2001), Slobodyan & Wouters (2012a), Gáti (2023) within representative-agent frameworks, and Evans & McGough (2024) for a comparison with SL. Branch & Evans (2011) consider a model where agents switch between two different adaptive learning rules and show how this expectation formation combined with adverse supply shocks can account for the dynamics of inflation in the 1970s and 1980s. Other related models of heterogeneous expectations include, *inter alia*, Andrade et al. (2019), Evans et al. (2024), Ozden (2024).

[8]Our paper builds upon the recent literature that estimates DSGE models under boundedly rational expectations using full information methods; see, *inter alia*, Slobodyan & Wouters (2012b), Hommes et al. (2023), Hajdini (2023). To the best of our knowledge, the most closely related works on heterogeneous expectation models estimated using Bayesian methods are Ozden (2024), where agents switch between two expectation formation mechanisms, and Branch & Gasteiger (2019), where agents can entertain Ricardian and non-Ricardian expectations. Our estimation exercise also uses expectations as observables in learning models; see Slobodyan & Wouters (2012a), Carvalho et al. (2023), Rychalovska et al. (2024) within NK models under homogeneous ex-

We first show how our model matches remarkably well the (untargeted) dynamics of the cross-sectional dispersion in inflation expectations over the business cycle, for instance as measured by the interquartile range of individual SPF forecasts. We further illustrate how this heterogeneity in expectations further impairs the transmission of monetary policy and amplifies the effects of shocks compared to homogeneous-agent frameworks, even those featuring boundedly rational expectations. We proceed by decomposing the macroeconomic variables as a function of the shocks introduced in the model. We show how the expectation dynamics contributes to explaining a substantial part of the missing inflation in the 2010 decade and the recent surge in inflation, along with an overall loose monetary policy stance and adverse cost-push and demand shocks. We then perform counterfactual policy simulations by varying the timing and strength of the policy reaction to the recent inflationary episode.

Our main policy message is that, in stark contrast with the FIRE exercises, the timing of the Fed's reaction to inflation developments is the main key to the management of inflation expectations and the ensuing inflation and output courses. On the contrary, varying the strength of the reaction – provide that it is not too weak – makes little difference.[9] The further the delay in the tightening cycle in 2021, the stronger the unanchoring movement in inflation expectations, and the more severe the resulting inflation scare and inflation surge. A further delay of about three quarters would have resulted in an expectation-driven entrenchment of high inflation.

Furthermore, we find that the Fed did fall behind the curve in 2021 and early 2022 since embarking on a preemptive tightening would have reduced the inflation peak without engineering a recession. While the recent inflation surge did create an inflation scare, it would have turned out more severe had the Fed further delayed the increase in interest rates because inflation expectations would have drifted further away from the target. This situation would have complicated the stabilization of inflation and increased the ensuing output losses. We further illustrate how even a single earlier cut in the interest rate would have aggravated this inflation scare and resulted in annual inflation landing close to 2% above steady state by the end of 2023.

The rest of the paper is organized as follows. Section 2 develops the model, Section 3 presents the resulting estimated framework, Section 4 discusses the policy exercises and Section 5 concludes.

2 A micro-founded HENK model

We introduce a stylized model of the economy where the only ingredient that differs from the RE model pertains to the way agents form their inflation expectations. Appendix B shows how the HENK model can be micro-founded with a set of standard ingredients[10] independently of the distribution of the agents' subjective beliefs, as long as

pectations and Cornea-Madeira et al. (2019) on a Phillips curve with two types of expectations. Among other things, we add to this literature the estimation of a general equilibrium model under heterogeneous expectations using empirical forecasts data, a focus on the cross-sectional dynamics in these expectations and its role in the transmission channel of monetary policy, and counter-factual policy simulations.

[9] Whether the stabilization trade-off requires a more or a less aggressive response in a non-FIRE world is unsettled. Earlier work based on adaptive learning concludes that a more aggressive rule is required than under RE (Orphanides & Williams 2004), which has been corroborated in lab experiments (Mauersberger 2021, Assenza et al. 2021). By contrast, under some form of rational inattention, Gabaix (2020) shows that the CB may stabilize the economy with smaller reaction parameter values than under RE.

[10] These elements include complete financial markets, a non-separable utility function on the households' side and a price-setting behavior *à la* Rotemberg together with a linear production technology where labor is the only input; see also Arifovic et al. (2024).

the population of heterogeneous expectations is discrete. We could straight-forwardly extend our framework, e.g. to include wage dynamics, since our Dynare toolbox – available in Grimaud et al. (forthcoming) – allows us to incorporate this class of beliefs into any dynamic model with expectations. Nevertheless, we use a baseline model that is sufficient to adequately capture the dominant role of price and money shocks over the period that we study, as emphasized in the literature (Gagliardone & Gertler 2023, Bernanke & Blanchard forthcoming).

2.1 A baseline NK model with subjective inflation expectations

Let us consider a textbook log-linear macroeconomic model in which inflation π and output gap x are determined by the following system of equations:

$$\hat{\pi}_t = \kappa y_t + \beta E_t^{SL}(\hat{\pi}_{t+1}) + u_t, \tag{1}$$

$$\hat{y}_t = E_t(\hat{y}_{t+1}) - \sigma^{-1}\left(\hat{\imath}_t - E_t^{SL}(\hat{\pi}_{t+1})\right) + g_t, \tag{2}$$

where $\hat{z}$ refers to the deviation of variable z from its steady-state value. Equation (1) is an aggregate supply relationship otherwise named NK Phillips curve (NKPC) with slope $\kappa > 0$, $\beta \in (0,1)$ is the utility discount factor of the households and $E_t^{SL}(\hat{\pi}_{t+1})$ refers to the *subjective* aggregate expectation of inflation in period $t+1$ subject to SL, which needs not coincide with RE. In other words, the aggregate subjective inflation beliefs need not coincide with the true conditional expectations of inflation according to the probability distribution implied by the model, that we denote by the usual operator $E_t(\cdot)$. Variable u_t is a cost-push shock.

Equation (2) describes aggregate demand in terms of output gap, denoted by $\hat{y}$; $\sigma > 0$ is the inverse of the intertemporal elasticity of substitution of consumption; $\hat{\imath}_t$ is the nominal interest rate set by the CB; g_t is a real shock and we assume rational expectation of the output gap in period $t+1$. While it would be straight-forward to include lagged inflation in Eq. (1) and lagged output gap in Eq. (2) perhaps as a result of consumption habits and price indexation,[11] we refrain from doing so here because such a less parsimonious version does not improve the quality of the estimation in Section 3.

The shock processes g and u are to be specified below and we add an interest-rate feedback rule of the following form:

$$\hat{\imath}_t = \rho_i \hat{\imath}_{t-1} + (1-\rho_i)\left(\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t\right) + v_t, \tag{3}$$

where $\phi_\pi, \phi_y > 0$ reflect a concern for inflation and output gap stabilization under a flexible inflation-targeting regime, $\rho_\imath \in (0,1)$ models interest-rate smoothing and v_t is a monetary policy shock.

Under RE, $E_t^{SL}(\hat{\pi}_{t+1}) = E_t(\hat{\pi}_{t+1})$, and the model is entirely summarized by Eqs.(1), (2) and (3) together with some shock processes and it is straightforward to obtain a closed-form solution using standard methods. The minimum state-variable (MSV) solution reads as follows:

$$\hat{z}_t = P\hat{z}_{t-1} + Q\epsilon_t, \tag{4}$$

where $\hat{z}_t \equiv \left(\hat{\pi}_t, y_t, \hat{\imath}_t\right)'$ gathers the endogenous variables, $\epsilon_t \equiv \left(u_t, g_t, v_t\right)'$ collects the exogenous shocks, matrices P and Q depend on the parameters of the model and the intercept is always zero because variables are expressed in deviation from their steady-state values.

[11] Grimaud et al. (forthcoming) show how subjective and dispersed expectations that evolve under SL can be introduced in any DGSE model and provide a Dynare toolbox to solve the resulting HENK models, which we use in the present paper.

In the HENK model, $E_t^{SL}(\hat{\pi}_{t+1}) \neq E_t(\hat{\pi}_{t+1})$ and to close the model, one needs to specify how subjective inflation expectations are formed. This information friction is the only departure from the otherwise standard RE textbook model.

2.2 Dispersed information and subjective inflation beliefs

We use a model of dispersed information driven by social interactions to introduce information friction regarding the long-run value of inflation. Therefore, there is disagreement among heterogeneous agents who differ from each other only regarding their subjective beliefs about long-run inflation. We focus on heterogeneous inflation expectations because households have admittedly more information about their own future income, as incorporated in the output gap expectations in the Euler equation (2), than about aggregate future inflation (see Walsh (2022) for a similar argument). More importantly, our focus is on the possibility of long-lasting drifts in inflation expectations away from the target.

Consider a discrete population of agents, indexed by $j \in \{1, 2, ..., J\}$, which differ only by their subjective inflation expectation, denoted by $E_{j,t}^{SL}(\hat{\pi}_{t+1})$. Without loss of generality and to facilitate the implementation of social interactions (see below), J is an even number. Appendix B shows how the present model can be micro-founded when individual expectations are aggregated into the set of equations (1)-(2) using the arithmetic mean, i.e. $E_t^{SL}\hat{\pi}_{t+1} \equiv \frac{1}{J}\sum_{j=1}^{J} E_{j,t}^{SL}(\hat{\pi}_{t+1})$ (see also Arifovic et al. 2024).

Before making explicit the evolution of these beliefs in the HENK model, we introduce two assumptions, namely steady-state learning and internal rationality. Under steady-state learning, agents' beliefs concern only the low frequency component of inflation, i.e. the intercept in the MSV solution of the economy; see, e.g., Carvalho et al. (2023) for a similar assumption. The rest of their expectations, that is the effects of the transitory shocks and the lagged endogenous variables on inflation, coincides with RE.

As to internal rationality, it implies that agents are aware that this deviation from RE impacts the law of motion of the economy and they internalize its effects when forming their forecasts; see Adam & Marcet (2011) for a similar approach. Mathematically, the deviation from RE introduces an additional state variable, which could equivalently be conceived as an additional shock, into the model.[12] This variable, which we denote by φ, summarizes the learning dynamics, i.e. the distance between the dispersed subjective inflation beliefs and their rational counterpart. We make φ explicit below but it is not necessary at this stage. The nested case of $\varphi_t = 0, \forall t$ corresponds to the model without SL beliefs and boils down to the RE model in Section 2.1. In the HENK model, the MSV solution (4) is therefore augmented as follows to account for the presence of subjective beliefs:

$$\hat{z}_t = P\hat{z}_{t-1} + \tilde{Q}\mathscr{E}_t, \tag{5}$$

where $\tilde{Q} \equiv (Q\ R)$ is now a 3×4 matrix whose first three columns correspond to Q in Eq. (4), R is a 3×1 vector that also depends on the structural parameters of the model, and $\mathscr{E}_t \equiv \left(u_t, g_t, v_t, \varphi_t\right)' = \left(\epsilon_t, \varphi_t\right)'$.

Under the aforementioned assumptions, agents use forecasting models, or perceived laws of motion (PLM), that have the same form as the MSV solution (5). In other words, their PLM is well-specified and coincide with the MSV solution of inflation dynamics, up

[12]Note that internal rationality implies that the rational output gap expectations that we assume in the model correspond to the true conditional expectations according to the probability distribution implied by the HENK model, i.e. a rational forecaster takes into account the effect of non-RE inflation beliefs on the output gap.

to a constant. Agent j uses the following forecasting model:

$$\hat{\pi}_t^{(j)} = a_{j,t} + P_{1,\bullet}\hat{z}_{t-1} + \tilde{Q}_{1,\bullet}\mathscr{E}_t, \tag{6}$$

where $a_{j,t}$ is their subjective belief about steady-state inflation, and matrices $P_{1,\bullet}$ and $\tilde{Q}_{1,\bullet}$ refer to the first row of matrices P and $\tilde{Q}$ as defined in Eq. (5). Since inflation is expressed in deviation from the target, any non-zero intercept $a_{j,t}$ in Eq. (6) corresponds to a subjective misperception of steady-state inflation. An individual belief $a_{j,t}$ may be interpreted as the relative anchoring of agent j's inflation expectation with respect to the CB's inflation target, where $a_{j,t} > 0$ denotes an upward-biased view and $a_{j,t} < 0$ a downward-biased perception of the target and $a_{j,t} = 0$ indicates a perfect anchoring at the target. Agents hold dispersed beliefs $\{a_{j,t}\}_{j\in J}$, which need not coincide with the inflation target (i.e., with zero) and agents constantly revise these beliefs (see Section 2.3).

Once a stochastic process for the exogenous shocks g, u and v is assumed, the only component that remains to be specified – besides the evolution of the idiosyncratic beliefs $\{a_{j,t}\}$ – is the state variable φ_t. Note that agents do not need to know how exactly φ is aggregated from the individual beliefs $\{a_{j,t}\}$ to use their PLM (6) but they need to form a forecast for φ_{t+1} to forecast inflation in $t+1$. We make two information assumptions to obtain a closed-form solution of the model and be able to use standard full-information estimation methods to bring our model to the data. The first assumption establishes that individual beliefs are dispersed and, hence, private:

Assumption 1 (Dispersed idiosyncratic beliefs) *Beliefs are private, agents cannot observe the beliefs of the others beyond the one of their mate during one-to-one social interactions: agent k does not observe any other beliefs $a_{j,t}$, $j \neq k, \ell$ in the population, where an agent ℓ is randomly selected without replacement in the entire population of J agents to be the matched mate of agent k in a given period (see Section 2.3 for details).*

Under Assumption 1, agents in the HENK model use their private beliefs to proxy for the aggregate deviation of inflation expectations from RE in $t+1$, i.e. $E^{SL}_{j,t}(\varphi_{t+1}) = a_{j,t}$. Together with explicit shock processes and a definition of the information set in time t, it is then straightforward to iterate forward the forecasting model (6) and take the mathematical expectation to obtain the idiosyncratic inflation expectation of agent j:

$$E^{SL}_{j,t}(\hat{\pi}_{t+1}) = a_{j,t} + P_{1,\bullet}E_t(\hat{z}_t) + \tilde{Q}_{1,\bullet}(E_t(\varepsilon_{t+1})\ a_{j,t})'. \tag{7}$$

Contemporaneous variables $\hat{z}_t$ may or may not be observable depending on the assumption pertaining to the information set in t.[13] We also assume that agents use their latest available belief $a_{j,t}$ in their forecasting model.

The second assumption is a consequence of Assumption 1 and pertains to the information set under RE. By RE, we have in mind an observer who forecasts the endogenous variables given their information set in line with their true conditional expectations according to the probability distribution implied by the HENK model. If individual beliefs were observable, an RE observer could predict future aggregate beliefs by anticipating the news shocks and the selected beliefs out of all possible pairwise comparisons in the tournament. However, since individual beliefs are private, this computation is unavailable. Instead, the RE observer could estimate the aggregate outcomes of the learning

[13] Grimaud et al. (forthcoming) show how the model can be straight-forwardly aggregated and solved while keeping the consistency between the subjective beliefs $\{a_{j,t}\}$ and the aggregate deviation from RE φ_t under both information assumptions. Ozden (2024) further discusses timing assumptions when estimating NK models under learning.

process detailed in Section 2.3 hereafter by using a forecasting model involving the state variables of the model. However, this would add approximations and learning from the RE observer, on top of the information frictions in the agents' beliefs, which would make the framework substantially more complex and jeopardize the consistency between the agents' individual expectations and the macroeconomic view of the observer. We instead define a quasi-RE observer and make the following assumption:

Assumption 2 (Random-walk beliefs) *In the absence of information about the formation of individual beliefs, in particular the functioning of the social interactions among dispersed agents, a quasi-RE observer of the SL economy treats the aggregate beliefs φ as a random-walk process in expectations (i.e., a martingale).*

One can think of Assumption 2 as depicting a CB surveying household or firm expectations and unsure about the precise way they form their expectations. The CB (or the macro observer) instead incorporates the latest observable information about their beliefs to predict the next period. Under Assumptions 1-2, we may write $E_t(\varphi_{t+1}) = \varphi_t$.[14] The aggregate inflation expectation, i.e. the average of the distribution of subjective beliefs given by (7), reads as:

$$\begin{aligned} E_t^{SL}(\hat{\pi}_{t+1}) \equiv \frac{1}{J}\sum_{j\in J} E_{j,t}^{SL}(\hat{\pi}_{t+1}) =& \varphi_t + P_{1,\bullet}E_t(\hat{z}_t) + \tilde{Q}_{1,\bullet}(E_t(g_{t+1}), E_t(u_{t+1}), E_t(v_{t+1}), \varphi_t)' \\ =& \varphi_t + P_{1,\bullet}E_t(\hat{z}_t) + \tilde{Q}_{1,\bullet}E_t\mathscr{E}_{t+1} \\ =& \varphi_t + E_t(\hat{\pi}_{t+1}). \end{aligned} \tag{8}$$

where we explicitly define $\varphi_t \equiv \sum_{j\in J} a_{j,t}$ as the average subjective belief about steady-state inflation in the economy, consistently with the definition of φ in Eq. (5) as the gap between aggregate subjective beliefs and their RE counterparts. The last equality in Eq. (8) stems from the forward iteration of Eq. (5) which gives the rational inflation forecast in the HENK model. Note that Assumption 1 is not essential: given our aggregation procedure, the same aggregate expectation (8) would result if SL agents were to forecast φ_t in the same way as the quasi-RE observer, i.e. using Assumption 2 – for instance if the average belief in the economy were to be released at the end of each period, perhaps through the publication of survey outcomes. In such case, only the individual expectations and, hence, the cross-section would be affected by the use of Assumption 1.

2.3 Evolution of inflation beliefs under SL

To conclude the model presentation, it remains to be determined how the cross-section distribution of beliefs $\{a_{j,t}\}$ evolves between time $t-1$ and t. Intuitively, beliefs evolve under SL, which involves decentralized interactions during which agents: after receiving news, they are grouped by pairs and compare their long-run inflation beliefs. The belief of the agent whose forecasting model produces the lowest forecast errors over the recent inflation history is adopted by the other member of the pair.

Formally, SL is a recursive non-linear process that we summarize at the agent-level by a function $s(\cdot)$ as follows:

$$a_{j,t} - a_{j,t-1} = s\left(\lambda_t, \iota_{j,t}, \hat{\pi}_{t-1}, \ldots, \hat{\pi}_1\right). \tag{9}$$

[14] We use the $E_t(\cdot)$ operator although the resulting expectations are model-consistent only when the *aggregate* belief φ, which is explicitly defined in Eq. (8) hereafter as the average belief over the J agents, is zero. It is the case at the REE, where $a_{j,t} = 0$, $\forall i$, but it is sufficient that the *average* belief across all agents is zero. In general, the derived closed-form solution represents a quasi-REE, and the $E_t(\cdot)$ operator a quasi-RE, since all expectations are model-consistent, except the expected aggregate outcome of social interactions after news are received, i.e. S_{t+1} (see Section 2.3 for details).

Let us unpack the process $s(\cdot)$. Information shocks λ_t and $\iota_{j,t}$ are independently and normally distributed (their variances, denoted respectively by σ^2_λ and σ^2_ι, are estimated in Section 3). λ_t is an aggregate information shock which affects all individual beliefs in the cross-section distribution to a same extent and could be interpreted as a common trend in the inflation expectations or the "mood of the market." Shocks $\iota_{j,t}$ are the idiosyncratic news shocks, drawn for each agent j independently. The presence of the whole history of past inflation rates in $s(\cdot)$ reflects the selection mechanism of the news during the social interactions: agents are more likely to adopt the forecasting models (6) that yield relatively better past forecasting performances, and discard those giving relatively higher forecast errors.

Precisely, SL follows a three-step process and encompasses two operators inspired by evolutionary learning and the use of genetic algorithms (GAs) for optimization purposes. First, at the beginning of period t, the economy-wide and individual information shocks are realized. Each agent j receives both an idiosyncratic and a common news which modify their belief with a probability μ. This probability is akin to a "Calvo" probability where information, rather than prices, is sticky (Mankiw & Reis 2002). Formally, individual beliefs are modified as:

$$m_{j,t} = a_{j,t-1} + \mathbb{1}_{\varpi_{j,t}\leq\mu}(\iota_{j,t} + \lambda_t) \tag{10}$$

where $\varpi_{j,t} \sim \mathcal{U}(0,1)$, so that $\mathbb{1}_{\varpi_{j,t}\leq\mu}$ is a dummy variable which equals one if agent j in period t has received the news (which occurs with probability μ), and zero otherwise.[15]

Second, after possibly updating the intercept in their forecasting model (6) with its new value given in Eq. (10), each agent j evaluates their associated forecasting performances over the whole history of inflation, but by putting more weights on recent forecast errors relative to the ones from the more distant past:

$$F_{j,t} = -\sum_{\tau=0}^{t-1} \rho^\tau (\hat{\pi}_{t-\tau} - E^{SL}_{j,t-\tau-1}(\hat{\pi}_{t-\tau}|m_{j,t}))^2 \tag{11}$$

where $E^{SL}_{j,t-\tau-1}\left(\hat{\pi}_{t-\tau}|m_{j,t}\right)$ is the forecast of inflation in period $t-\tau$ formed with the information set in $t-\tau-1$ and the forecasting model (6) where the intercept $a_{j,t}$ is $m_{j,t}$. Lower values of $\rho \in (0,1)$ imply a faster discounting of past errors, and higher values of $F_{j,t}$ indicate greater forecasting accuracy (or, equivalently, lower forecast errors).

Third, social interactions take place and are modeled *via* a tournament process during which all agents are paired, which results in $J/2$ couples. The best forecasting model within the couple (i.e. the one with the highest accuracy (11)) is adopted by the two members, and the least accurate one is therefore discarded. Formally, if agents k and ℓ are paired, we have:

$$(a_{k,t}, a_{\ell,t}) = \mathbb{1}_{F_{k,t}>F_{\ell,t}}(m_{k,t}, m_{k,t}) + (1 - \mathbb{1}_{F_{k,t}>F_{\ell,t}})(m_{\ell,t}, m_{\ell,t}). \tag{12}$$

where $\mathbb{1}_{F_{k,t}>F_{\ell,t}}$ is a dummy variable equal to one if the forecasting model of agent k utilizing $a_{k,t}$ has smaller forecast errors than the model of agent ℓ utilizing $a_{\ell,t}$ so that agent ℓ copies the forecasting model of agent k, and zero otherwise (when agent k then copies the forecasting model of agent ℓ).

Once these three steps are completed, the distribution of subjective beliefs is updated. To collect the model's equations and present the estimation method in Appendix

[15]The notation m and F hereafter are inherited from the *m*utation process and *f*itness measurement in the GA language.

A.1, it is convenient to express the evolution of beliefs at the aggregate level by summing Eq. (9) as follows:

$$\varphi_t = \varphi_{t-1} + S\left(\lambda_t, \{\iota_{j,t}\}, \hat{\pi}_{t-1}, \ldots, \hat{\pi}_1\right) \equiv \varphi_{t-1} + S(\cdot), \tag{13}$$

where $S(\cdot)$ is a non-linear function without an analytical expression that describes how the SL algorithm modifies the average subjective belief about steady-state inflation in the economy as a function of the aggregate and idiosyncratic shocks and the history of inflation.

2.4 Summary of the model

Collecting all equations gives the summary of the model with disperse and subjective inflation expectations:

$$\begin{aligned}
&\hat{\pi}_t = \kappa \hat{y}_t + \beta E_t^{SL}(\hat{\pi}_{t+1}) + u_t,\\
&\hat{y}_t = E_t(\hat{y}_{t+1}) - \sigma^{-1}\left(\hat{\imath}_t - E_t^{SL}(\hat{\pi}_{t+1})\right) + g_t,\\
&\hat{\imath}_t = \rho_i \hat{\imath}_{t-1} + (1-\rho_i)\left(\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t\right) + v_t,\\
&E_t^{SL}\hat{\pi}_{t+1} = \varphi_t + E_t\hat{\pi}_{t+1},\\
&\varphi_t = \varphi_{t-1} + S(\cdot),\\
&\varphi_t = E_t(\varphi_{t+1}),\\
&\varphi_t = \sum_{j\in J} a_{j,t},\\
&g_t = \rho_g g_{t-1} + \varepsilon_t^g - \mu_g \varepsilon_{t-1}^g,\\
&u_t = \rho_u u_{t-1} + \varepsilon_t^u - \mu_u \varepsilon_{t-1}^u,\\
&v_t = \varepsilon_t^v,
\end{aligned}$$

where the fourth to the seventh equations describe the subjective steady-state beliefs. The seventh equation is relevant to define the time-varying distribution of the idiosyncratic subjective beliefs. The last three equations of the model specify the processes of the fundamental shocks for the purpose of the estimation (ε_t^g, ε_t^u and ε_t^v are i.i.d. zero-mean Gaussian processes).

As usual, this model can be stacked into a matrix system as follows:

$$\mathbb{E}_t\left(F\hat{z}_{t+1} + G\hat{z}_t + H\hat{z}_{t-1} + M\mathscr{E}_t\right) = 0, \tag{14}$$

Identifying the terms with the MSV solution (5) solves for the policy function and pins down matrices P and $\tilde{Q}$.[16] Note that we only solve for well-specified models by considering the stable root P of the problem, which implies that the RE counterpart of the HENK model is always determinate.

3 The estimated model

The model is estimated using a Bayesian approach applied to a non-linear environment thanks to the inversion filter (Cuba-Borda et al. 2019), the technical details of which are deferred to Appendix A.1. In this section, we describe the data used, the estimated parameter values, the shocks identified, and the heterogeneity in expectations generated by the estimated model.

[16] To do so, we use the toolbox presented in Grimaud et al. (forthcoming) developed for Dynare (Adjemian et al. 2024).

3.1 Data description

Our sample comprises four macroeconomic time series that span the periods from 1985Q1 to 2023Q4 for the US economy. As the model's endogenous variables are expressed in percentage deviation from the steady state, a transformation of the data is necessary to make it consistent with its representation within the model. The first variable is the output gap. We take the real domestic product from the FRED database (*GDPC1*) and express it in percentage deviation from its trend as measured by an adjusted one-sided HP filter.[17] The second variable is the quarterly inflation rate, also taken from the FRED database (*CPIAUCSL*), and measured based on the percentage change of the US consumption price index (CPI). This measure of inflation is used to maintain consistency with the third time series – namely, inflation expectations, that are taken from the SPF which consists of CPI forecasts. We consider the average one-quarter-ahead inflation expectation (*CPI3*). Because realized and expected inflation rates are linearized around the same steady-state value, we demean inflation and inflation expectation values with the sample average of realized inflation. Lastly, the conduct of monetary policy is summarized by the nominal interest rate. To consider the effects of unconventional policies in stylized structural models, we use the proxy funds rate of Choi et al. (2022) that captures both quantitative easing (QE) and tightening (QT).

The resulting mapping between the sample and the model is given by:

$$\begin{bmatrix} \text{Output gap} \\ \text{Inflation rate} \\ \text{Inflation expectations} \\ \text{Nominal rate} \end{bmatrix} = \begin{bmatrix} \hat{y}_t \\ \hat{\pi}_t \\ \mathbb{E}_t^{SL}(\hat{\pi}_{t+1}) \\ \hat{\imath}_t \end{bmatrix}$$

3.2 Parameter values

As usual in the business-cycle literature, the discount factor is calibrated and we use $\beta = 0.99$. The rest of the parameters are estimated through Bayesian techniques.

The prior distributions of the parameters are listed in the first three columns of Table 1 and are based on the existing literature. In detail, for exogenous disturbances, the standard deviations are given by an inverse gamma "type 2" distribution with a prior mean of 0.001 and a standard error of 0.1. Our prior is inspired by Christiano et al. (2014) but we allow for a less informative prior to capture possibly larger fluctuations induced by the COVID-19 recession. The persistence of stochastic processes is taken from Smets & Wouters (2007) with a beta distribution, a prior mean of 0.5 and a standard error of 0.2. The MA(1) prior distributions for the demand and cost-push shocks are also reminiscent of the ARMA markup shocks of Smets & Wouters (2007).

Regarding the structural parameters, the means of the parameters related to utility and monetary policy are in line with Smets & Wouters (2007), with slightly tighter priors to circumvent possible sources of weak identification. We estimate the slope of the NKPC as a composite parameter[18] whose prior mean and standard deviation are set to match the usual intervals of price stickiness found in the literature.

As for the learning parameters, we set the priors' means in line with the SL literature exemplified in Arifovic et al. (2013, 2018, 2024), Grimaud et al. (forthcoming). On a

[17]This allows us to avoid the well-documented bias at the end of the sample when using the standard HP filter (Hamilton 2018).

[18]From the estimated value, it is straightforward to retrieve the values of the Rotemberg (1982) adjustment costs through the microfoundations given some values of the Frisch elasticity of labor and the elasticity of substitution between goods, see Appendix B.6.

		Prior Distribution			Posterior Distribution		
		Shape	Mean	Std	Mode	Mean [5%:95%]	
Panel A: Shock processes							
Demand shock std	σ_g	$\mathscr{IG}_2$	0.001	1	0.0062	0.0067	*[0.0058:0.0076]*
Cost-push shock std	σ_u	$\mathscr{IG}_2$	0.001	1	0.0043	0.0043	*[0.0039:0.0048]*
Monetary policy shock std	σ_v	$\mathscr{IG}_2$	0.001	1	0.0023	0.0023	*[0.002:0.0025]*
AR(1) of the demand shock	ρ_g	$\mathscr{B}$	0.5	0.2	0.7424	0.7455	*[0.7048:0.7869]*
AR(1) of the cost-push shock	ρ_u	$\mathscr{B}$	0.5	0.2	0.6554	0.6328	*[0.5934:0.6771]*
MA(1) of the demand shock	μ_g	$\mathscr{B}$	0.5	0.1	0.4222	0.4579	*[0.3959:0.5395]*
MA(1) of the cost-push shock	μ_u	$\mathscr{B}$	0.5	0.1	0.5367	0.4887	*[0.4354:0.5386]*
Panel B: Structural parameters							
Slope of the NKPC	κ	$\mathscr{G}$	0.3	0.05	0.1913	0.2032	*[0.1741:0.24]*
Intertemp. elasticity	σ	$\mathscr{G}$	1.75	0.25	1.7	1.6993	*[1.6914:1.7091]*
Inflation stance	ϕ_π	$\mathscr{G}$	1.5	0.1	1.6983	1.7131	*[1.6952:1.733]*
Output stance	ϕ_y	$\mathscr{G}$	0.15	0.01	0.1617	0.1564	*[0.144:0.1723]*
MPR smoothing	ρ_ι	$\mathscr{B}$	0.7	0.1	0.8159	0.8179	*[0.7925:0.8411]*
Panel C: Social learning parameters							
Idiosyncratic news shock std.	σ_ι	$\mathscr{G}$	0.1	0.05	0.0006	0.0006	*[0.0006:0.0006]*
Aggregate news shock std	σ_λ	$\mathscr{IG}_2$	0.001	1	0.0004	0.0004	*[0.0003:0.0004]*
Decay in past forecast errors	ρ	$\mathscr{B}$	0.8	0.1	0.7978	0.7745	*[0.7041:0.8599]*
Frequency of news shocks	μ	$\mathscr{B}$	0.3	0.05	0.426	0.4357	*[0.3937:0.4808]*
Log marginal data density						-2818.12	

Table 1: Priors and posteriors of the estimated model from 1985Q1-2023Q4

technical note, those priors are set rather tight due to the non-linearity in the SL process which makes the minimization problem of the estimation rather difficult, even for the global optimization algorithm we use. The prior for the standard deviation of the aggregate news shock under SL is set in the same fashion as for the rest of the exogenous shocks.

The last three columns of Table 1 provide the posterior distributions of the estimated parameters. Looking at Panel A reveals that all the standard deviations of the shocks are different from zero and none of the processes exhibit unit-root behaviors. From Panel B, all posterior means are different from their prior means at a 10% confidence level, except for ϕ_y.

Panel C of Table 1 presents the values of the SL parameters. The standard deviations of the aggregate and idiosyncratic news shocks are strictly higher than zero. The decay in past forecast errors ρ used to evaluate each agent's subjective forecasting model implies that the half-life of forecast errors in the performance measure is $\frac{\ln \rho}{\ln 0.5} \approx 3$ quarters. The value of parameter μ implies that approximately 40% of the agents receive an inflation news every quarter.

We now examine the underlying forces driving fluctuations by breaking down the variance of each time series into distinct combinations of the shocks. The detail of the model validation and the comparison with the RE framework are presented in, respectively, Appendix A.2 and A.3.

3.3 Historical decomposition

Figures 2, 3 and 4 show the decomposition of the variance of, respectively, inflation, inflation expectations and the interest rate between the four types of shocks we consider, namely cost-push, demand, monetary policy and expectation shocks, over the period 2007Q1-2023Q4.[19] We defer the decomposition of the output gap variance to Fig. C.6 in Appendix since it is not the focus of the analysis, together with the decomposition of the other variables over the entire sample in Figs. C.3-C.5.

Fig. 2 shows the large and negative contribution of dispersed expectations to inflation over the 2010 decade of below-target inflation and in the aftermath of the COVID-19 pandemic until the beginning of 2022 (see the blue bars below the zero line). This effect operates through the inflation expectations channel, as evidenced by the overwhelming contribution of these shocks to the variance of subjective inflation expectations in our model (see Fig. 3).

Importantly, since our model nests the FIRE benchmark, it is always a possible outcome of the estimation exercise: time periods over which the expectation shocks play no substantial role in explaining the macroeconomic dynamics indicate data that are consistent with the FIRE model. This is notably the case in the mid-nineties (see Fig. C.3), around 2007, 2016-2017 and 2019 (Fig. 2). In this respect, the contribution of the expectation shocks to the great financial crisis (GFC), up until the end of 2009, appears negligible compared to the role played by the fundamental and the policy shocks. This is in stark contrast with the 'missing inflation' 2010 decade and the recent experience.

Furthermore, Fig. 4 illustrates the negative contribution of all shocks to the movements in the policy instrument since the GFC until the COVID-19 area. Despite our use of the proxy funds rate which accounts for the effects of unconventional monetary policy, the large contribution of monetary policy shocks over the 2010 decade until 2022 reflects a looser monetary policy stance compared to the interest-rate path compatible with the estimated Taylor rule; see the yellow vars below the zero line.

Focusing on the post-2020 area, inflation is initially subject to negative demand and monetary policy shocks at the onset of the COVID-19 pandemic (2020Q2-Q3); see, respectively, the orange and yellow areas in Fig. 2. Supply shocks then ensue at the beginning of 2021 and continue into 2022 (see the green bars). In the wake of the COVID-19 pandemic, the contribution of dispersed expectations to inflation first remains negative, which translates the sluggishness of below-target expectations inherited from the 2010s in our model (see the blue bars below the zero line). The contribution of expectation shocks then turns positive and explains most of the variance of inflation in the last two years of the sample (see the blue bars above the zero line).

Fig. 3 shows the corresponding drivers of subjective inflation expectations. In 2020Q2, the movements in expectations are overwhelmingly initiated by negative demand and monetary policy shocks (see the orange and yellow bar below the zero line), before the expectations shocks kick in and drive first expectations below, and less than two years later, above the target (see the thick blue area below the zero line until the beginning of 2022, and above afterwards). This pattern is accompanied by negative and then positive shocks in the CB's instrument, indicating a too loose and then too tight monetary policy stance compared to the one implied by the Taylor rule (see the yellow area since 2020Q2 on Fig. 4).

[19]The contributions of the demand, supply and monetary policy shocks are extracted by simulating the model considering one shock at a time under RE, which is the standard procedure in Dynare. The contribution of SL expectations is treated as the residual of this exercise and includes the effects of both the idiosyncratic and the aggregate news shocks and the history-dependent learning dynamics.

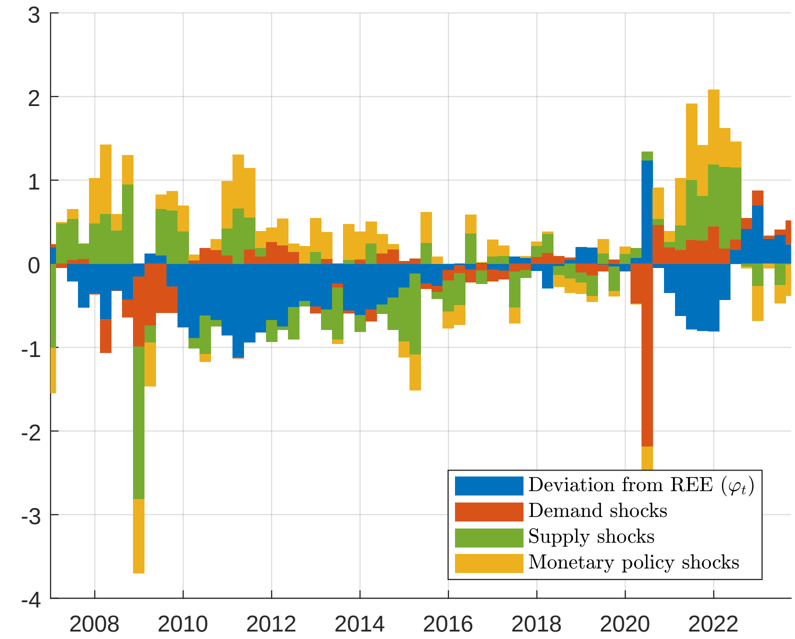


Figure 2: Historical decomposition of the inflation rate (2007Q1-2023-Q4)

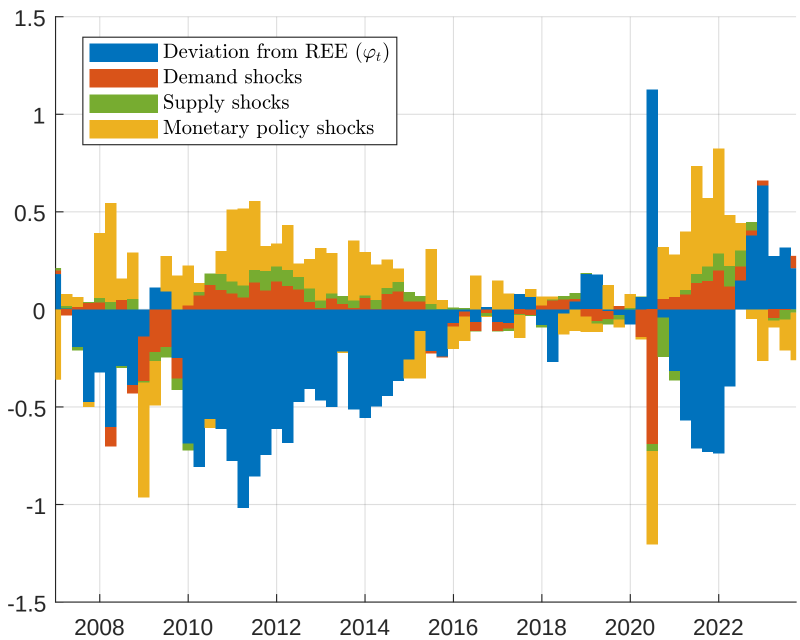


Figure 3: Historical decomposition of one quarter ahead of aggregate inflation expectation (2007Q1-2023-Q4)

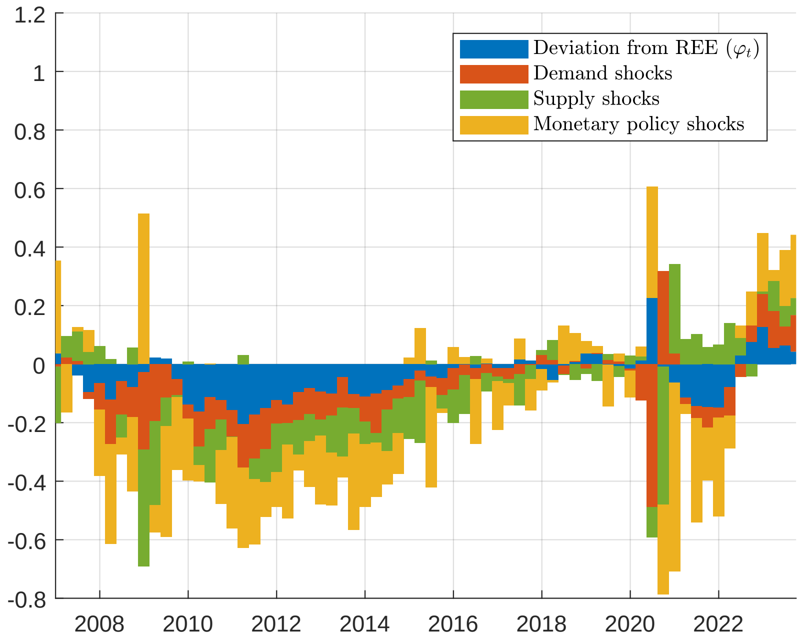


Figure 4: Historical decomposition of the policy rate (2007Q1-2023-Q4)

The recent disinflation appears therefore to be mostly due to a combination of transitory shocks and monetary policy actions, namely the easing of supply pressure, the normalization of monetary policy and the fading out of expectational shocks, i.e. the re-anchoring of inflation expectations to the Fed's target.

Before exploring the macroeconomic effects of these recent policy shocks in greater detail in Section 4, we take a look at the implied heterogeneity in expectations in our model with respect to its empirical counterpart, and explore its role in the transmission of monetary policy.

3.4 Heterogeneity in expectations over the business cycle

The estimation only takes the SPF average inflation expectations as an input, but our model generates the whole time-varying distribution of idiosyncratic subjective beliefs. It is therefore possible to compare this distribution with its (untargeted) empirical counterpart to further evaluate the empirical relevance of the model. It is then valuable to discuss the distinct role of heterogeneous expectations in the transmission of monetary policy in our framework.

Fig. 5 reports the time-varying 50 and 90% dispersion along with the median expectations since 2007 in the SPF (in red) and in the estimated model (in blue); see Fig. C.7 in Appendix C for the whole sample (the data in the figure are expressed in p.p. deviation from the target). It is revealing to see how much the two patterns overlap. The cross-sectional distributions display a similar dynamics during the GFC, where disagreement was stronger than during the following decade. The dispersion in both datasets is minimal over the 2010s, in particular towards the end of the decade, where inflation expectations are concentrated a little below the inflation target, both in the SPF and in the model. This observation is in line with these time periods being well-accounted for by the FIRE benchmark, as discussed in Section 3.3. In both datasets, disagreement then

	HENK MODEL	DATA [0.05;0.95]
Cross sections		
Average standard deviation of forecasts	0.2131	*[0.1740;0.2029]*
Average interquartile range of forecasts	0.2327	*[0.1908;0.2210]*
Average skewness of forecasts	0.1494	*[-0.4451;-0.1466]*
Average kurtosis of forecasts	4.6631	*[4.2470;5.1970]*
Correlations		
Correlation of output gap and standard deviation of forecasts	-0.3668	*[-0.3975;-0.1034]*
Correlation of inflation and standard deviation of forecasts	0.0663	*[-0.0260;0.2830]*
Correlation of output gap and interquartile range of forecasts	-0.1773	*[-0.3459;-0.0438]*
Correlation of inflation and interquartile range of forecasts	0.1914	*[0.0403;0.3428]*

Notes: Model-implied moments are computed across the sequence of filtered shocks that replicate the business cycle at the aggregate level.

Table 2: Empirical and model-implied cross-section moments along the filtered shocks

increases at the onset of the COVID-19 pandemic, and the distribution of inflation expectations shifts upwards and peaks above target in the beginning of 2022.

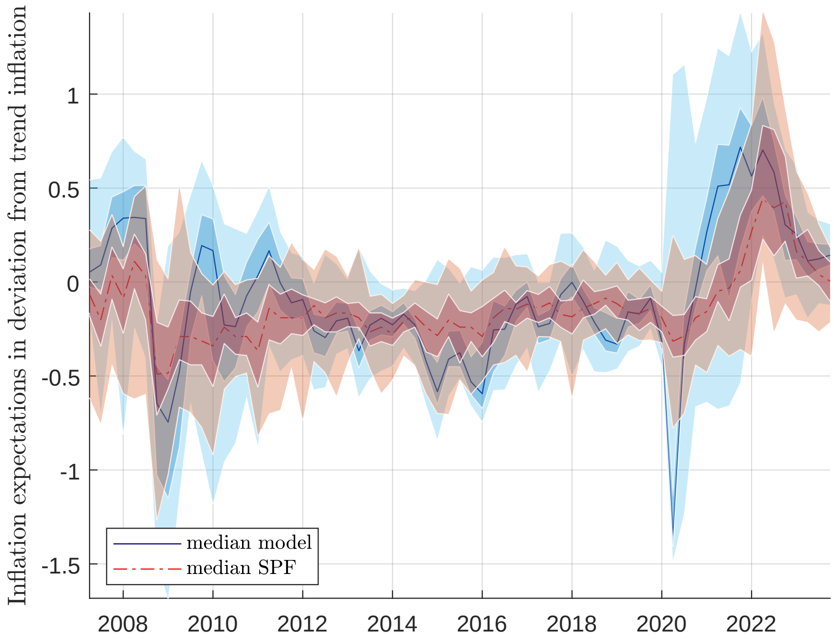
Fig. 6 provides another look at the heterogeneity in the SPF forecasts and in the model-generated expectations by comparing two measurements of their dispersion. The top and the bottom panels report the time series of, respectively, the interquartile range and the cross-sectional standard deviation of expectations in the SPF (dotted lines) and in the model (plain lines). The similarity in the time-variation pattern between the two data sources is striking, with a significant correlation of about 50% between the two time series for each of the two measurements considered. Our model captures particularly well the increase in dispersion that accompanied the GFC and the COVID-19 area, as well as the low dispersion over the Great Moderation period.

To go deeper into the validation exercise, Table 2 compares (untargeted) moments of the distribution of expectations in the SPF and in the model, in particular the co-movement of the cross-section with the business cycle. In this respect, we find that all magnitudes and signs observed in the data are also observed in the model. Our model further accounts for all other (untargeted) measurements of dispersion, to the exception of the average skewness which indicates that the model-generated distribution of beliefs is more skewed to the left compared to the average pattern seen in the SPF.

The collection of evidence presented so far along with Appendix A.2 suggests that our model is well-suited to account for both the recent macroeconomic developments and the dispersion of micro-level beliefs.

Importantly, moving beyond describing the heterogeneity in expectations, we may now use our framework to investigate the distinct role of this heterogeneity in the transmission of shocks and monetary policy. As alluded to in the introduction, this feature is suggested by the data but may not be systematically examined in homogeneous agent settings, such as the FIRE benchmark, adaptive learning models or most rational inattention frameworks.

To shed light on this matter, we simulate IRFs of the model under varying degrees of dispersion in inflation expectations. Specifically, at the onset of a shock in $t = 0$, each individual belief $a_{j,0}$ is assumed to follow a normal distribution with a mean of zero (i.e., $\varphi_0 = 0$, corresponding to the FIRE benchmark), and a standard deviation set as a proportion of the maximum model-generated standard deviation under the filtered shocks.



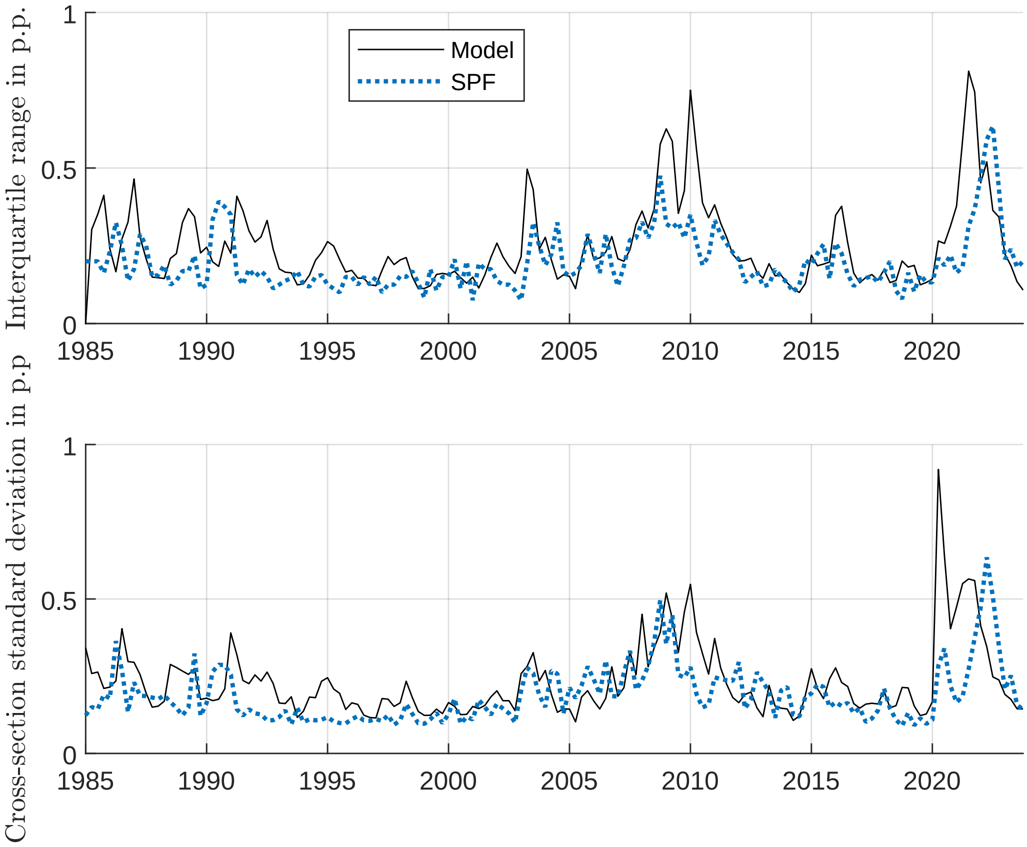
Notes: Here, and in all subsequent figures, data are expressed in p.p. on a quarterly basis and in deviation from the steady state value, which is 2.76% for inflation – i.e., the sample average. Hence, a peak in 2022 at 0.5% indicates $2.76 + 4 \times 0.5 = 4.76\%$ inflation on a yearly basis. Appendix Fig. C.2 displays the transformed inflation paths in year-over-year rates on a quarterly frequency for most figures presented in Section 4. The red and the blue areas represent respectively the distribution of the forecasts from the SPF and the model-generated subjective inflation expectations . The dark shaded areas contain 50% of the individual expectations in any period, and the light-shaded area 90% of them. The plain line is the median of the model-generated expectations and the dashed line is the median of the SPF.

Figure 5: Time-varying cross-sectional dispersion of inflation expectations

Fig 7 presents the mean dynamics of the endogenous variables over 1,000 Monte Carlo simulations after a cost-push shock. Figs 7, C.8, C.9 and C.10 in Appendix C report on the same exercise after, respectively, a demand, a monetary policy, and an expectational shock.

In all these figures, it is evident that the greater the initial dispersion in expectations, the larger and more persistent the ensuing fluctuations in all variables. This is particularly striking when looking at the effect of dispersion on aggregate inflation expectations, where a greater dispersion is also associated with larger deviations from the target. In turn, this larger unanchoring of expectations calls for higher-for-longer nominal rates to stabilize larger inflation gaps, at the price of deeper and more persistent output gaps. Therefore, the heterogeneity in inflation expectations amplify the effects of the shocks on the economy and complicates the stabilization task of the CB.

Within the context of our model, the intuition behind the effect of heterogeneous expectations on monetary policy transmission may be expressed as follows. A greater dispersion in beliefs leads to a more diverse population of forecasting strategies which then contains more 'extreme' beliefs, i.e. beliefs that are further away from the target. In face of a shock that temporarily moves inflation away from the target, these 'extreme' beliefs are temporarily validated since they result in lower errors when it comes to forecast recent inflation than beliefs that are closer to the target. Consequently, these 'extreme' beliefs may gain momentum during social interactions and diffuse in the population



Notes: Data are expressed in p.p. deviation from the target, see Fig. 5 for interpretation. The top panel represents the interquartile range of the distribution of inflation expectations in any period (i.e. $Q3-Q1$), in the SPF data (dashed blue line) and as generated by the model plain line). The correlation between the two time series is 0.538 (with a corresponding p-value < 0.001). The bottom panel displays the standard deviations of these distributions per period, with a correlation of 0.483 (and a corresponding p-value < 0.001).

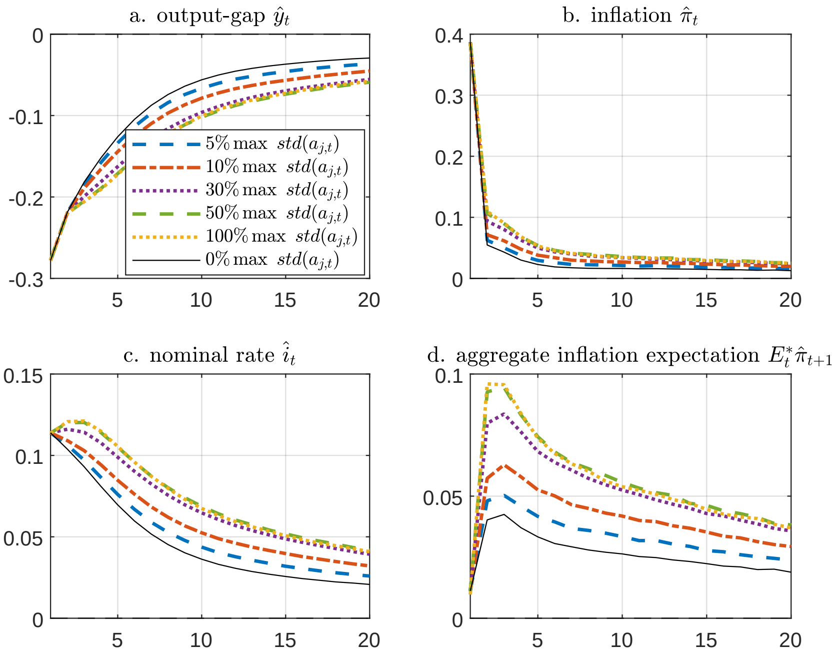
Figure 6: Model-generated vs. empirical heterogeneity in inflation expectations

of agents, which contributes to steer the average inflation expectation away from the target, which amplifies the effect of the exogenous shock. This mechanism enables a perpetual feedback loop, wherein higher macroeconomic volatility is associated with a greater heterogeneity in expectations which, in turn, contributes to amplify shocks and increase aggregate volatility.

We now proceed to policy simulations within the context of this model.

4 Policy counter-factual simulations

In what follows, we experiment with counter-factual policy scenarios to assess the effect of the timing and the strength of the Fed's reaction to the recent inflation surge on inflation expectations and the macroeconomic variables. Table 3 contrasts these policy scenarios with respect to the historical benchmark using the micro-founded welfare measure derived in Appendix B.7 and an ad-hoc aggregate loss function as commonly used in policy institutions.



Notes: Initial dispersion of beliefs is assumed to be centered on zero, normally distributed and measured as shared of the maximum standard deviation generated by the sequence of shocks mapping the sample. Each line is the mean dynamics of 1000 MC simulations.

Figure 7: Response to a one standard deviation supply shocks as function of initial heterogeneity in expectation

4.1 Delayed tightening cycles and unanchoring of expectations

We start by illustrating how inflation expectations may become unanchored or, in other words, how an inflation scare can develop and entrench above-target inflation in the model. To do so, in Fig. 8, we simulate what would have happened had the Fed delayed its tightening cycle in 2021 by one, two, four or, even, eight quarters.[20]

From the simulations we can clearly see that delays in the tightening cycle generate higher inflation expectations and realizations than under the historical scenario. This worsening of economic outcomes is also reflected in the lower welfare values in these counter-factual scenarios in Table 3 (see the first four rows of Panel B, where above-unity values represent a welfare loss over the historical benchmark and below-unity values an improvement). Longer delays in initiating the tightening cycle result in larger upward movements in expectations, which translates into a more severe unanchoring of these expectations and a stronger inflation surge. The extreme case of a two-year delay generates runaway inflation and a large overheating of the economy, as evident from the dynamics of the output gap in this scenario.

Within the context of our model, a delayed tightening implies higher output, marginal costs and inflation, which contributes to the diffusion of above-target inflation expectations through the social interactions. Due to the self-referential nature of the NK Phillips Curve, these higher inflation expectations feed higher inflation, which further increases

[20] In our simulations, the path of the nominal rate is constrained through unexpected monetary policy shocks. Considering anticipated shocks instead would only strengthen the inflationary effect of these shocks on the expectations and the macroeconomic variables.

Scenario	Ad-Hoc measure $\mathscr{L}_{2021Q1,2023Q4}$ (as a ratio of the benchmark)	Microfounded measure $\mathscr{U}_{2021Q1,2023Q4}$ (as a ratio of the benchmark)
Benchmark	1	1
A. Alternative MP shocks		
No MP shocks after 2021Q1	**0.385**	**0.955**
B. Timing of the tightening		
One-quarter delay	1.0890	1.0040
Two-quarter delay	1.2140	1.0090
Four-quarter delay	2.1890	1.0510
Eight-quarter delay	5.7600	1.1670
+100 basis point in 2021Q1	**0.336**	**0.949**
+400 basis point in 2021Q1	**0.304**	**0.932**
+800 basis point in 2021Q1	**0.555**	**0.922**
$\rho_\iota + 10\%$ more inertia	2.591	1.142
$\rho_\iota - 10\%$ less inertia	**0.643**	**0.967**
C. Hawk and dove		
$\phi_\pi - 10\%$ more dovish	1.057	1.052
$\phi_\pi + 10\%$ more hawkish	**0.960**	**0.981**
D. Crossed scenarios		
$\rho_\iota + 10\%$ and $\phi_\pi - 10\%$	3.060	1.299
$\rho_\iota + 10\%$ and $\phi_\pi + 10\%$	2.534	1.102
$\rho_\iota - 10\%$ and $\phi_\pi - 10\%$	**0.640**	**0.994**
$\rho_\iota - 10\%$ and $\phi_\pi + 10\%$	**0.645**	**0.950**
E. Early cuts		
100 basis point 2022Q3	1.200	1.005
100 basis point 2022Q4	1.102	1.003
100 basis point 2023Q1	1.097	1.002
100 basis point 2023Q2	1.079	1.001

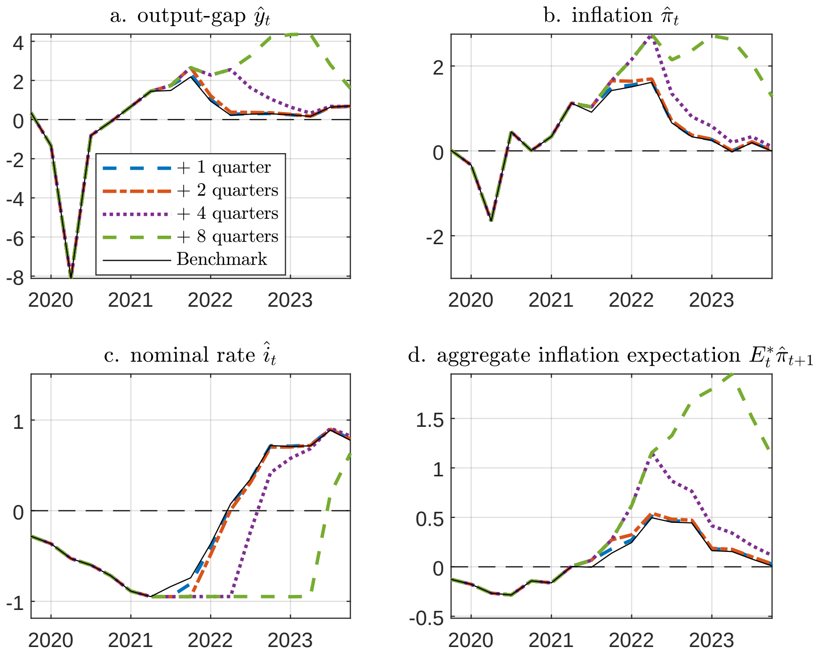
Note: Lower-than-one (resp. higher-than-one) ratios indicate that the scenario considered leads to higher (resp. lower) welfare than the historical benchmarks scenario. Numbers in bold indicate welfare-improving scenarios compared to the benchmark.

Table 3: Welfare ratios to the benchmark (2021Q1-2023Q4)

these expectations.

Fig. 9 digs further into the dynamics of the distribution of inflation expectations under the different scenarios. It represents the time-varying share of inflation beliefs (i.e., the idiosyncratic variables $a_{j,t}$) that lie between 1% and 3% on an annualized basis. Our interpretation of the $\{a_{j,t}\}$ as the idiosyncratic biases toward the long-run anchor is intuitive: the case $a_{j,t} = 0, \forall j$, represents the RE case where expectations are anchored at the target. Therefore, Fig. 9 illustrates a measure of the anchoring of inflation expectations or, in other words, the credibility of the CB's target or the size of the inflation scare: the lower the share of agents that hold long-run beliefs between 1 and 3%, the more unanchored expectations, the lower the credibility of the target, and the more severe the inflation scare.

In the historical scenario, this share falls to 40% in mid-2022 before recovering above 90% by the end of 2023. Consistent with the insights of Fig. 8, all delayed-tightening scenarios result in further unanchoring of expectations compared to the historical scenario. The extreme case of a two-year delay results in a 100% loss in the target's credibility,



Notes: See Fig. 5.

Figure 8: Aggregate dynamics under delayed tightening cycle

which explains the runaway inflation observed in this scenario.

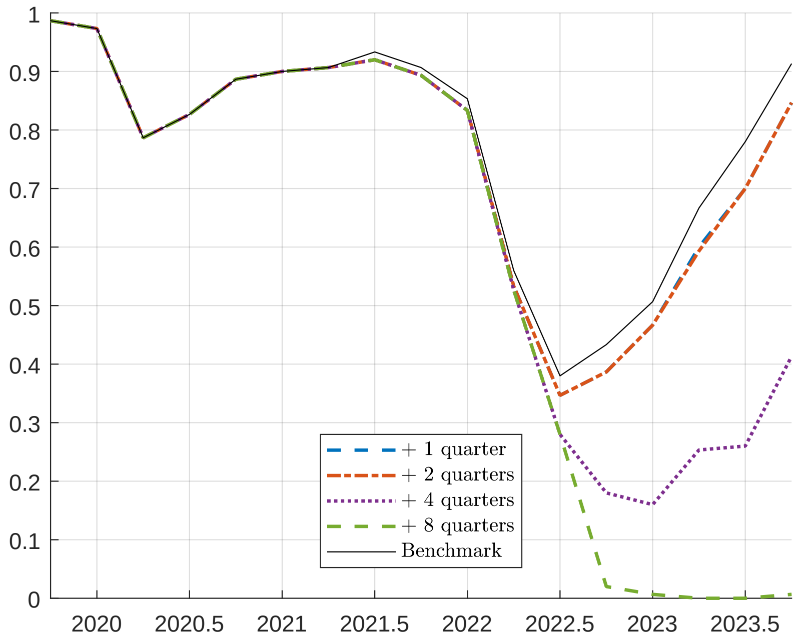
This first set of counter-factual simulations demonstrates the possibility of endogenous inflation scares in the model. We now investigate a scenario absent monetary policy shocks.

4.2 What would have happened had the Fed strictly followed the Taylor rule?

The series of substantial negative monetary policy shocks from mid-2020 until mid-2022 discussed in Section 3.3 and depicted in Fig. C.1 in Appendix C indicates that the Fed was too accommodative during these two years compared to what the Taylor rule would have prescribed. In other words, interest rates, even once accounting for unconventional monetary policy *via* the use of the proxy funds rate, were below the path compatible with the estimated interest-rate setting rule. Could the inflation surge in 2021-2022 have been avoided had the Fed strictly followed the Taylor rule? If so, at which costs in terms of economic activity?

We simulate a counterfactual scenario using the estimated model where we remove the monetary policy shocks after 2020Q4 so that the Fed sets the proxy rate according to the Taylor rule instead – see the dashed blue line in Fig. C.1. This scenario is contrasted with the actual economic developments under the benchmark scenario in Fig. 10. The Taylor rule prescribes an earlier and more aggressive tightening than the actual rate path, as evidenced from the plot of the proxy funds rate on Panel C.

The resulting inflation surge in 2021 is of substantially smaller magnitude compared to the benchmark; see Panel B. The difference reaches about 0.75 p.p. in deviation from the target on a quarterly basis and persists for several periods. Had the Fed tightened



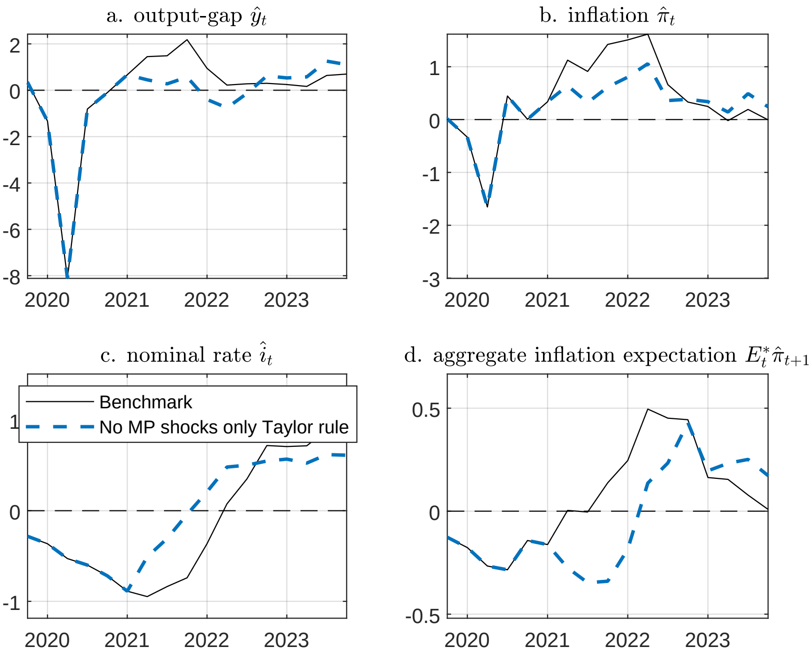
Notes: We convert the $a_{j,t}$ in annualized level in order to ease the assessment of the anchoring of expectation.

Figure 9: Share of $a_{j,t}$ between 1% and 3% in annualized rate under delayed tightening cycle

earlier in accordance with the Taylor rule, inflation would have been peaked about 3% lower on a yearly basis. This earlier tightening contributes to keeping inflation expectations lower for longer (for almost a year longer) and once the increase in expectations materializes, it is more gradual, does not peak as high and for not as long as under the benchmark scenario; see Panel D. In other words, if we measure expectations anchoring as the cumulative distance between the average inflation expectations and the target (zero in the model), expectations would have remained better anchored under the Taylor rule scenario than in the benchmark historical scenario.

What is most compelling for the case for an earlier and stronger tightening is the dynamics of the output gap under this alternative. While the reduction in the inflation gap is intuitive, the prominence of cost-push shocks in the inflation dynamics in 2021, as identified in Fig. 2, may suggest that such preemptive increases in the nominal rate would have been detrimental to output. Looking at Panel A of Fig. 10, we see that this is not the case. While the post-COVID-19 expansion that accompanied the relatively loose monetary policy in 2021 and early 2022 would have been milder under a preemptive tightening, the output gap would only have been briefly negative (in 2022Q2), which would not have been enough to trigger a recession. This is because under the scenario of an earlier tightening, the CB can exploit the sluggishness in subjective expectations and delay the over-shooting of these expectations above the target which eventually results from the positive cost-push shocks in 2021 and 2022. A better anchoring of inflation expectations in turn softens the CB's trade-off between the stabilization of inflation and economic activity, which explains the relatively modest output cost in the Taylor rule scenario for bringing down inflation earlier and by a greater magnitude.

As a conclusion, the strict implementation of the Taylor rule would have prescribed a preemptive tightening which would have delayed and dampened the unanchoring of inflation expectations, thus limiting further the inflation scare in the wake of the recent



Notes: See Fig. 5.

Figure 10: Aggregate dynamics under historical and counterfactual MP shocks

inflationary pressures. The macroeconomic outcome would have involved a smaller inflation surge at a negligible cost in terms of output. This beneficial outcome is evident from the welfare measures in Panel A in Table 3, where the gain over the historical scenario is among the strongest of all scenarios considered.

Next, we analyze various monetary policy counter-factual simulations in the spirit of Walsh (2022), where we vary first the timing and then the strength of the monetary policy reaction to the recent inflationary pressures.

4.3 Earlier tightening cycles and soft-landing

We have shown that the Fed fell behind the curve. One way to assess the effects of preemptive tightening is to simulate positive monetary policy shocks that front-load right in 2021Q1 part of the subsequent tightening cycle. Fig. 11 shows three examples of such scenarios: the dashed red lines represent a 100 basis-point increase in the Fed's instrument, the dotted green line a 400 basis-point increase, which represents half of the increase over the entire historical tightening cycle, and the nested case where the whole tightening is front-loaded at once in the dotted-dashed blue line.[21] These exercises illustrate the substantial effect on expectations and the economy of such a preemptive tightening (panel D). Under the half- and full-tightening front-load scenarios, it is clear that the departure of inflation from the target is minimal (Panel C), but comes at the cost of a double-dip in the output gap given the cost-push shocks identified in Section 3.3 (Panel A). Note that front-loading only half of the tightening eventually results in a lower peak in the interest rate for a much smaller inflation gap than the historical scenario.

[21]This extreme scenario is only considered for completeness. Note that our stylized framework allows us to shed light on the interplay between interest rate adjustments and expectations coordination and anchoring but it does not account for the effects of monetary policy on the other parts of the economy.

Most relevantly, a smaller and more realistic preemptive tightening of 100 basis points at the beginning of 2021 results in a proxy funds rate about 1% smaller on an annual basis than the benchmark scenario, a flattening of the interest rate path towards the end of 2022, contrary to the further increase that historically occurred mid-2023, and an inflation gap about half of what was actually observed. The difference in the inflation dynamics stems from lower-for-longer inflation expectations, which then only rebound beyond the target amid the series of supply shocks that affected inflation in 2022 (see, again, Fig. 2). However, even a 100 basis-point front-loading of the tightening cycle comes at the cost of two quarters of negative output gaps that invariably occur in each front-loading scenario as a result of this series of cost-push shocks.

Yet, in the case of the 100 basis-point shock, the output gap remains positive up until early 2022. A preemptive tightening therefore appears to exploit the sluggishness in subjective expectations stemming from the information frictions and prolong the below-target trend in expectations, which helps stabilize inflation at a relatively modest output cost.

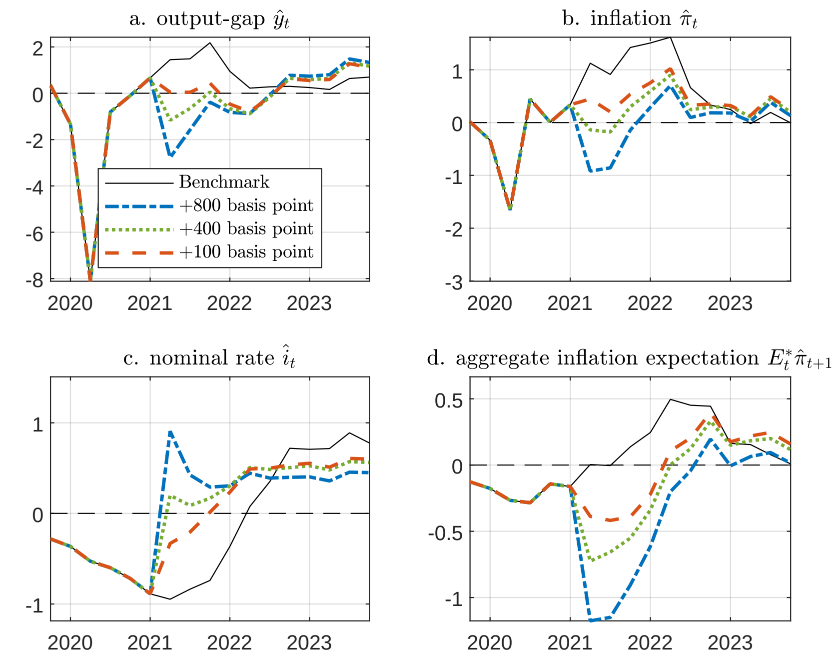
The fifth to the seventh rows of Panel B in Table 3 compare welfare measurements under these three scenarios. Both the ad-hoc and the micro-founded welfare measures indicate a U-shaped relationship between the size of the early tightening and the welfare gain over the benchmark:[22] Front-loading larger interest rate increases initially dampens the inflation surge but the output costs become increasingly prohibitive. These counter-factual exercises illustrate the merits and limits of preemptive actions of the CB to manage expectations and soften the stabilization trade-off in the wake of adverse fundamental shocks.

Another way to assess the timing of the monetary policy reaction is to simulate what would have happened under a different Taylor rule, which would involve more or less inertia compared to the estimated coefficient of $\rho_\iota = 0.83$ (see, again, Table 1). A higher smoothing parameter can be interpreted as a delayed response, and a smaller one as an earlier reaction of the Fed to the inflation build-ups. We therefore compare scenarios where we increase or decrease this coefficient by 10%, resulting in a scenario where $\rho_\iota^{\text{late}} = 0.91$ and one with $\rho_\iota^{\text{early}} = 0.74$. Would a faster or slower reaction have helped soften the stabilization trade-off in the wake of the post-pandemic inflation surge?

The results of these policy experiments are presented in Fig. 12, and the welfare comparisons are given in the last two rows of Panel B in Table 3. The 'early' scenario (see the yellow dotted lines and the last row of Panel B) confirms the main message of Section 4.2: less inertia prescribes a quicker increase in the interest rate once inflation starts rising beyond the target in early 2021, which results in lower inflation expectations, a lower inflation gap and a more moderate output expansion than in the historical scenario. The difference in inflation between the 'early' and the historical scenario amounts to a non-negligible 0.25% on a quarterly basis, which represents 1% on an annual basis. In this 'early' scenario, output gap crosses the zero line only in a single quarter (2022Q2) where it barely reaches -0.1%, which is not enough to yield a recession.

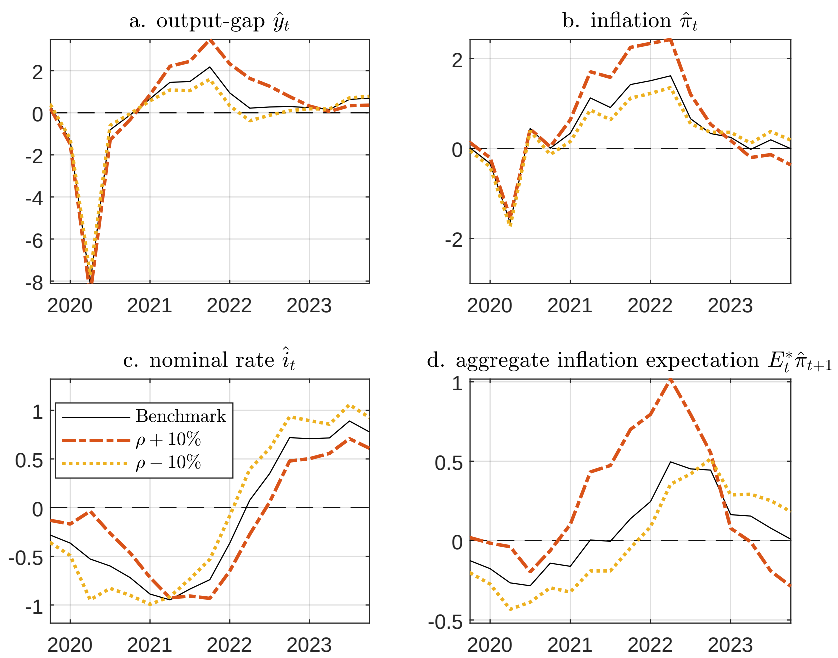
By contrast, under the 'late' scenario (see the red dashed lines in Fig. 12), the tightening of monetary policy is delayed compared to the historical scenario. This delayed scenario would have resulted in a (much) more salient unanchoring of inflation expectations, i.e. a (much) more severe inflation scare, and an annual inflation rate of about 4% higher than in the historical scenario. These worse economic outcomes are reflected in the (much) lower welfare achieved in this scenario with respect to the benchmark (see the one-but-last row of Panel B in Table 3).

[22] The maximum welfare gain arises for a stronger early increase – not displayed in the table – in the micro-founded than in the ad-hoc welfare criterion. This is due to the absence of a financial sector in our model and our use of the proxy fund rate, which is more volatile than the Fed fund rate.



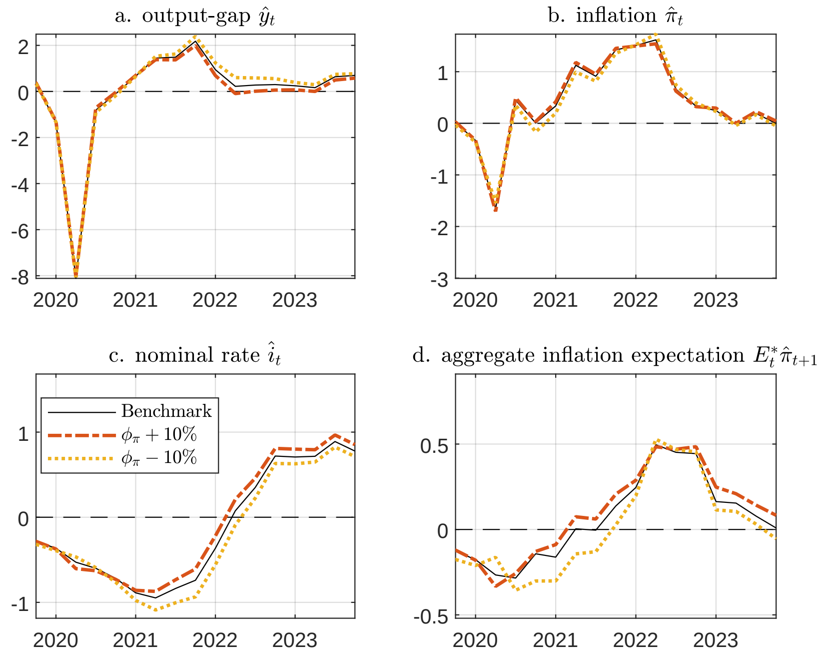
Notes: See Fig. 5. The largest shock of 8% represents the front-load of the full-cycle of subsequent tightening.

Figure 11: Aggregate dynamics with preemptive tightening



Notes: See Fig. 5.

Figure 12: Aggregate dynamics under historical and counterfactual inertia in the Taylor rule



Notes: See Fig. 5.

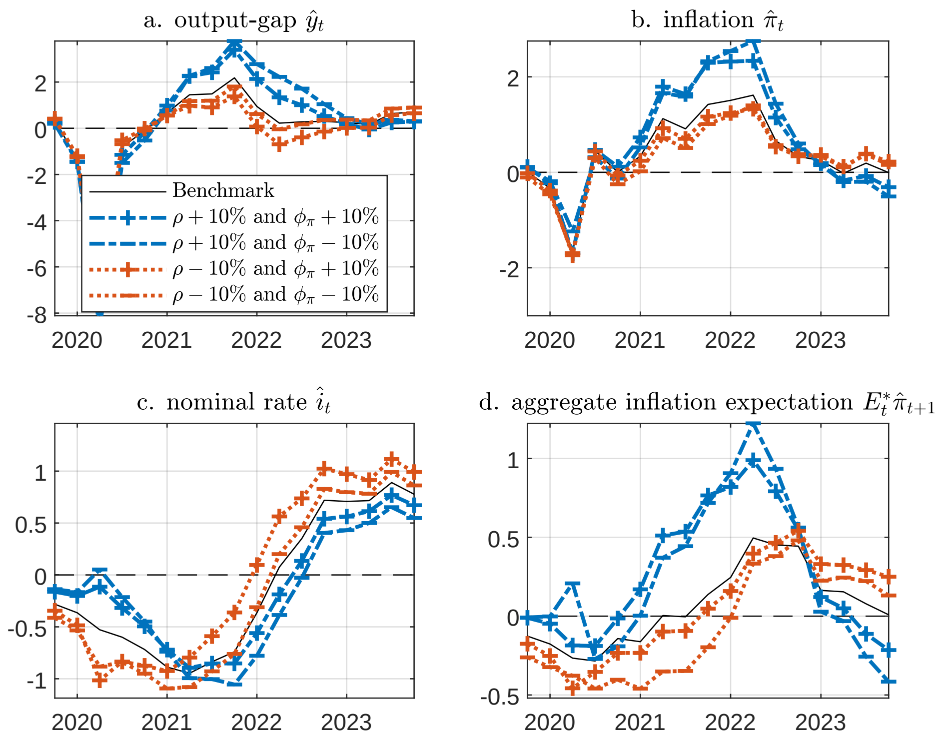
Figure 13: Aggregate dynamics under historical and counterfactual monetary policy stances on inflation

4.4 More of an hawk or more of a dove?

Beyond the timing of the reaction to the recent inflationary pressures, could the Fed have compensated the fall behind the curve by implementing a more hawkish monetary policy? Alternatively, what would have been the outcome of a more dovish scenario? To address these questions, we now vary the strength of the policy reaction to inflation by experimenting with the coefficient ϕ_π in the Taylor rule. We model an hawkish scenario with a 10%-increase in ϕ_π beyond its estimated value of 1.82 (see again Table 1), i.e. $\phi_\pi^{\text{hawk}} = 2$, and a dovish scenario with a 10%-lower-than-estimated reaction to inflation, i.e. $\phi_\pi^{\text{dove}} = 1.64$.

These policy experiments are presented in Fig. 13. The striking insight is that the three scenarios do not dramatically differ from each other. While the 'hawk' scenario features an earlier rise in the interest rate (red dashed-dotted lines) and the 'dove' scenario delivers the opposite pattern (orange dotted lines), the magnitude of the differences in terms of inflation and output gaps appears negligible. Despite higher rates earlier, the hawkish scenario does not reduce the inflation gap with respect to the historical one, but closes the output gap as of early 2022. The corresponding welfare measures, given in Panel C of Table 3, are in line with this observation: the more hawkish scenario achieves a welfare gain and the more dovish one a welfare loss but the differences with respect to the benchmark are small.

The hawkish scenario suggests that there is no way to compensate for falling behind the curve by reacting more strongly to the build-up of inflation. To shed more light on the matter, we experiment by varying simultaneously the timing and the strength of the monetary policy reactions. Fig. 14 presents the outcomes of these policy counter-factual scenarios. It is evident from this figure that the timing, rather than the size of the reaction



Notes: See Fig. 5.

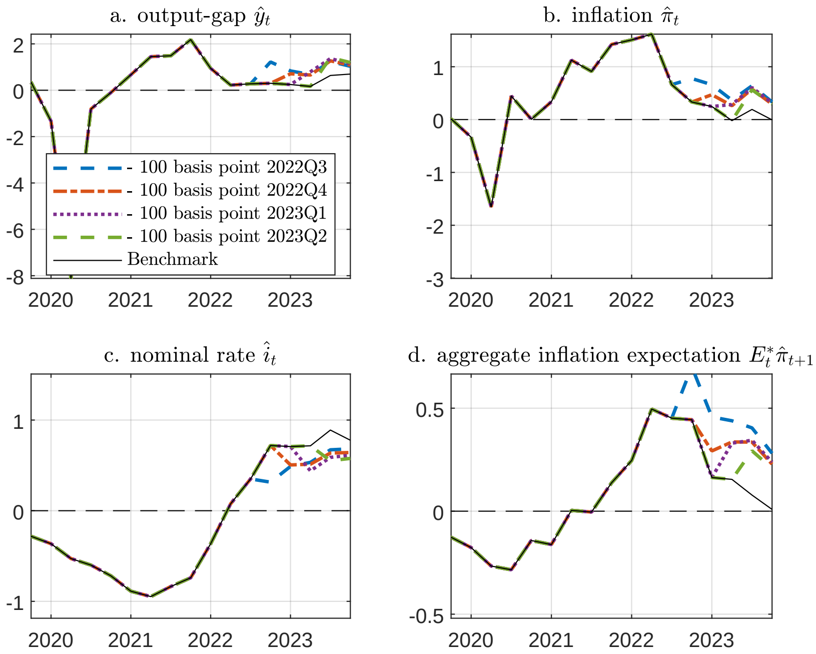
Figure 14: Aggregate dynamics under historical and counterfactual Taylor rules

to the inflation build-up, is what matters: see how close the two red lines (representing an earlier reaction) and the two blue lines (picturing a delayed tightening) are from each other.

Considering the two scenarios with an earlier reaction (i.e. the two red dotted scenarios in Fig. 14), the magnitudes of the differences in terms of inflation and output with respect to the historical benchmark are not large. The only noticeable difference concerns the dynamics of inflation expectations: with an earlier but slightly less aggressive tightening (the red dotted curve with '-'), expectations remain lower for longer compared to the historical scenario, and they do not exceed the steady state before 2022. This scenario is therefore the most favorable in terms of exploiting the sluggishness of expectations to further reduce the inflation peak (by about 1% on a yearly basis compared to the benchmark peak), while avoiding any quarter of negative output gap. However, under this alternative scenario, expectations are also more sluggish on the way down back to the target and remain above the steady state at the end of the sample. Welfare comparisons in Panel D of Table 3 confirm this discussion: adjusting the strength of the reaction makes little differences in terms of welfare and the key welfare-improving policy is an earlier policy adjustment (i.e. a lower ρ_ι than in the benchmark scenario).

4.5 To cut or not to cut?

In September 2024, the Fed changed its instrument for the first time since the last increase in 2023Q3 and decreased it by 25bps – although the proxy funds rate that we use did go down by about 50 basis points in the last quarter of 2023, which is our last data point. In this section, we look into the consequences on inflation expectations and



Notes: See Fig. 5. The largest shock of 8% represents the front-load of the full-cycle of subsequent tightening.

Figure 15: Aggregate dynamics under counter-factual early interest rate cuts

macroeconomic variables of earlier interest rate cuts. We consider illustrative counter-factual scenarios of a 100 basis point cut in any quarter since 2022Q3.

The results are reported in Fig. 15 where we clearly see that the earlier the cut, the larger the ensuing re-bounce in inflation expectations and, hence, inflation and output gap beyond their targeted levels. The scenario of a cut as early as 2022Q3 is most telling of the risk of inflation scare when the stance of monetary policy is too loose (blue dashed line). In this scenario, inflation expectations resume their increase by about 1% further up (on an annual basis) right after the cut, and annual inflation remains on average 2% higher than the historical scenario. Interestingly, no matter the timing of the cut, neither inflation nor inflation expectations would have converged back to the target by the end of 2023 had the Fed cut rates earlier. In any of the earlier cut scenarios we have simulated, inflation would have invariably landed close to 2% above target by the end of 2023 instead. Panel E of Table 3 confirms that no early cut scenario is welfare-improving with respect to the historical course of action.

5 Conclusion

The recent surge in inflation has been resolved without major output losses. What is the origin of this soft landing? Previous historical episodes, such as the Volcker disinflation in the 1980s, could have suggested otherwise, although both the nature of the shocks and the institutional frameworks substantially differ from the ones prevailing in the 1970s-1980s. In other words, to what extent was the soft landing a matter of policy, i.e. a timely adjustment of the Fed's instrument that kept inflation expectations anchored in the face of inflation build-ups?

To investigate this question, we develop a macroeconomic framework with information frictions that result in subjective and heterogeneous beliefs about the low-frequency component of inflation. The rest of the model is otherwise standard and estimated with

Bayesian methods. Our model opens the door to the possibility of inflation scares which may arise whenever inflationary shocks validate above-target beliefs so that they diffuse in the population of heterogeneous expectations. As a result, in our framework, the dispersion of expectations *per se* is costly, these expectations may not always mean-revert to the target as shocks dissipate, and can instead become unanchored. We identify the contributions of price, demand, policy, and expectation shocks to the recent economic dynamics.

Counter-factual policy simulations indicate that monetary policy was instrumental in limiting to a large extent the unanchoring of expectations, although an inflation scare arose. The timing, rather than the strength of the reaction to inflation, was the key to limit the depth of this inflation scare. Delaying further the tightening cycle would have resulted in a higher inflation peak, and a delay of about three quarters at least would have resulted in a sizable unanchoring of inflation expectations. Nevertheless, we find that the Fed did fall behind the curve in 2021, since a preemptive tightening could have resulted in inflation peaking nearly 3% below the experienced level at a negligible cost for output. A preemptive tightening could have exploited the sluggishness in expectations due to information frictions, which could have kept expectations lower for longer and softened the stabilization trade-off when responding to the series of adverse supply shocks in 2022.

Beyond this policy insight, we offer a flexible yet rigorous macroeconomic framework with heterogeneous expectations that can be estimated using the state-of-the-art empirical methods. This framework can be applied to a variety of contexts where time-varying heterogeneity and subjective beliefs may play a role in shaping aggregate dynamics. In particular, our framework can straight-forwardly be extended to larger-scale models, for instance to include labor market dynamics, thanks to our related toolbox. This endeavor is left for future research.

References

Adam, K. & Marcet, A. (2011), 'Internal rationality, imperfect market knowledge and asset prices', Journal of Economic Theory **146**(3), 1224 – 1252. 6

Adjemian, S., Juillard, M., Karamé, F., Mutschler, W., Pfeifer, J., Ratto, M., Rion, N. & Villemot, S. (2024), Dynare: Reference Manual, Version 6, number 80.
URL: *https://ideas.repec.org/p/cpm/dynare/080.html* 10

An, S. & Schorfheide, F. (2007), 'Bayesian analysis of DSGE models', Econometric Reviews **26**, 113–172. 34

Andrade, P., Gaballo, G., Mengus, E. & Mojon, B. (2019), 'Forward guidance and heterogeneous beliefs', American Economic Journal: Macroeconomics **11**(3), 1–29. 3, 40

Arifovic, J. (1995), 'Genetic algorithms and inflationary economies', Journal of Monetary Economics **36**(1), 219–243. 3

Arifovic, J. (1996), 'The Behavior of the Exchange Rate in the Genetic Algorithm and Experimental Economies', Journal of Political Economy **104**(3), 510–541. 3

Arifovic, J., Bullard, J. & Kostyshyna, O. (2013), 'Social learning and monetary policy rules', The Economic Journal **123**(2001), 38–76. 3, 11

Arifovic, J., Grimaud, A., Salle, I. & Vermandel, G. (2024), 'Social Learning and Monetary Policy at the Effective Lower Bound', Journal of Credit, Money and Banking . 3, 4, 6, 11, 48, 49

Arifovic, J., Schmitt-Grohé, S. & Uribe, M. (2018), 'Learning to live in a liquidity trap', Journal of Economic Dynamics and Control **89**, 120–136. 3, 11

Assenza, T., Heemeijer, P., Hommes, C. & Massaro, D. (2021), 'Managing self-organization of expectations through monetary policy: A macro experiment', Journal of Monetary Economics **117**, 170–186. 4

Bernanke, B. & Blanchard, O. (forthcoming), 'What Caused the U.S. Pandemic-Era Inflation?', American Economic Journal:Macroeconomics . 2, 5

Branch, W. A. & Evans, G. W. (2011), 'Monetary policy and heterogeneous expectations', Economic Theory **47**, 365–393. 3

Branch, W. A. & Gasteiger, E. (2019), Endogenously (Non-)Ricardian Beliefs, ECON WPS - Working Papers in Economic Theory and Policy 03/2019, TU Wien, Institute of Statistics and Mathematical Methods in Economics, Economics Research Unit. 3

Carvalho, C., Eusepi, S., Moench, E. & Preston, B. (2023), 'Anchored inflation expectations', American Economic Journal: Macroeconomics **15**(1), 1–47. 3, 6

Choi, J., Doh, T., Foerster, A. & Martinez, Z. (2022), 'Monetary Policy Stance Is Tighter than Federal Funds Rate', FRBSF Economic Letter **2022**(30), 1–5. 11

Christiano, L. J., Motto, R. & Rostagno, M. (2014), 'Risk shocks', American Economic Review **104**(1), 27–65. 11

Clarida, R., Gali, J. & Gertler, M. (1999), 'The Science of Monetary Policy: A New Keynesian Perspective', Journal of Economic Literature **37**(4), 1661–1707. 1

Cornea-Madeira, A., Hommes, C. & Massaro, D. (2019), 'Behavioral Heterogeneity in U.S. Inflation Dynamics', Journal of Business & Economic Statistics **37**(2), 288–300. 4

Cuba-Borda, P., Guerrieri, L., Iacoviello, M. & Zhong, M. (2019), 'Likelihood evaluation of models with occasionally binding constraints', Journal of Applied Econometrics **34**(7), 1073–1085. 10, 34

Evans, G., Gibbs, C. & McGough, B. (2024), 'A unified model of learning to forecast', American Economic Journal: Macroeconomics . forthcoming. 3

Evans, G. & McGough, B. (2024), 'Social learning and expectational stability', Journal of Economic Dynamics & Control . forthcoming. 3

Evans, G. W. & Honkapohja, S. (2001), Learning and Expectations in Macroeconomics, Princeton University Press. 3

Fair, R. C. & Taylor, J. B. (1983), 'Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models', Econometrica **51**(4), 1169–1185. 34

Fernández-Villaverde, J., Rubio-Ramírez, J. F. & Schorfheide, F. (2016), Solution and estimation methods for DSGE models, in 'Handbook of Macroeconomics', Vol. 2, Elsevier, pp. 527–724. 34

Gabaix, X. (2020), 'A Behavioral New Keynesian Model', American Economic Review **110**(8), 2271–2327. 4

Gagliardone, L. & Gertler, M. (2023), Oil Prices, Monetary Policy and Inflation Surges, NBER Working Papers 31263, National Bureau of Economic Research, Inc. 2, 5

Galí, J. (2015), Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications, Princeton University Press. 47

Gáti, L. (2023), 'Monetary policy & anchored expectations – An endogenous gain learning model', Journal of Monetary Economics **140**, S37–S47. 3

Goodfriend, M. (1993), 'Interest rate policy and the inflation scare problem: 1979-1992', Economic Quarterly (Win), 1–24. 2

Goodfriend, M. & King, R. G. (2005), 'The incredible Volcker disinflation', Journal of Monetary Economics **52**(5), 981–1015. 2

Greenwood, J., Hercowitz, Z. & Huffman, G. W. (1988), 'Investment, capacity utilization, and the real business cycle', American Economic Review pp. 402–417. 40

Grimaud, A., Salle, I. & Vermandel, G. (forthcoming), 'A Dynare Toolbox for Social Learning Expectations', Journal of Economic Dynamics and Control . 5, 7, 10, 11

Guerrieri, L. & Iacoviello, M. (2017), 'Collateral constraints and macroeconomic asymmetries', Journal of Monetary Economics **90**, 28–49. 34

Hajdini, I. (2023), Mis-specified Forecasts and Myopia in an Estimated New Keynesian Model, Technical Report 22-03R, Federal Reserve Bank of Cleveland. Working Paper. 3

Hakamada, M. & Walsh, C. (2024), Policy Delays under Rational and Behavioral Expectations, Working Papers 24-42, IMF. 2

Hamilton, J. D. (2018), ‘Why you should never use the hodrick-prescott filter’, Review of Economics and Statistics **100**(5), 831–843. 11

Hommes, C., Mavromatis, K., Özden, T. & Zhu, M. (2023), ‘Behavioral learning equilibria in new keynesian models’, Quantitative Economics **14**(4), 1401–1445. 3

Jia, C. & Wu, J. C. (2023), ‘Average inflation targeting: Time inconsistency and ambiguous communication’, Journal of Monetary Economics **138**, 69–86. 3

Kiley, M. T. & Roberts, J. M. (2017), ‘Monetary policy in a low interest rate world’, Brookings Papers on Economic Activity **2017**(1), 317–396. 49

King, R. G. & Lu, Y. K. (2022), Evolving reputation for commitment: The rise, fall and stabilization of us inflation, Technical Report 30763, National Bureau of Economic Research. 2

Kydland, F. E. & Prescott, E. C. (1977), ‘Rules Rather than Discretion: The Inconsistency of Optimal Plans’, Journal of Political Economy **85**(3), 473–491. 1

Lucas, R. E. & Sargent, T. (1981), Rational Expectations and Econometric Practice, University of Minnesota Press. 1

Mankiw, G. & Reis, R. (2002), ‘Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve’, Quarterly Journal of Economics **117**(4), 1295–1328. 9

Mauersberger, F. (2021), ‘Monetary policy rules in a non-rational world: A macroeconomic experiment’, Journal of Economic Theory **197**, 105203. 4

Meeks, R. & Monti, F. (2023), ‘Heterogeneous beliefs and the Phillips curve’, Journal of Monetary Economics **139**, 41–54. 3

Orphanides, A. & Williams, J. C. (2004), Imperfect Knowledge, Inflation Expectations, and Monetary Policy, in B. S. Bernanke & M. Woodford, eds, ‘The Inflation-Targeting Debate’, University of Chicago Press, pp. 201–246. 4

Ozden, T. (2024), ‘Heterogeneous Expectations and the Business Cycle at the Effective Lower Bound’, International Journal of Central Banking . forthcoming. 3, 7

Reifschneider, D. (2024), US monetary policy and the recent surge in inflation, Working Papers 24-13, Peterson Institute for International Economics. 2

Rotemberg, J. J. (1982), ‘Sticky prices in the United States’, Journal of Political Economy **90**(6), 1187–1211. 11, 43

Rychalovska, Y., Slobodyan, S. & Wouters, R. (2024), Survey Expectations, Adaptive Learning and Inflation Dynamics, Technical Report 781, CERGE-EI. 3

Sargent, T. J. (2022), Rational Expectations and Volcker’s Disinflation, in R. G. King & A. L. Wolman, eds, ‘Essays in Honor of Marvin Goodfriend: Economist and Central Banker’, Federal Reserve Bank of Richmond, pp. 279–288. 1

Slobodyan, S. & Wouters, R. (2012a), ‘Learning in a medium-scale DSGE model with expectations based on small forecasting models’, American Economic Journal: Macroeconomics **4**(2), 65–101. 3

Slobodyan, S. & Wouters, R. (2012b), ‘Learning in an estimated medium-scale DSGE model’, Journal of Economic Dynamics and Control **36**(1), 26–46. 3

Smets, F. & Wouters, R. (2007), ‘Shocks and frictions in US business cycles: A Bayesian DSGE approach’, American economic review **97**(3), 586–606. 11, 40

Walsh, C. E. (2022), Inflation Surges and Monetary Policy, IMES Discussion Paper Series 22-E-12, Institute for Monetary and Economic Studies, Bank of Japan. 2, 6, 23

A Model estimation and validation

A.1 Methodology

The model is estimated using Bayesian techniques (see An & Schorfheide 2007 for an overview). The posterior distribution associated with the vector of observable variables is computed numerically using a Markov chain Monte Carlo sampling approach. However, taking non-linear models to the data is a challenge because non-linear filters, which are required to form the likelihood function, are computationally expensive. An inversion filter has emerged as a computationally cheap alternative (see, e.g., Guerrieri & Iacoviello 2017). Pioneered by Fair & Taylor (1983), this filter recursively extracts the sequence of innovations by inverting the observation equation for a given set of initial conditions. Unlike other commonly used filters (such as the Kalman or particle filters), the inversion filter relies on an analytic characterization of the likelihood function.[23] Therefore, we first need to rewrite the non-linear model for the purpose of the estimation.

To perform the inversion, it is computationally convenient to separate the linear part of the model that is solved using standard methods from the non-linear part that depends on the SL algorithm. Therefore, the inversion is performed based on the following linear recursive solution of the HENK model:

$$Z_t = A(\Theta)Z_{t-1} + B(\Theta)\Omega_t \tag{A.1}$$

with $Z_t \equiv (z_t\ \varphi_t)'$ the vector of stacked endogenous variables which includes the deviation of aggregate expectations from RE and $\Omega_t \equiv (\varepsilon_t\ \lambda_t)'$ the vector of structural shocks augmented by the common expectational shock, while $A(\Theta)$ and $B(\Theta)$ are matrices that depend on set of structural parameters Θ. It is easy to see that the recursive form (A.1) is equivalent to the presentation in Section 2:

$$A(\Theta) = \begin{bmatrix} P(\Theta) & R(\Theta) \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}, B(\Theta) = \begin{bmatrix} Q(\Theta) & R(\Theta)S(\cdot) \\ \mathbf{0}_{1\times 3} & S(\cdot) \end{bmatrix},$$

where matrices P, Q and R are solved with standard techniques embedded in our Dynare toolbox and depend on the structural parameters gathered in the set Θ. $S(\cdot)$ summarizes the aggregate outcome of the SL learning process detailed in Section 2.3 that depends on the aggregate news shock λ_t contained in the shock matrix Ω_t. The idiosyncratic news shocks, the fitness evaluation and the tournament are implicit in $S(\cdot)$.

Next, consider a sample Y_t of size T, the inversion filter maps a subset of endogenous variables to the set of observable, $Y_t = \omega Z_t$, where ω is a selection matrix of the subset. By replacing Eq. (A.1), one can determine the set of shocks as follows:

$$\Omega_t = (\omega B(\Theta))^{-1}(Y_t - \omega A(\Theta)Z_{t-1}) \tag{A.2}$$

The Bayesian method seeks to characterize the likelihood function of the model conditional on the matrix of observations through time T, denoted by $\mathscr{L} = l(\theta|Y_{1:T})$, where θ corresponds to a subset of estimated parameters $\theta \subset \Theta$, while non-estimated parameters are calibrated. We incorporate prior information about the structural parameters θ by specifying a prior distribution $f(\theta)$. The posterior distribution of $f\left(\theta|y_{1:T}\right)$ is then determined by Bayes theorem: $f(\theta|Y_{1:T}) = f(Y_{1:T}|\theta)f(\theta) / \int f(Y_{1:T}|\theta)f(\theta)\,d\theta$. Finally, we use the Metropolis-Hastings algorithm as a sampler to draw from the parameter uncertainty. We obtain a random draw of 800,000 from the posterior distribution of the

[23] For a presentation of alternative filters to calculate the likelihood function, see Fernández-Villaverde et al. (2016). See also Cuba-Borda et al. (2019) for details on the relative gains of the inversion filter.

parameters. In detail, we use eight parallel chains drawing 100,000 iterations, with a common jump scale parameter to match an acceptance rate of approximately 30%.

This next section discusses the dynamic properties of the model.

A.2 Model evaluation

We present impulse response functions (IRFs) of several variables of interest to the shocks considered in the model. This analysis is useful in assessing how shocks to economic variables propagate in the HENK model and checking that the model adequately captures the statistical properties of the macroeconomic data. Figure A.1 displays response of the economy to a 1% increase in each of the four shocks: demand, cost-push, monetary policy, and news shocks. We immediately remark that the IRFs are overall consistent with the business cycle theory.

For instance, the effect of the monetary policy shock in the first column of Fig. A.1 illustrates the standard policy transmission in the model: a higher nominal rate triggers intertemporal substitution where the households trade less consumption now for higher returns on their saving, which compresses aggregate demand and generates a drop in marginal costs and inflation. Upon observing the shock, agents revise their inflation expectations downwards because their forecasts are model-consistent. When lower inflation rates materialize, the selection process during social interactions further favors more pessimistic inflation expectations over the pre-shock ones, which contributes to the sluggishness of expectations. Because of the self-referential nature of inflation in the NKPC, this temporary unanchoring of short-run inflation expectations contributes to the prolongation of the effect of the shock[24] but eventually, all variables converge back to their steady state.

Similar effects on aggregate inflation expectations arise in the IRFs to the demand and supply shocks in the second and third column of Fig. A.1. The estimated ARMA(1,1) supply shocks being less autocorrelated than the demand shocks, their effect on inflation expectations is weaker than the effect of the demand or monetary policy shocks.

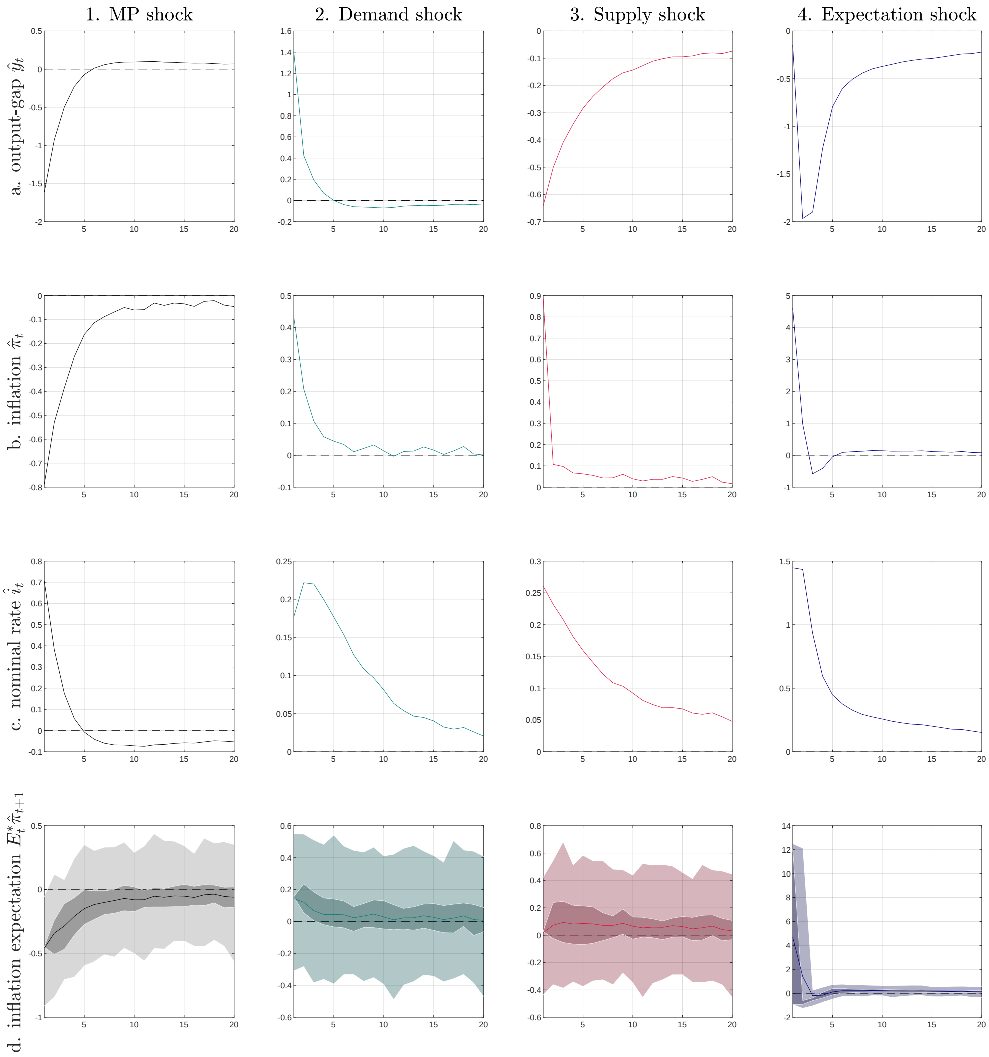
Looking at the expectation shock in the fourth column of Figure A.1, higher inflation expectations trigger higher inflation through the NKPC relation. Thanks to the Taylor rule, the CB reacts by tightening the policy rate, which pushes the real rate up and output down. It is striking to see that despite the white-noise nature of the expectation shock considered, it generates a persistent deviation of the endogenous variables from the steady state. This is because higher inflation beliefs temporarily lead to smaller forecast errors amid the temporary rise of inflation triggered by the shock, and therefore tend to spread through social interactions, which shifts upward the distribution of subjective beliefs. Eventually, the monetary policy tightening brings the economy back to steady state and aggregate expectations on target. Thus, we observe the typical result that learning amplifies the effects of shocks and adds persistence in the endogenous variables.

Finally, we compare the HENK model with its RE counterpart.

A.3 Model comparison

Capturing salient features of the formation process of expectations remains a central challenge and the vast majority of estimated macroeconomic models have remained tied to FIRE. The structural inference utilized in this paper is able to determine from a

[24] This additional persistence is a well-documented feature in the learning literature in macroeconomics that is referred to in the introduction for example.



Notes: Each column corresponds to the IRFs to a 1% increase in the four shocks i.e., from the left to the right, demand, supply, monetary policy, and expectational shocks, or the four variables i.e., in row from the top to the bottom, the output gap, inflation, the interest rate the aggregate inflation expectations (data are expressed in p.p. on a quarterly basis and in deviation from their steady state). In this last row, the shaded areas represent the distribution of the subjective beliefs, where the dark-shaded areas contain 50% of them in any period, and the light-shaded area 90%.

Figure A.1: Impulse response functions of the model

statistical standpoint which of the two models, i.e. the RE or the HENK model, is most likely to have generated the *whole* data sample. Because the RE model is nested into its HENK counterpart, one can formally test the null hypothesis $H_0 : \varphi_t = 0$, i.e. our structural model implies RE.[25] Employing an uninformative prior distribution over models (i.e., a 50% prior probability for each model), we find that the null hypothesis is rejected ($p < 0.0001$) and our HENK specification is statistically more relevant than the RE formulation to explain the whole sample.

[25]To perform this comparison exercise, the competing models have to be estimated using the same data sample. Because expectations are observable in the HENK model, we suppose that in the RE counterpart, expectations are subject to white-noise news shocks. For this reason, forecast errors under RE may be auto-correlated, see Table A.1 below, because these news shocks propagate through the lagged endogenous variables.

To further compare the two models, one can derive the moment statistics to evaluate which observable variables one model can better account for than the other. Table A.1 provides the empirical second moments of our four observable variables generated by the RE and the HENK specifications and the 95% confidence interval in the sample.

Consider first the standard deviations generated by the two models in the first panel. Both models are able to account for output gap and inflation volatility but fall (slightly) outside the confidence interval for the interest rate and expectations volatility. In the second panel, we can see that both models exhibit similarly good performances at capturing the autocorrelations of output gap and interest rate, but only the HENK model replicates the positive auto-correlation in inflation, inflation expectations and forecast errors. This difference highlights a key property of the HENK model: the model of subjective beliefs involves information frictions which create sluggishness in their updating process when new macroeconomic realizations are released and aggregate and idiosyncratic news shocks occur. This mechanism opens up the possibility of inflation expectations durably drifting away from the target as a result of the steady-state learning. By contrast, under RE, expectations are completely model-consistent and the (low) persistence in their dynamics directly stems from the persistence implied by the estimated exogenous shock processes. This low auto-correlation in expectations under RE systematically implies fast mean reversion in the wake of disturbances.

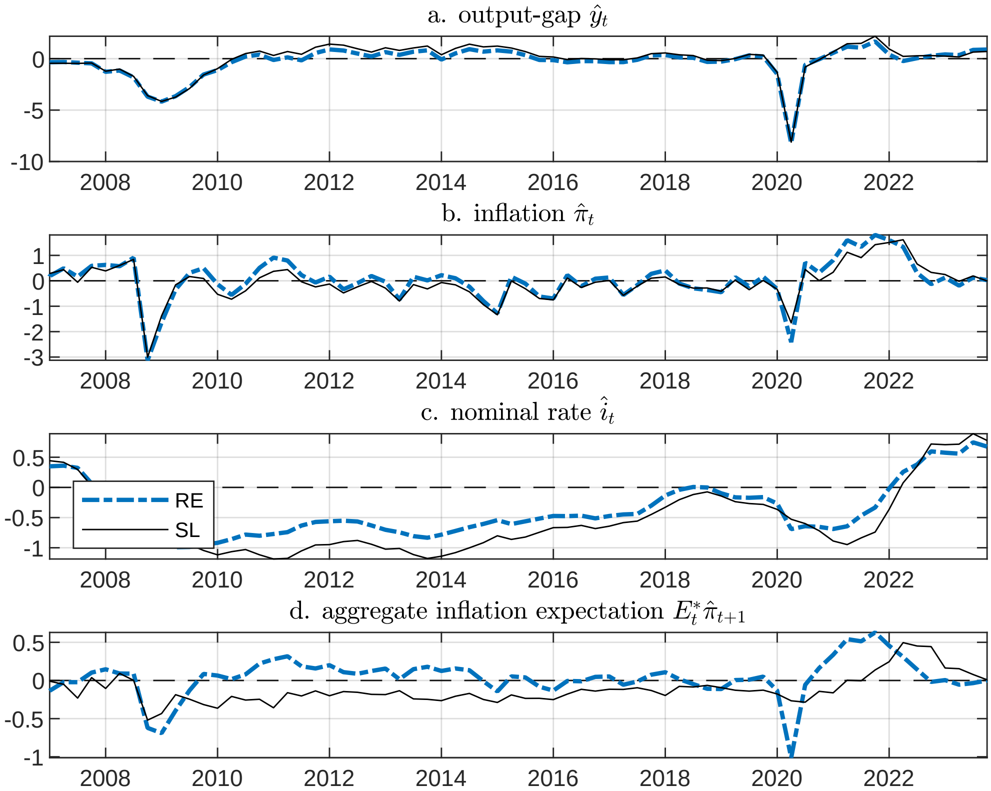
Finally, Figure A.2 compares the dynamics of the HENK model (solid black line) with its RE counterpart (blue dashed line) over the period 2007Q1-2023Q4, and Fig. A.3 does so over the entire sample. The RE model is not estimated but uses the posterior parameter values found in the HENK model together with its filtered shocks. Hence, the RE simulation needs to be understood as a counterfactual scenario absent learning dynamics and news shocks. The difference between the RE and the HENK models isolates the role of SL expectations in the model dynamics.

Per design, model-consistent inflation expectations under RE are strikingly better anchored (i.e., their deviations from the target are closer to zero) than in the HENK model where they err below target during the entire 2010 decade. Our estimated framework therefore accounts well for the below-target inflation of the 2010s thanks to persistently below-target expectations. Under RE, higher inflation expectations imply higher inflation, higher nominal rates, and lower output than in the HENK model. Furthermore, at the onset of the COVID-19 pandemic, RE produce a deeper plunge and a quicker overshooting of expectations than observed in the HENK model where expectations are sluggish. The unanchoring of short-run expectations is delayed in the HENK model compared to RE, but they remain elevated for a longer time, which produces higher inflation towards the end of the sample.

	DATA [0.05;0.95]	HENK MODEL	RE MODEL
Standard deviations			
Output gap	*[1.195;1.494]*	1.234	1.269
Inflation rate	*[0.490;0.613]*	0.598	0.605
Interest rate	*[0.660;0.826]*	0.554	0.445
Inflation expectations	*[0.207;0.259]*	0.391	0.365
Inflation forecast errors	*[0.4547;0.5685]*	0.562	0.803
Autocorrelations			
Output gap	*[0.640;0.826]*	0.515	0.501
Inflation rate	*[0.239;0.579]*	0.428	0.162
Interest rate	*[0.969;0.987]*	0.892	0.835
Inflation expectations	*[0.839;0.927]*	0.824	0.161
Inflation forecast errors	*[0.120;0.412]*	0.122	-0.145

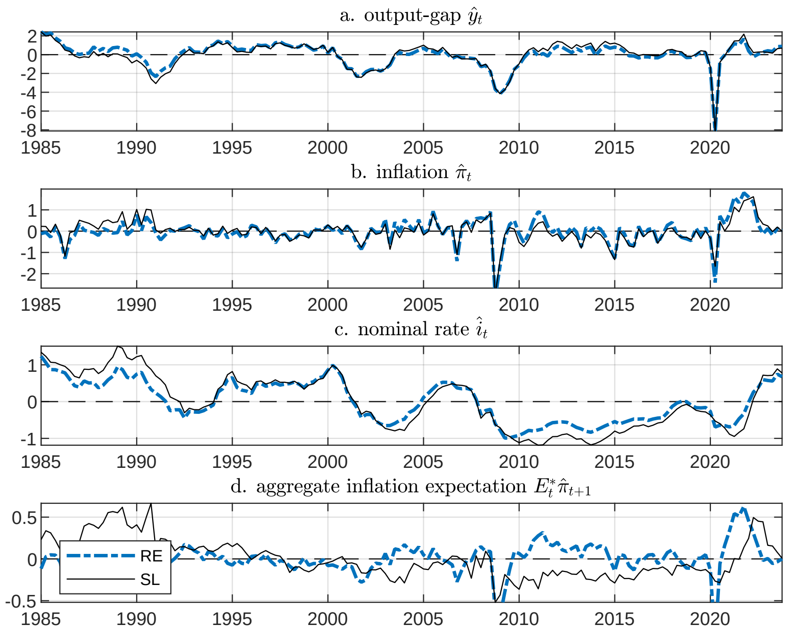
Notes: Model-implied moments are computed across 100 Monte Carlo iterations of the same duration as the sample.

Table A.1: Empirical and model-implied moments



Notes: See Fig. 5.

Figure A.2: Contractual exercises under FIRE and SL (2007Q1-2023Q4)



Notes: See Fig. 5.

Figure A.3: Comparison of the model's dynamics under SL and RE

B Microfoundation of the HENK model

We develop here the microfundation of the three-equation HENK model. For that, the time and the number of agents are discrete. Specifically, we assume that the economy is populated by an infinitely-living family composed by a discrete number J of members, indexed by $j \in [1,...,J]$. Each member operates an intermediary-good-sector firm so that indexes J indifferently refer to either firms or households, which we may then call agents, and firms and households have the same discount factor. These agents are identical (in particular in terms of preferences and technology, etc.) except when it comes to their inflation expectations that are heterogeneous. In what follows, $\mathbb{E}^*_{j,t}$ is the generic expression of the expectation in t of agent j. This expectation can be rational, in which case the star-superscript is dropped, or can follow social learning, in which case $* = SL$.

B.1 Households

Each agent j decides about their consumption, labor, and saving plans in order to maximize the household's welfare:

$$\mathbb{E}^*_{j,t}\sum_{\tau=0}^{\infty}\beta^{\tau}u(c_{j,t+\tau},h_{j,t+\tau}) = \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}^*_{j,t}\sum_{\tau=0}^{\infty}\beta^{\tau}\frac{1}{1-\sigma'}\left(c_{j,t+\tau}-\chi\frac{h^{1+\varphi}_{j,t+\tau}}{1+\varphi}\right)^{1-\sigma'}. \tag{B.1}$$

Utility $u(\cdot)$ is increasing in consumption $c_{j,t}$ and decreasing in labor $h_{j,t}$ where σ' and φ are curvature parameters and β is the non-stochastic discount factor.

The non-separable utility function is based on the GHH utility function of Greenwood et al. (1988).

Agents face an intertemporal problem: they determine the value of their consumption $c_{j,t}$, hours worked $h_{j,t}$ and real bond holdings $b_{j,t}$ so as to maximize the welfare of the family under the following budget constraint which binds in every period:

$$c_{j,t}+b_{j,t} = \frac{i_{t-1}}{\pi_t}\frac{b_{j,t-1}}{\exp\left(\varsigma_g g_t\right)} + w_t h_{j,t} + T_{j,t} + z_{j,t} + \Pi_{j,t}, \tag{B.2}$$

where w_t is the real wage (which is symmetric across all agents because they all have the same marginal product of labor under constant returns to scale); i_{t-1} the nominal interest rate payable on nominal bond holdings; π_t the inflation rate between periods $t-1$ and t; $\Pi_{j,t}$ the share of agent j of the real profits from monopolistic competition (see Section B.2); $T_{j,t}$ lump-sum government transfers that may be positive or negative; $z_{j,t}$ zero-sum intra-household transfers[26] and g_t an exogenous source of aggregate fluctuations (referred to as the risk-premium shock in Smets & Wouters (2007)) and affected by the elasticity parameter $\varsigma = \sigma'\frac{(1-\chi)}{\vartheta}$ which normalizes the shock with respect to the formulation in the textbook three-equation NK model.

Agents choose their consumption and savings plans conditional on their inflation and output-gap expectations. Hence, heterogeneity in expectations may entail heterogeneous consumption and wealth values, which poses a challenge for aggregation, in particular when saving is used in the production of capital goods. In the textbook NK

[26] These transfers impose homogeneous end-of-period wealth across households through an intra-risk sharing plan (see, e.g., Andrade et al. 2019). In each period, after making possibly heterogeneous consumption choices due to possible heterogeneous inflation expectations, each member's bonds holding is made identical *via* an agreement to distribute the number of bonds equally across the family. In detail, a transfer plan $z_{j,t} = b_{j,t} - B_t/J$ to each member j in each period t, where $B_t = \sum_{j\in J} b_{j,t}$ is the aggregate bonds holdings, equalizes post-transfer wealth. In equilibrium, the sum of transfers is zero $\sum_{j=1}^{J} z_{j,t} = 0$.

model, labor is the only input in the production function, and under the usual hypothesis of no restriction to the access to asset markets, the level of consumption is solely determined by the Euler equation so that the idiosyncratic bond holdings do not affect current consumption. Therefore, the usual permanent income hypothesis holds at the agent-level. Note that the idiosyncratic saving stock measured by $b_{j,t}$ may be positive if savings or negative if borrowing, and that the family member is not allowed to run a Ponzi scheme, namely:

$$\lim_{T\to\infty} \mathbb{E}^*_{j,t}\left(\Lambda_{j,t,T}\frac{B_T}{P_T}\right) \geq 0 \tag{B.3}$$

with $\Lambda_{j,t,T} \equiv \beta\frac{\partial u(c_{j,T},h_{j,T})/\partial c_{j,T}}{\partial u(c_{j,t},h_{j,t})/\partial c_{j,t}}$ the stochastic discount factor of the household. The aggregate demand for government bonds reads as follows: $B_t = \sum_{j=1}^{J} b_{jt}$.

Each household j solves the following problem:

$$\max_{\{c_{j,t},h_{j,t},b_{jt,}\}} \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}^*_{j,t}\sum_{\tau=0}^{\infty}\beta^{\tau}\left[\frac{1}{1-\sigma'}\left(c_{jt+\tau}-\chi\frac{h_{j,t+\tau}^{1+\varphi}}{1+\varphi}\right)^{1-\sigma'}\right.$$
$$\left.+\lambda'_{j,t+\tau}\left(\frac{i_{t-1+\tau}}{\pi_{t+\tau}}\frac{b_{j,t-1+\tau}}{\exp(\varsigma_g g_{t+\tau})}+w_{t+\tau}h_{j,t+\tau}+z_{j,t+\tau}+\Pi_{j,t+\tau}-c_{j,t+\tau}-b_{j,t+\tau}\right)\right].$$

The first-order conditions are given by

$$w_t\lambda'_{j,t} = \chi h_{j,t}^{\varphi}\left(c_{j,t}-\chi\frac{h_{j,t}^{1+\varphi}}{1+\varphi}\right)^{-\sigma'},$$
$$\lambda'_{j,t} = \left(c_{j,t}-\chi\frac{h_{j,t}^{1+\varphi}}{1+\varphi}\right)^{-\sigma'},$$
$$\exp\left(\varsigma_g g_t\right)\lambda'_{j,t} = i_t\beta\mathbb{E}^*_{j,t}\frac{\lambda^c_{j,t+1}}{\pi_{t+1}}.$$

Log-linearizing each first-order condition yields

$$\hat{w}_t = \varphi\hat{h}_{j,t}, \tag{B.4}$$

and:

$$\hat{\lambda}'_{j,t} = -\sigma'\left(\bar{c}_j-\chi\frac{\bar{h}_j^{1+\varphi}}{1+\varphi}\right)^{-1}\left(\bar{c}_j\hat{c}_{j,t}-\chi\bar{h}_j^{1+\varphi}\hat{h}_{j,t}\right), \tag{B.5}$$

$$\hat{\lambda}'_{j,t} = \hat{\imath}_t-\varsigma_g g_t+\mathbb{E}^*_{j,t}\left\{\hat{\lambda}'_{j,t+1}-\pi_{t+1}\right\}, \tag{B.6}$$

where variables with a hat denote deviations from their deterministic steady-state values denoted by a bar. Eq. (B.4) shows that wages equal the marginal product of labor which is the same across all agents j due to their non-separable preferences and the linear production technology.

Moreover, at the deterministic steady state, all agents have the same information[27] and, therefore, consume the same amount of goods. It follows that: $\bar{c}_j = \bar{c}$, $\bar{h}_j = \bar{h}$, $\forall j$.

Equalizing Eqs. (B.5) and (B.6) yields:

$$\bar{c}\hat{c}_{j,t}-\chi\bar{h}^{1+\varphi}\hat{h}_{j,t} = -\frac{\vartheta}{\sigma'}\left(\hat{\imath}_t-\mathbb{E}^*_{j,t}\hat{\pi}_{t+1}\right)+\frac{\vartheta\varsigma_g}{\sigma'}g_t+\mathbb{E}_t\left[\left(\bar{c}\hat{c}_{jt+1}-\chi\bar{h}^{1+\varphi}\hat{h}_{j,t+1}\right)\right], \tag{B.7}$$

with $\vartheta = \bar{c}_j-\chi\frac{\bar{h}^{1+\varphi}}{1+\varphi}$.

[27] At the deterministic steady-state, there is no news shocks, neither idiosyncratic nor aggregate, and all agents hold homogeneous beliefs, i.e. $\bar{a}_j{}^{\pi} = 0$, $\forall j$.

B.2 Firms

To introduce a monopolistic-competition framework, the production process of goods is divided between two types of firms: intermediate and final firms. Intermediate firms produce different types of goods which are imperfect substitutes. We assume that each member j owns an intermediate-sector firm j that produces an intermediate good y_j and generates the profit $\Pi_{j,t}$ (in Eq. (B.2)). Final firms produce a homogeneous good by combining all intermediate goods $\{y_j\}$, $j = 1, \ldots, J$.

Final sector. The final-good producers are retailers. They buy the intermediate goods and package them into the aggregate supply of goods which, in equilibrium, equals the aggregate good demand from households, denoted by Y_t^D. On a perfectly competitive market, final producers take the price P of the goods as given and maximize profits as follows:

$$P_t Y_t^D - \sum_{j=1}^{J} p_{j,t} y_{j,t}, \tag{B.8}$$

subject to the supply constraint:

$$Y_t^D = \left(J^{-1/\epsilon} \sum_{j=1}^{J} y_{j,t}{}^{(\epsilon-1)/\epsilon} \right)^{\epsilon/(\epsilon-1)}, \tag{B.9}$$

which is the counterpart of the well-known Dixit-Stiglitz index with a finite amount of firms (the same holds for the price index (B.13) below). This supply constraint implies that the final-good producers have a technology that aggregates non-perfectly substitutable goods. This imperfect substitutability between all types of varieties j introduces monopolistic competition in the intermediate good market. Each good j is an imperfect substitute of degree $\epsilon > 1$, allowing intermediate firms to gain positive profits through a gap between their selling and producing prices. The intensity of the monopolistic competition is driven by $\epsilon/(\epsilon-1)$, which is the markup over the marginal costs of intermediate firms.

The optimization problem of the final-good producers reads as follows

$$\max_{\{y_{j,t}\}} P_t Y_t^D - \sum_{j=1}^{J} p_{j,t} y_{j,t} + \varrho_t \left[J^{-1/\epsilon} \sum_{j=1}^{J} y_{j,t}{}^{(\epsilon-1)/\epsilon} - \left(Y_t^D\right)^{(\epsilon-1)/\epsilon} \right].$$

The associated first-order condition of the maximization problem is given by

$$P_t - \varrho_t (\epsilon-1)/\epsilon \left(Y_t^D\right)^{-1/\epsilon} = 0. \tag{B.10}$$

Plugging B.9 into B.10 we can rewrite

$$-p_{j,t} + \varrho_t (\epsilon-1)/\epsilon J^{-1/\epsilon} y_{j,t}{}^{(-1)/\epsilon} = 0. \tag{B.11}$$

which can be rewritten as the standard CES downward-sloping demand function per firm j:

$$y_{j,t} = \frac{Y_t^D}{J} \left(\frac{p_{j,t}}{P_t} \right)^{-\epsilon} = y_t^{av} \left(\frac{p_{j,t}}{P_t} \right)^{-\epsilon}. \tag{B.12}$$

where $y_t^{av} \equiv \frac{Y_t^D}{J}$ is the average demand.[28]

[28] The introduction of this variable comes from the discretization of the population of agents. In the textbook version of the model, a continuum of firms between 0 and 1 is considered so that the average coincides with the sum of the individual variables.

The aggregate price index is given by

$$P_t = \left[\frac{1}{J}\sum_{j=1}^{J} p_{j,t}^{1-\epsilon}\right]^{1/(1-\epsilon)}. \tag{B.13}$$

Log-linearizing Eqs. (B.12) and (B.13) leads to the following expressions:

$$\hat{y}_{j,t} = \hat{y}_t^{av} - \epsilon\left(\hat{p}_{j,t} - \hat{P}_t\right), \tag{B.14}$$

$$\hat{P}_t = \frac{1}{J}\sum_{j=1}^{J} \hat{p}_{j,t}. \tag{B.15}$$

Expressing Eq. (B.15) in growth rates provides the expression for the inflation rate:

$$\hat{\pi}_t = \frac{1}{J}\sum_{j=1}^{J} \hat{\pi}_{j,t} \tag{B.16}$$

Intermediate sector. Each firm j in the intermediate sector has a linear production technology:

$$y_{j,t} = h_{j,t}^d, \tag{B.17}$$

where $y_{j,t}$ is their production and $h_{j,t}$ is their labor input. Intermediate-goods producers solve a two-stage problem. In the first stage, taken the labor price w_t as given, firms hire labor $h_{j,t}^d$ in a perfectly competitive labor market in order to minimize their costs subject to the production constraint (B.17).

Stage 1: The first stage can be expressed as a profit-maximization problem:

$$\max_{\{y_{j,t}h_{j,t}^d\}} mc_{j,t}y_{j,t} - w_t h_{j,t}^d + \lambda_t\left[h_{j,t}^d - y_{j,t}\right], \tag{B.18}$$

where $mc_{j,t}$ denotes the real marginal cost of producing one additional good. The first-order condition leads to the expression of the real marginal cost:

$$mc_{j,t} = mc_t = w_t. \tag{B.19}$$

Because households exhibit the same labor productivity, all firms hire households at the same wage rate w_t. Once the firms have determined their marginal cost, the next step is to determine their mark-up over this marginal cost mc_t from the imperfect substitution of the good varieties.

Stage 2: In the second-stage problem, the firms operate under a Rotemberg (1982) price-setting mechanism. We define the Rotemberg price adjustment cost by:

$$\Theta_{j,t} = \frac{\xi'}{2}\left(\frac{p_{j,t}}{p_{j,t-1}} - \bar{\pi}\right)^2 y_t^{av}, \tag{B.20}$$

where $\xi' > 0$ is the price stickiness parameter and $\bar{\pi}$ is the CB target.

The profit maximization becomes dynamic because of the adjustment costs over prices that span over two periods. In a monopolistic-competition setting, firms face the individual demand for goods given by Eq. (B.12). The problem faced by firms is then given by[29]

$$\max_{\{p_{j,t}\}} \mathbb{E}_{j,t}^* \sum_{\tau=0}^{\infty} \Lambda_{j,t,t+\tau}\left(\left(1-\iota_{j,t+\tau}\right) y_{j,t+\tau}\frac{p_{j,t+\tau}}{P_{t+\tau}} - e^{\varsigma_u u_{t+\tau}} mc_{t+\tau}y_{j,t+\tau} - \Theta_{j,t+\tau}\right), \tag{B.21}$$

[29] The menu costs setup make it possible to bypass the infinite sum and any concerns regarding the use of the law of iterated projection with boundedly rational expectations because the FOC involves only t and $t+1$.

where $\Lambda_{j,t,t+\tau} = \beta^{\tau}\lambda'_{t+\tau}/\lambda'_t$ corresponds to the discount factor of agent j as previously defined in Eq. (B.3); $p_{j,t}$ is the individual price set by firm j, P_t is the aggregate price which sets the problem of firms in real terms. Variable u_t is an exogenous cost-push shock that captures exogenous changes in the cost structure of the firms. Parameter ς_u normalizes to one the parameter on this shock in the log-linearized form of the aggregate-supply equation (see Eq. (B.44) below). Note that the tax rate on the added value, $\iota_{j,t}$, is typically used in the NK literature to offset some market distortions and simplify the analysis of optimal policy. In the present case, given the presence of heterogeneity with respect to the benchmark textbook model, we assume that this tax is set by the government so as to offset the effect of the relative price dispersion on profits, i.e. $\iota_{j,t} = \frac{p_{j,t}-P_t}{p_{j,t}}$.

Plugging in the demand function (B.12), the objective function of the firms becomes

$$\max_{\{p_{j,t}\}} \mathbb{E}^*_{j,t} \sum_{\tau=0}^{\infty} \Lambda_{j,t,t+\tau} \left(\left(1-\iota_{j,t+\tau}\right) \left(\frac{p_{j,t+\tau}}{P_{t+\tau}}\right)^{1-\epsilon} y^{av}_{t+\tau} - e^{\varsigma_u u_{t+\tau}} mc_{t+\tau} \left(\frac{p_{j,t+\tau}}{P_{t+\tau}}\right)^{-\epsilon} y^{av}_{t+\tau} - \Theta_{j,t+\tau} \right). \tag{B.22}$$

The first-order condition reads as:

$$\begin{aligned} \left(1-\iota_{j,t}\right) \frac{(1-\epsilon)}{p_{j,t}} \left(\frac{p_{j,t}}{P_t}\right)^{1-\epsilon} y^{av}_t + e^{\varsigma_u u_t} \epsilon \frac{mc_t}{p_{j,t}} \left(\frac{p_{j,t}}{P_t}\right)^{-\epsilon} y^{av}_t - \frac{\xi'}{p_{j,t-1}} \left(\frac{p_{j,t}}{p_{j,t-1}} - \bar{\pi}\right) y^{av}_t \\ + \mathbb{E}^*_{j,t} \Lambda_{j,t,t+1} \frac{p_{j,t+1}}{p^2_{j,t}} \xi' \left(\frac{p_{j,t+1}}{p_{j,t}} - \bar{\pi}\right) y^{av}_{t+1} = 0. \end{aligned} \tag{B.23}$$

Log-linearizing this expression, assuming symmetry among firms at the steady state and using the definition of the tax rate, we can write

$$(\epsilon-1)\,\hat{p}_{j,t} + (\epsilon-1)/\epsilon \left(\varsigma_u u_t + \widehat{mc}_t - \epsilon \hat{p}_{j,t}\right) = \frac{\xi'}{\epsilon} \bar{\pi} \hat{\pi}_{j,t} - \frac{\xi'}{\epsilon} \bar{\pi} \beta \mathbb{E}^*_{j,t} \hat{\pi}_{j,t+1}, \tag{B.24}$$

with $\bar{mc} = (\epsilon-1)/\epsilon$.

Rearranging terms leads to the final inflation equation for each producer j:

$$\hat{\pi}_{j,t} = (\epsilon-1) \frac{\varphi}{\bar{\pi}\xi'} \hat{h}_{j,t} + \beta \mathbb{E}^*_{j,t} \hat{\pi}_{j,t+1} + u_t. \tag{B.25}$$

Note that setting $\varsigma_u \equiv \xi'\bar{\pi}/(\epsilon-1)$ normalizes the shock in the linear equation because the marginal cost is the same across firms, $\widehat{mc}_t = \hat{w}_t$, and across households, $\hat{w}_t = \varphi \hat{h}_{jt}$ (from the wage-setting equation (B.4)).

B.3 Authorities

Monetary policy. The monetary policy authority, namely the CB, sets the nominal interest rate i as a function of the deviation of inflation from its target and of the output gap. We use here a general formulation of the monetary policy that allows for history-dependent inflation targets.

Precisely, the monetary policy rule reads as:

$$\frac{i_t}{\bar{\imath}} = \left(\frac{i_{t-1}}{\bar{\imath}}\right)^{\rho} \times \left(\left(\frac{\pi_t}{\bar{\pi}}\right)^{\phi_\pi} \times \left(\frac{Y^D_t}{\bar{Y}^D}\right)^{\phi_y} \right)^{1-\rho} \times \exp(v_t), \tag{B.26}$$

where parameters $\{\phi_\pi, \phi_y\}$ are the two reaction coefficients to, respectively, inflation and output gap, ρ the smoothing coefficient, v_t a monetary policy shocks and $\bar{\pi}$ the CB's target.

The log-linearization of the monetary policy rule yields:

$$\hat{\imath}_t = \rho \hat{\imath}_{t-1} + \left(1-\rho\right)\left(\phi_\pi \hat{\pi}_t + \phi_y \hat{y}_t\right) + v_t. \tag{B.27}$$

Government. The government implements the tax $\iota_{j,t}$ on the added-value of firms and borrows B_t from the households, while the expenditure side includes interest payments and lump-sum transfers (where it is assumed that all household members receive the same amount). The budget constraint of the government is then:

$$\sum_{j=1}^{J} \iota_{j,t} y_{jt} + B_t = \sum_{j=1}^{J} T_{jt} + B_{t-1} i_{t-1} / \pi_t. \tag{B.28}$$

B.4 Equilibrium conditions

Intermediate sector. Assuming a symmetric equilibrium in the intermediate-good market we can write

$$\sum_{j=1}^{J} y_{j,t} = \sum_{j=1}^{J} y_t^{av} \left(\frac{p_{j,t}}{P_t} \right)^{-\epsilon}, \tag{B.29}$$

Linearizing and using the definition in Eq. (B.15) which rules out the effect of price dispersion on production, we obtain

$$\sum_{j=1}^{J} \hat{y}_{j,t} = \hat{Y}_t^D, \tag{B.30}$$

where $J \times \bar{y} = \bar{Y}^D$.

Final goods sector. The resource constraint is given by:

$$Y_t^D = \sum_{j=1}^{J} \left(c_{j,t} + \frac{\xi'}{2} \left(\pi_{j,t} - \bar{\pi} \right)^2 y_t^{av} \right), \tag{B.31}$$

where the second term on the right-hand side corresponds to the price adjustment costs.

Log-linearizing Eq. (B.31) yields:

$$\hat{Y}_t^D = \frac{1}{J} \sum_{j=1}^{J} \hat{c}_{j,t}, \tag{B.32}$$

with $\bar{c} = \bar{y} = \bar{Y}^D / J$.

Labor market. Equilibrium in the labor market is reached when the aggregate labor demand from firms satisfies:

$$\sum_{j=1}^{J} h_{j,t}^d = \sum_{j=1}^{J} h_{j,t}, \tag{B.33}$$

or, after log-linearization:

$$\sum_{j=1}^{J} \hat{h}_{j,t}^d = \sum_{j=1}^{J} \hat{h}_{j,t}, \tag{B.34}$$

because firms and households share the same steady-state values for the hours worked.

B.5 Aggregation

Labor demand dispersion. In the presence of non-separable preferences, the labor supply in Eq. (B.4) is the same across households:

$$\hat{h}_{j,t} = \hat{h}_t. \tag{B.35}$$

Combining Eqs. (B.17) and (B.35) with the firm-specific demand for intermediate inputs (B.14) at the symmetric equilibrium leads to the following expression

$$\hat{h}^d_{j,t} = \hat{y}^{av}_t - \epsilon\left(\hat{p}_{j,t} - \hat{P}_t\right) \tag{B.36}$$

One can note that pricing strategies that are different from the average, i.e. $\hat{p}_{j,t} \neq \hat{P}_t$, imply a dispersion in both labor demands and outputs across firms. This dispersion in labor demand needs to be matched with the homogeneous labor supplies from the households in Eq. (B.35). To do so, we use the following assumption:

Assumption B.1 *To map potentially heterogeneous labor demands with homogeneous labor supplies, households evenly split their working hours across all firms at no cost, which translates into:*

$$h_{j,t} = \sum_{j=1}^{J} \frac{h^d_{j,t}}{J}. \tag{B.37}$$

Log-linearizing Eq. (B.37) gives

$$\hat{h}_{j,t} = \sum_{j=1}^{J} \frac{\hat{h}^d_{j,t}}{J},$$

which shows that Assumption B.1 ensures that the equilibrium in the labor market given by Eq. (B.34) holds even in the presence of heterogeneous labor demands.

Aggregate demand. Consider the agent-level log-linearized Euler equation (B.7) and aggregate across all household members as:

$$\sum_{j=1}^{J}\left[\bar{c}_j\hat{c}_{j,t} - \chi\bar{h}^{1+\varphi}\hat{h}_{j,t}\right] = \sum_{j=1}^{J}\left[-\frac{\vartheta}{\sigma'}\left[\hat{\imath}_t - \mathbb{E}^*_{j,t}\hat{\pi}_{j,t+1}\right] + \mathbb{E}_t\left(\bar{c}_j\hat{c}_{j,t+1} - \chi\bar{h}^{1+\varphi}\hat{h}_{j,t+1}\right) + \frac{\vartheta\varsigma}{\sigma'}g_t\right],$$

which may be rearranged into:

$$\bar{c}_j\hat{c}_t - \chi\bar{h}^{1+\varphi}\hat{h}_t = -\frac{\vartheta}{\sigma'}\hat{\imath}_t + \mathbb{E}_t\left(\bar{c}_j\hat{c}_{t+1} - \chi\bar{h}^{1+\varphi}\hat{h}_{t+1}\right) + \mathbb{E}^*_t\left(\frac{\vartheta}{\sigma'}\hat{\pi}_{t+1}\right) + \frac{\vartheta\varsigma}{\sigma'}g_t,$$

where $\mathbb{E}^*_t\hat{\pi}_{t+1} = \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}^*_{j,t}\hat{\pi}_{j,t+1}$, with $\mathbb{E}^*_{j,t}\hat{\pi}_{t+1} \equiv \mathbb{E}^{SL}_{j,t}\hat{\pi}_{t+1}$ and $\mathbb{E}^*_t\hat{\pi}_{t+1} \equiv \mathbb{E}^{SL}_t\hat{\pi}_{t+1}$, and output gap expectations are assumed to be identical across all agents and, in particular, rational (while taking into account the presence of subjective inflation expectations). The learning shock φ_t under SL is implicit to the aggregate inflation expectations $\mathbb{E}^*_t\hat{\pi}_{t+1}$.

Equilibrium in labor market allows one to write $\hat{h}_t = \hat{h}^d_t = \hat{y}_t$, while equilibrium in the intermediate-good market entails $\hat{y}_t = \hat{y}^D_t = \hat{c}_t$. The previous condition reads as:

$$\left(\bar{c} - \chi\bar{h}^{1+\varphi}\right)\hat{y}_t = -\frac{\vartheta}{\sigma'}\hat{\imath}_t + \mathbb{E}_t\left(\left(\bar{c} - \chi\bar{h}^{1+\varphi}\right)\hat{y}_{t+1}\right) + \mathbb{E}^*_{j,t}\left(\frac{\vartheta}{\sigma'}\hat{\pi}_{t+1}\right) + \frac{\vartheta\varsigma}{\sigma'}g_t,$$

Recall that at the steady state, the hours worked are normalized to one, thus $\bar{c} - \chi\bar{h}^{1+\varphi} = 1 - \chi$ and $\vartheta = 1 - \chi/1 + \varphi$. Recall also that $\varsigma = \sigma'\frac{(1-\chi)}{\vartheta}$. Hence, the aggregate Euler equation in a compact form reads as:

$$(1-\chi)\hat{y}_t = -\frac{\vartheta}{\sigma'}\hat{\imath}_t + \mathbb{E}_t\left((1-\chi)\hat{y}_{t+1}\right) + \mathbb{E}^*_t\left(\frac{\vartheta}{\sigma'}\hat{\pi}_{t+1}\right) + (1-\chi)g_t. \tag{B.38}$$

Aggregate supply. From the micro-level NK Phillips curve (B.25), we may aggregate over all firms j:

$$\sum_{j=1}^{J} \hat{\pi}_{jt} = \sum_{j=1}^{J} \left[(\epsilon - 1) \frac{\varphi}{\bar{\pi}\xi'} \hat{h}_{j,t} + \beta \mathbb{E}^*_{j,t} \hat{\pi}_{jt+1} + u_t \right],$$

which becomes:

$$\hat{\pi}_t = (\epsilon - 1) \frac{\varphi}{\bar{\pi}\xi'} \hat{h}_t + \beta \mathbb{E}^*_t \hat{\pi}_{t+1} + u_t. \tag{B.39}$$

Note that we can replace aggregate labor with aggregate output as in the usual textbook formulation, which results in:

$$\hat{\pi}_t = (\epsilon - 1) \frac{\varphi}{\bar{\pi}\xi'} \hat{y}_t + \beta \mathbb{E}^*_t \hat{\pi}_{t+1} + u_t. \tag{B.40}$$

The monetary policy rule (B.27) closes the model.

B.6 Convergence between separable and non-separable utilities

In this section, we show which restrictions on the parameters of the non-separable utility function allow the model to correspond to the one derived from separable preferences and reconcile the form of our three-equation NK model with heterogeneous expectations with the usual textbook formulation of NK models under RE (see, e.g., Galí 2015, Chap. 3).

Marginal utility of consumption. Let λ_t and λ'_t denote respectively the marginal utility of consumption under separable and non-separable preferences respectively. These are given by:

$$\hat{\lambda}_t = -\sigma \hat{c}_t \ \text{ and } \ \hat{\lambda}'_t = -\sigma' \left(\frac{1-\chi}{1-\chi/\left(1+\varphi\right)} \right) \hat{c}_t$$

Imposing $\hat{\lambda}_t = \hat{\lambda}'_t$ results in the condition on σ' under which both models exhibit the same marginal utilities of consumption:

$$\sigma' = \frac{1-\chi/\left(1+\varphi\right)}{1-\chi} \sigma. \tag{B.41}$$

We may substitute σ' by σ into (B.38) as follows:

$$\hat{y}_t = \mathbb{E}_t \left\{ \hat{y}_{t+1} \right\} - \frac{1}{\sigma} \mathbb{E}^*_t \left\{ \hat{\imath}_t - \hat{\pi}_{t+1} \right\} + g_t. \tag{B.42}$$

Slope of the New Keynesian Phillips Curve. Nonseparability in utility also affects the real-wage setting, and thus the marginal cost and the slope of the NK Phillips curve:

$$\kappa = (\epsilon - 1) \frac{\left(\sigma + \varphi\right)}{\bar{\pi}\xi} \text{ and } \kappa' = (\epsilon - 1) \frac{\varphi}{\bar{\pi}\xi'}.$$

By imposing $\kappa = \kappa'$, we derive ξ':

$$\xi' = \frac{\varphi}{\left(\sigma + \varphi\right)} \xi. \tag{B.43}$$

Under this second condition, the aggregate supply curve may be rewritten as follows:

$$\hat{\pi}_t = \kappa \hat{y}_t + \beta \mathbb{E}^*_t \hat{\pi}_{t+1} + u_t \tag{B.44}$$

with $\kappa = (\epsilon - 1) \frac{(\sigma+\varphi)}{\bar{\pi}\xi}$.

These parameter values are only used in the micro-founded welfare analysis (see Section B.7 hereafter), where we assume $\epsilon = 6$ and $\varphi = 2$, which is in the realm of the literature.

B.7 Approximating the welfare

In this section, we discuss the approximation of the welfare criterion to measure the cumulative utility loss or gains across scenarios.

B.8 Derivations

Microfounded measure The welfare function is the discounted sum of the average utility across family members:

$$\mathscr{W}_t = \frac{1}{j}\sum_{j=1}^{J} U_{j,t} + \beta \mathbb{E}_t^* \mathscr{W}_{t+1}, \tag{B.45}$$

where the individual utility function of each household member j reads as:

$$U_{j,t} = \frac{1}{1-\sigma'}\left(c_{j,t} - \chi \frac{h_{j,t}^{1+\varphi}}{1+\varphi}\right)^{1-\sigma\prime}.$$

or plugging for hours worked and ressource constraints we have

$$U_{j,t} = \frac{1}{1-\sigma'}\left(c_{j,t} - \chi \frac{\left(\frac{1}{J}\sum_{j=1}^{J}\left(c_{j,t} + \frac{\xi'}{2}\left(\pi_{j,t} - \bar{\pi}\right)^2 y_t^{av}\right)\left(\frac{1}{J}\sum_{j=1}^{J}\left(\frac{p_{j,t}}{P_t}\right)^{-\epsilon}\right)\right)^{1+\varphi}}{1+\varphi}\right)^{1-\sigma\prime}.$$

with

$$\begin{aligned} P_t &= \pi_t P_{t-1}, \\ p_{j,t} &= \pi_{j,t} p_{j,t-1}. \end{aligned}$$

Recalling that agents hold homogeneous (rational) expectations regarding future output and individual consumption levels so that $\mathbb{E}_{j,t}(\hat{c}_{j,t+1}) = \mathbb{E}_t(\hat{y}_{t+1})$, $\forall j, t$, we can use the microfoundations to write the agent-level inflation and consumption levels:

$$\begin{aligned} \hat{c}_{j,t} &= \mathbb{E}_t\left(\hat{y}_{t+1}\right) - \frac{1}{\sigma}\left(\hat{\imath}_t - \mathbb{E}_{j,t}^{SL}\hat{\pi}_{t+1}\right) + g_t, \\ \hat{\pi}_{j,t} &= \kappa \hat{y}_t + \beta \mathbb{E}_{j,t}^{SL}\hat{\pi}_{t+1} + u_t. \end{aligned}$$

By definition, we have:

$$\begin{aligned} c_{j,t} &\approx \overline{c}_j + \overline{c}_j \times \hat{c}_{j,t}, \\ y_t^{av} &\approx \overline{Y}/J + \overline{Y}/J \times \hat{y}_t, \\ \pi_t &\approx \overline{\pi} + \overline{\pi} \times \hat{\pi}_t, \\ \pi_{j,t} &\approx \overline{\pi} + \overline{\pi} \times \hat{\pi}_{j,t}. \end{aligned}$$

Because we focus on counterfactual scenarios happening at the very end of our sample (i.e. over the last 12 quarters), it is impossible and inadvisable to compute an unconditional welfare measure because it would require to extract asymptotic simulated moments under different policy regimes and insert them into a second-order approximation of Equation (B.45); see Arifovic et al. (2024) for such an approach. We thus use the average cumulative sum of the utility *conditional* on the path of aggregate and idiosyncratic variables between two periods $[t_1 : t_2]$ as

$$\mathscr{U}_{t_1,t_2} = \sum_{h=t_1}^{t_2}\left(\frac{1}{J}\sum_{j=1}^{J} U_{j,h}\right). \tag{B.46}$$

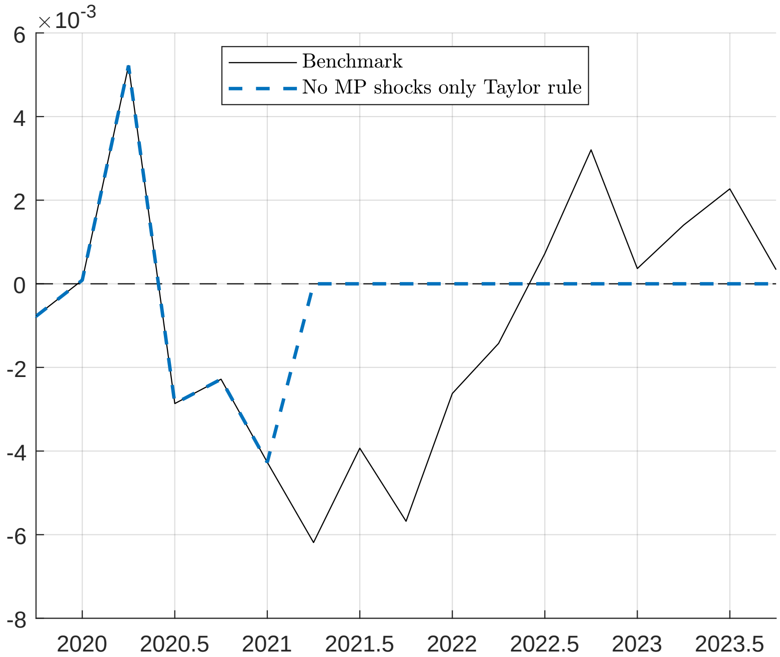
This indicator takes into account the cost implied by prices and consumption dispersion as a result of the heterogeneity in inflation expectations and quantifies the preference between inflation and output stabilization. To avoid history-dependent effects and focus on the sub-sample $[t_1, t_2]$, we normalize all prices at $P_{t_1} = p_{j,t_1} = 1 \ \forall j \in J$. Moreover, it is important to note that if the model were to be run under REE, there would not be prices and consumption dispersion. Therefore, the REE case is always welfare improving with respect to the SL model (Arifovic et al. 2024).

We also consider a more standard *ad-hoc* loss function such as

$$\mathscr{L}_{t_1,t_2} = -\gamma_\pi \sum_{h=t_1}^{t_2} (100 \times \hat{\pi}_h)^2 - \gamma_y \sum_{h=t_1}^{t_2} \left(100 \times \hat{y}_h\right)^2,$$

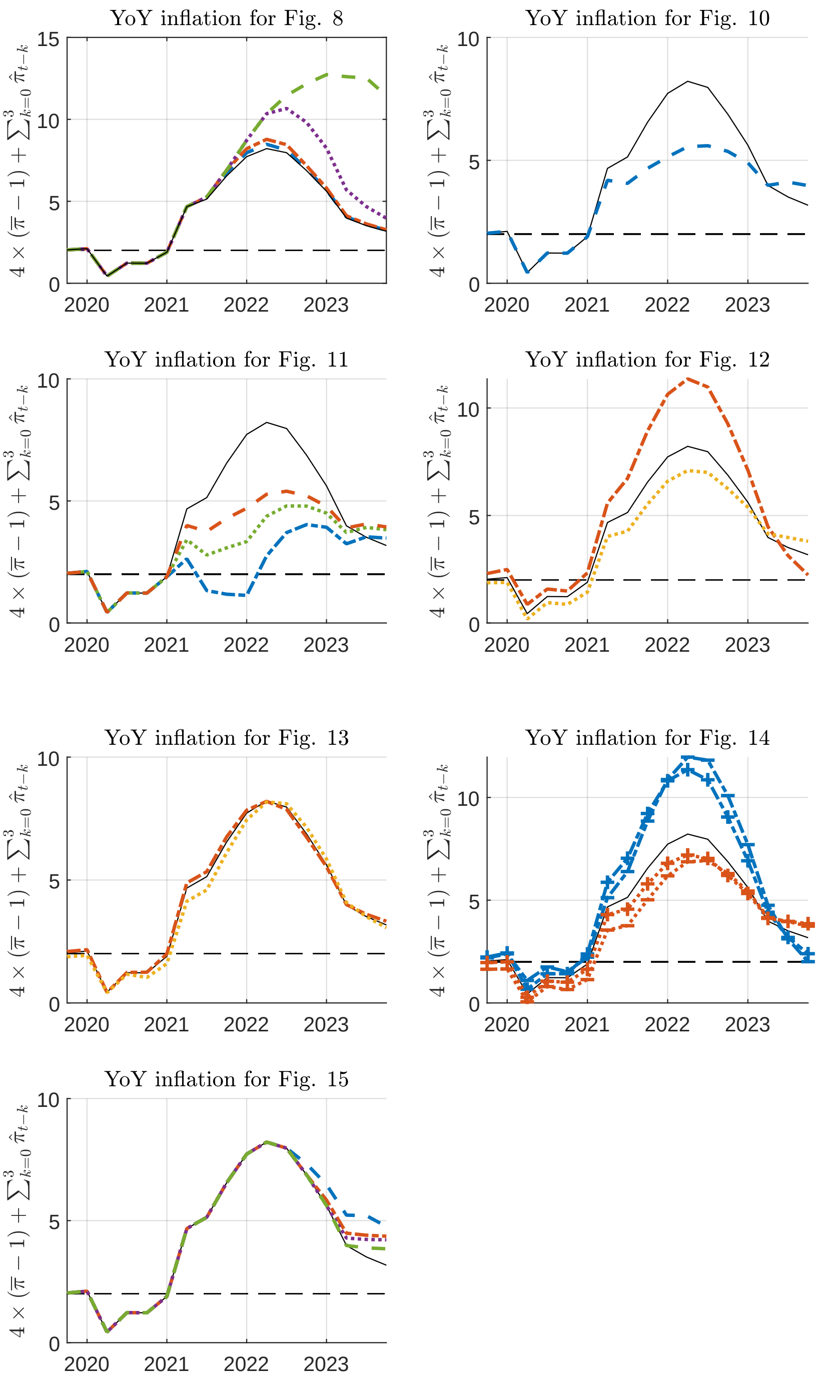
with $\gamma_\pi = 1$ and $\gamma_y = 0.25$ in line with policy institution practices and the literature such as Kiley & Roberts (2017). This function disregards the micro-foundations and, hence, the effects of the dispersion of prices and consumption values under heterogeneous expectations but quantifies the trade-off between inflation and output stabilization.

C Additional results



Notes: The solid black line depicts the monetary policy shocks, i.e. the deviation of the proxy funds rate from the one prescribed by the estimated Taylor rule. The blue dashed line represents the counterfactual scenario of Fig. 10 where the Taylor rule would have been strictly followed as of 2021Q1.

Figure C.1: Monetary policy shocks over 2019Q4-2023Q4



Notes: Note that the dashed target line is inconsistent between those level charts and the deviation charts. In the deviation charts the dashed target line is the 0 deviation from historical average. In those charts, this is 2% Fed target.

Figure C.2: Year-over-year inflation rate in the counter-factual simulations

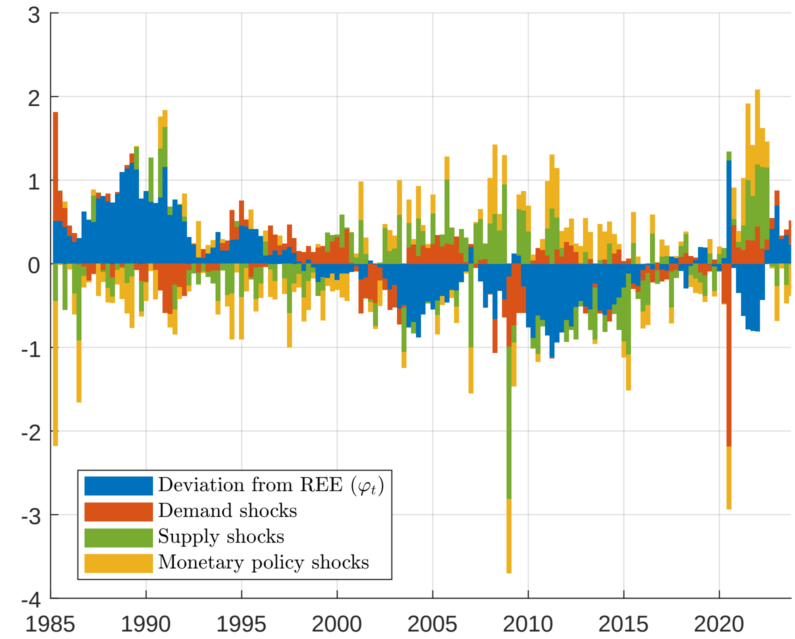


Figure C.3: Historical decomposition of the inflation rate over the whole sample

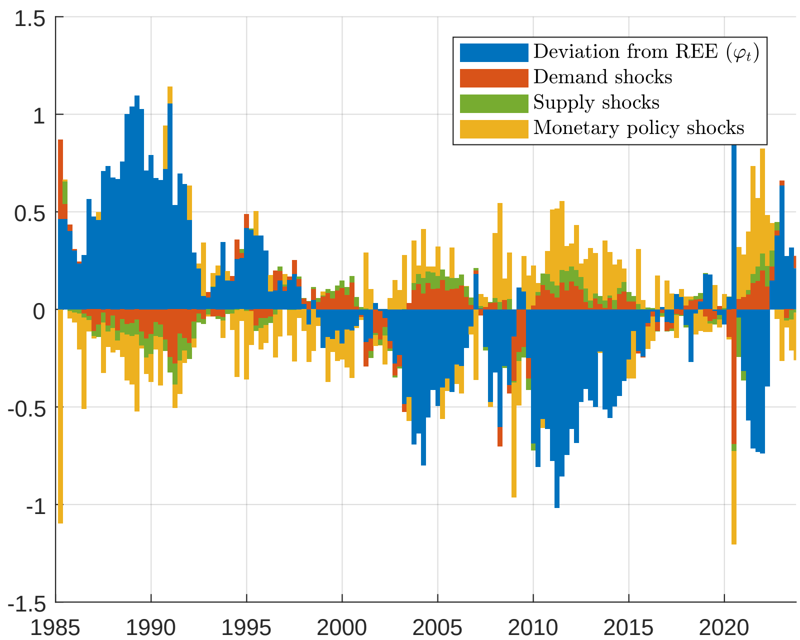


Figure C.4: Historical decomposition of one-quarter-ahead aggregate subjective inflation expectation over the whole sample

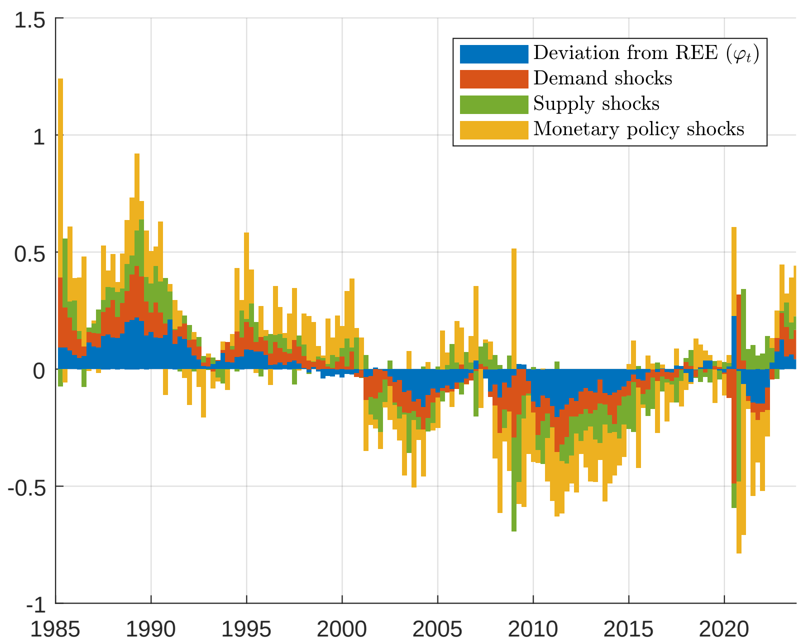


Figure C.5: Historical decomposition of the policy rate over the whole sample

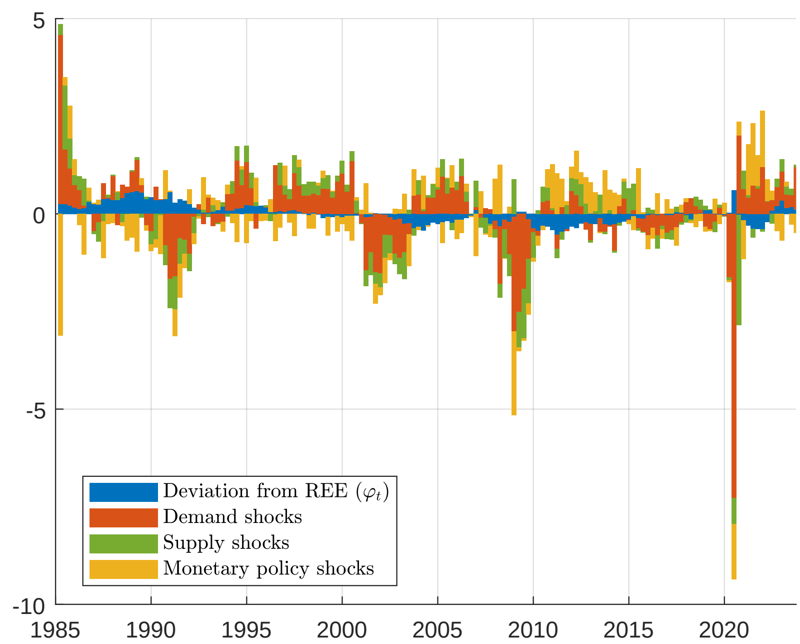
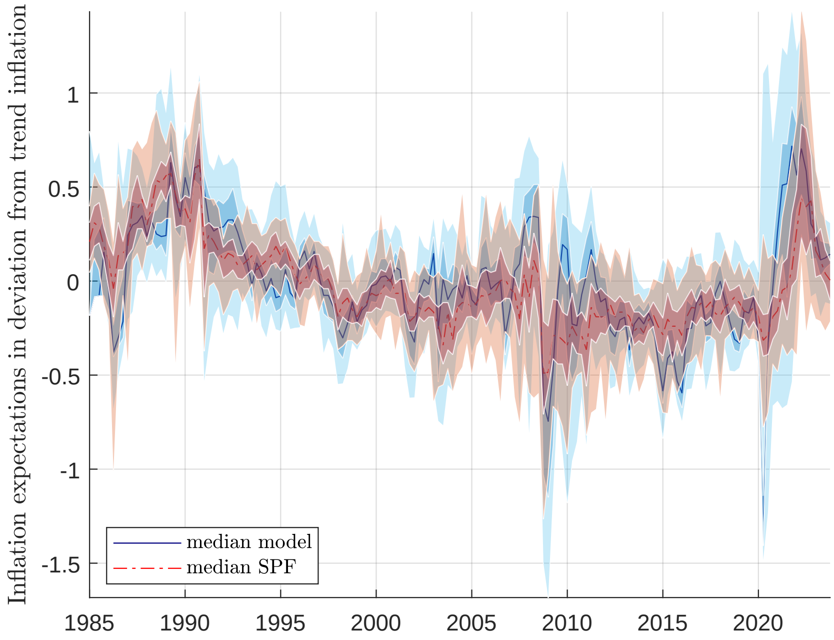
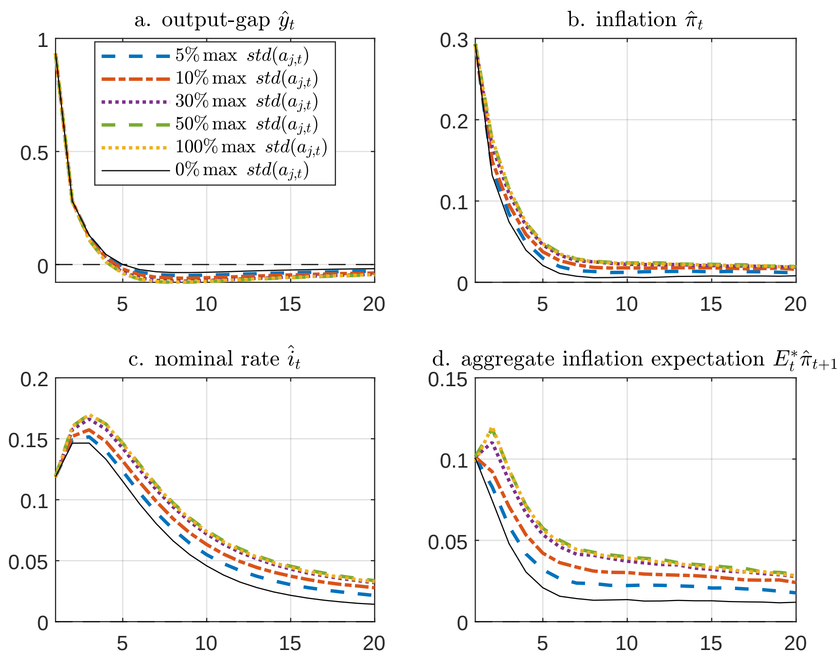


Figure C.6: Historical decomposition of the output gap



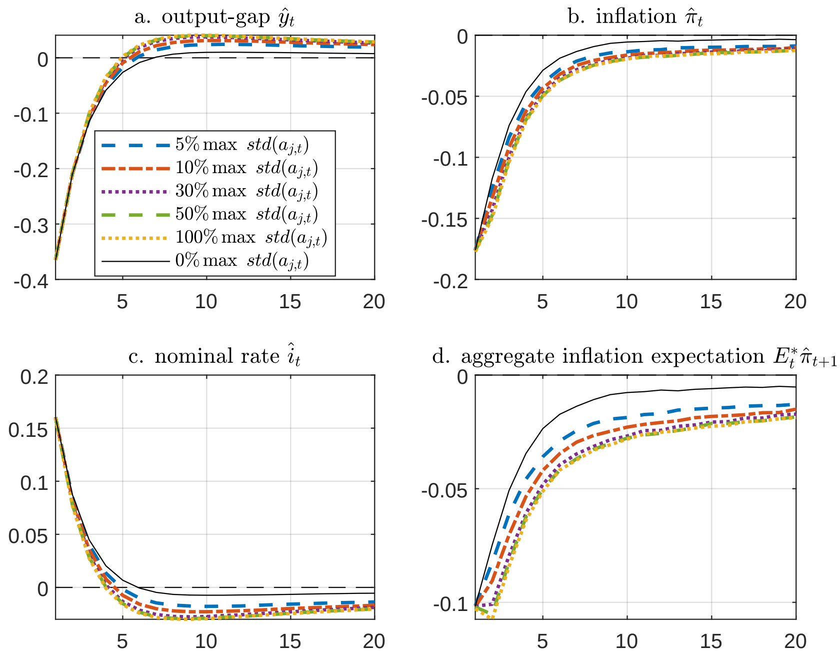
Notes: See Fig. 5.

Figure C.7: Time-varying cross-sectional dispersion of inflation expectations over the full sample



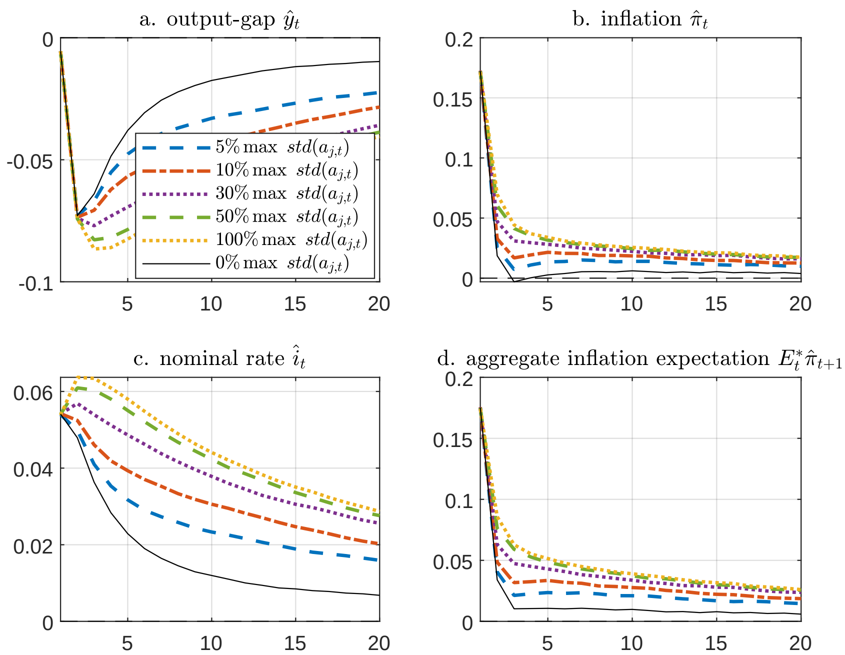
Notes: See Fig. 7.

Figure C.8: Response to a one standard deviation demand shock as a function of the heterogeneity in expectations



See Fig. 7.

Figure C.9: Response to a one standard deviation monetary policy shock as a function of the heterogeneity in expectations



Notes: See Fig. 7.

Figure C.10: Response to a one standard deviation news shock as a function of the heterogeneity in expectations