A Macroeconomy with Intuitive Thinkers

Preliminary draft

Maren Bartels[*] Johannes Binswanger[†] Albert Flak[‡] Manuel Oechslin[§]

April 13, 2025

Abstract

Expectation formation is a key element in macroeconomic models. A large literature documents that individuals' expectations are often influenced not only by rational but also by intuitive thinking. In this paper, we propose a tractable framework in which expectations are a priori formed in a rational thinking mode which, however, gets perturbed by intuitive thinking. The framework captures a range of phenomena. Our main focus is on subjective models of the economy that concern correlation patterns, in particular a subjective "stagflationary model", according to which higher inflation comes with lower output. This pattern has been frequently found in the recent empirical literature on subjective expectations. We derive impulse responses for an "intuitive thinking shock" and show how such a shock can have relatively persistent effects that resemble a supply shock, with no actual supply shock involved. We contribute to the development of tractable alternatives to fully rational expectations, and to the study of how non-fundamental factors can influence fluctuations in macroeconomic series.

JEL classification: E71, E31

Keywords: Behavioral macro, expectations, dual system processing, impulse responses

[*] University of Lucerne, Department of Economics, Switzerland. Email: maren.bartels@unilu.ch.
[†] University of St. Gallen, Department of Economics, Switzerland. Email: johannes.binswanger@unisg.ch.
[‡] University of St. Gallen, Department of Economics, Switzerland. Email: albert.flak@unisg.ch.
[§] University of Lucerne, Department of Economics, Switzerland. Email: manuel.oechslin@unilu.ch.

1 Introduction

Expectation formation is a key element in macroeconomic models. Typically, equilibrium conditions of DSGE models come in the form of stochastic difference equations with expectation terms playing a key role for the solution of the models. Moreover, expectations are also prominent in individual decision problems or partial-equilibrium applications, such as consumption-saving or investment decisions. An important question is therefore how agents form such expectations. The empirical literature on this topic has found strong evidence that some typical patterns in expectation formation cannot be captured completely by full-information rational expectation (FIRE) models (Andre, Haaland, Roth, & Wohlfahrt, 2023; Andre, Pizzinelli, Roth, & Wohlfahrt, 2022; Bordalo, Gennaioli, & Shleifer, 2018; Candia, Coibion, & Gorodnichenko, 2020, 2021; Coibion & Gorodnichenko, 2012, 2015a; Dräger, Lamla, & Pfajfar, 2016). Expectations often are forward-looking to an important degree. However, expectations often also show traits of intuitive thinking that would not appear in a FIRE world. For instance, agents may (partly) form expectations according to an intuitive model of how the economy works that does not correspond to the mechanisms of a FIRE economy. We may improve our understanding of economic relationships, and also make better predictions, if we have frameworks on how to include these traits of intuitive thinking. We may also ask the important question to what degree "wrong" expectations – from an ex ante FIRE perspective – may become partly self-fulfilling ex post, and hence contribute to fluctuations in macroeconomic series.

In light of the strong merits of the FIRE model, which has helped to shed light on many macroeconomic phenomena from monetary policy to fiscal policy, it seems important to have a framework that integrates elements from both, rational as well as intuitive thinking. There is a prominent framework in the psychological literature that does exactly this: the framework of dual-system processing (Evans & Stanovich, 2013; Ilut & Valchev, 2023; Kahneman, 2011). In the words of Kahneman (2011), there is a slow thinking system that is akin to rational-expectations thinking. Next to this, there is also a fast intuitive thinking system that is active whenever assessments of a situation are made quickly, and – more important for our purposes – when assessments become very challenging for the slow rigorous thinking system. In the latter case, intuitive thinking may perturb the rigorous thinking process and lead to so-called substitution effects. These entail that the fast thinking system substitutes bits of a hard problem by an easier one that is solved in an intuitive way, often unconsciously so. As an example, it may be hard to forecast inflation, but gasoline or grocery prices are rising. The intuitive system may then infer high inflation expectations, based on an observation that may be nearly meaningless in a FIRE world (Coibion & Gorodnichenko, 2015b; D'Acunto, Malmendier, Ospina, & Weber, 2021; D'Acunto & Weber, 2024).

In this paper, we develop a tractable framework of expectation formation under dual-system processing that we refer to as dual-system expectations (DSE). The starting point for this framework is a first-order stochastic difference equation such as the recursive Blanchard-Kahn equation (Blanchard & Kahn, 1980). We assume that agents understand the associated solution form of the difference equation, which typically consists of an infinite sum of expectation terms about future fundamental states (e.g. future total factor productivity levels, or future incomes in the case of a consumption-saving model).

In practice, predicting future states of fundamentals is hard. There is an overwhelming array of signals that may potentially be relevant, and agents have to determine the weight that they assign to each signal. We adopt the view that when individuals face the task of selecting signals and determining their weights, they are a priori in a focused mode of cognition which activates respective mental networks. In the language of the dual-processing model of cognition (Kahneman, 2011; Stanovich & West, 1991), individuals activate "System 2" or "slow" thinking. However, a key finding in modern neuroscience is that there are always multiple networks active in the brain (Gratton, Sun, & Petersen, 2018; Sporns & Betzel, 2016). Thus, to some degree, intuitive thinking may still influence the outcome of focused cognition, as it is not fully suppressed by the the brain's executive control system (Gratton et al., 2018). In the words of Kahneman (2011), System 2 is a "lazy controller". As a result, when agents think about future fundamental states, the answer may, to some degree, be shaped by "sentiments", resulting from intuitive or emotional judgment that is not fully suppressed by executive control. This may include a disliking of political party currently in charge, overreaction to recent changes in economic circumstances, movements in grocery prices that, a priori, would only loosely correlate with movements in the aggregate price level, or idiosyncratic experiences by acquaintances or peers. Evidence from professional forecasters shows that even these experts are not fully capable to adequately suppress irrelevant cues. In particular, Adam, Kuang, and Xie (2024) and Bianchi, Ludvigson, and Ma (2022) show that professional forecasters overreact to private information. This observation is also consistent with a finding that, in complex environments where individuals are uncertain about the weighting of different signals, they tend to overweight weak signals and underweight strong signals (Augenblick, Lazarus, & Thaler, 2024; Kieren & Weber, 2024)

In our formal DSE framework, there are two signals that agents obtain for forming expectations about future fundamental states. Both signals may be vectors. The first signal is identical to the one that FIRE agents would take into account for forming expectations. In many application, this is the current state of fundamentals (e.g. current productivity, or current income). The second signal is sent by agents' intuitive system. We refer to intuitive signals as *sentiments*. The overall dual-system then determines the weight of sentiments, or, to what degree agents "listen to their gut". In the case of a FIRE agent, which is nested by our approach, sentiments are fully suppressed. Arguably, this is a special case that, according to the empirical literature, is hard to achieve when forming expectations in practice.

Due to the generic nature of dual-system processing, and our way of capturing intuitive thinking by a generic signal, the DSE framework is highly versatile and can capture a large range of phenomena. However, we intentionally narrow somewhat the scope of possible applications by making the assumption that sentiments follow, as their law of motion, a stationary first-order vector autoregressive (VAR) process. More precisely, this reflects the *actual* law of motion of sentiments. Since agents react to sentiments unconsciously (otherwise, they would suppress them), agents de facto behave as if sentiments followed a "perceived" law of motion, according to which the current state of sentiments were permanent. As an example, this means that if the intuitive system is pessimistic about the inflation situation ("negative inflation sentiments"), this pessimism applies with the same degree to period $t + 1$ as it applies to $t + h$ for any $h > 1$. With this, our DSE framework can encompass any phenomena of intuitive thinking that can approximately be captured by these two laws of motion, actual and perceived. As we will elaborate in the paper, this includes intuitive subjective models of the economy, but also phenomena related to associative memory (Bordalo, Conlon, Gennaioli, Kwon, & Shleifer, 2023), diagnostic expectations (Bordalo et al., 2018), and more. Importantly, thanks to the stated assumptions about the law of motion of sentiments, the building blocks of DSE are modularized. The interplay between the rigorous and intuitive mental system does not depend on the specific mechanics of how these intuitions emerge. What really matters for an analysis, say, at the macroeconomic model, is sentiments' law of motion, not their microfoundation. Thus, if a researcher desires to abstract from the inner workings of the intuitive system, sentiments can be represented by just a set of exogenous state variables. Overall, we view the DSE framework as a potentially unifying approach for modeling perturbations originating from the intuitive mind.

After developing the generic DSE framework, we apply it to study the effect of subjective models about the economy. Specifically, we are interested in the intuitive "stagflationary" model of the economy that has been found prominently in the recent empirical literature (Andre et al., 2022; Bhandari, Borovička, & Ho, 2023; Hou & Wang, 2024; Kamdar, 2019). According to this subjective model, (relatively) high inflation is associated with a (relatively) bad state of the economy. The intuitive reasoning behind sees inflation either as a form of cost-push or negative productivity shock, which explains an association of inflation with low output growth. Alternatively, high inflation may be seen as inflicting pain on consumers, which then cut back on spending, which in turn also leads to low output growth (according to intuitive reasoning). We show how this reasoning can be captured using the DSE framework, which we integrate into the standard New Keynesian model of Galí (2015). We analyze the effect of the intuitive stagflationary model on equilibrium outcomes. Moreover, based on data from the Michigan Survey of Consumers, and on the structure of the New Keynesian model, we reconstruct the sentiment series and derive impulse responses for a sentiment shock. With standard parameter values, and relying on the DSE data, such a shock has relatively persistent effects. For our main specification, upon impact, output decreases by roughly 1 percentage point, whereas inflation increases by 0.6 percentage points. The cumulative effect over the first year is -2.72 percentage points for output, and 2.11 percentage points for inflation.

In the last main section of the paper, we illustrate a specific microfoundation for the subjective stagflationary model captured by sentiments. The model is based on associative memory and diagnostic expectations (Bordalo et al., 2023, 2018). According to this model, when individuals obtain a cue that makes them think about (high) inflation, associative memory brings to mind other economic events that are assessed as typical – or rather stereotypical – for (high) inflation. Such stereotypical associations are characterized by what we call a high diagnostic value. We show that in historical data, low output growth is in fact characterized by a high diagnostic value, and thus stereotypical for high inflation. The high diagnostic value means that for high inflation, the frequency of low output growth outcomes is particularly high, *relative* to a situation where inflation is not high. Low output growth may not happen very frequently in either inflation situation scenario. However, what matters for diagnostic expectations is whether low output growth are relatively more frequent in high-inflation states than in other states. We show that diagnostic expectations can explain the negative correlation between empirical output and inflation expectations in a rather straightforward way. To our knowledge, we are the first to show this diagnostic pattern in historical output and inflation data, and to link them to subjective expectations.

The rest of this paper is organized as follows. The next section discusses the related literature. Section 3 presents motivating evidence for the importance of studying intuitive thinking components in expectations, with a focus on documenting traces of a subjective economic model that exhibits a "stagflationary" pattern. Section 4 presents the DSE framework. Section 5 discusses a set of assumptions that guarantee that the results from the previous section indeed hold when DSE are directly applied to equilibrium conditions. Section 6 embeds the framework into a simple New Keynesian model, with a focus on a stagflationary subjective model of the economy. Section 7 provides a microfoundation for the stagflationary model based on diagnostic expectations. Section 8 concludes.

2 Literature

Deviations from the FIRE benchmark are widespread. People generally underappreciate publicly available information when forming expectations and overweigh private information (Adam et al., 2024; Bianchi et al., 2022). Even experts deviate from FIRE when making forecasts, either under-reacting (Coibion & Gorodnichenko, 2012, 2015a) or over-reacting (Bordalo, Gennaioli, Ma, & Shleifer, 2020) to news about fundamentals. Households, in turn, often hold expectations for macroeconomic variables, and the relationships among them, that conflict with experts' expectations and the macroeconomic record, as a growing body of survey-based research shows (e.g., Bhandari et al., 2023; Candia et al., 2020; Dräger et al., 2016; Kamdar, 2019). Sometimes, as Candia et al. (2020) note, "it's as if the Phillips curve perceived by households is upward-sloping." (p. 9). Focusing on autonomous innovations in expectations estimated via a VAR, we similarly find traces of a stagflationary subjective model of the economy among US households, even after taking account of TFP, oil price, and IST (investment-specific technology) shocks. We further relate to the literature on the stagflationary subjective model by demonstrating how this can be captured by our DSE framework, and how it may contribute to explain patterns of fluctuations in macroeconomic series.

Ilut and Valchev (2023) provide a seminal application of dual-system processing for consumption and savings decisions. In their model, the agent initially only has a vague prior about the policy function relating optimal consumption levels to income states. The intuitive system provides a default decision modeled as a prior belief that is available based on experience from the past without any effortful thinking. If this prior features high uncertainty, the agent finds it worthwhile to engage in rigorous thinking and so obtaining a posterior belief that is more informative. These dynamics have the effect that if agents encounter a state that they have already experienced multiple times, they feel familiar with that state and stop engaging in rigorous thinking. This can happen even if the consumption choice suggested by the intuitive system deviates strongly from the optimal choice of a FIRE agent. Ilut and Valchev (2023) refer to this outcome as a learning trap. By contrast, our dual-system model is not a learning framework. We take the stance that many of the decisions underlying macroeconomic series, such as consumption-saving decisions, are important enough that individuals are motivated to achieve a good outcome. As such, the rational system may always be involved to some degree. However, agents may not manage to fully suppress "sentiment signals" sent by the intuitive system (i.e. their gut feelings), and may hence react to them. We view these sentiment signals as contributors to macroeconomic fluctuations.

Kamdar (2019) uses the rational inattention framework to offer a potential explanation for how completely rational consumers could arrive at such a stagflationary view. The paper demonstrates that when there are information costs of intermediate size, rationally inattentive consumers optimally gather information in such a way that the covariance of the posterior beliefs about labor market slackness and the price level is positive, even though the true covariance is modeled as zero. This result is intriguing; however, it is obtained for a rather specific combination of technical assumptions regarding the nature of the signals that can be obtained, the nature of the shocks to labor market slackness and the price level, and the utility function of the consumers. Moreover, the desirable feature that the rational inattention model approaches the FIRE-benchmark as information costs go to zero is only obtained if there is a correspondence between the number of choice variables and the number of unknown states. It is thus unclear whether the rational inattention model is robust enough to generate a stagflationary view (and to have the

FIRE-benchmark as a limiting case) across a wide range of linear difference models à la Blanchard and Kahn (1980).

Hou and Wang (2024) provide a very insightful analysis of the stagflationary pattern in expectation data from an empirical perspective. They derive a structural test that discriminates between noisy information vs. a deeper subjective model of the economy as the origin of the stagflationary pattern. Their analysis rejects noisy information and strongly supports the subjective model channel. Moreover, Hou and Wang (2024) also shed light on the causal pathways of reasoning that underlies subjective models by studying how invididual expectations react to news. They find that when individuals are exposed to news about inflation, this affects their expectations about inflation and also unemployment. However, when they hear news about unemployment (independent of news about inflation), inflation expectations do not react. They find evidence for a similar pathway in association patterns of news topics. We rely on these findings in our identification of an inflation sentiment shock.

Binetti, Nuzzi, and Stantcheva (2024) elicit subjective models and reasoning about inflation in a conjoint experiment. They find that the perceived causal pathways of subjective models substantially differ from a FIRE economy. Stantcheva (2024) analysis survey data on the perception of inflation and its nexus to other economic variables. The study provides evidence that, among macroeconomic variables, inflation capture individuals' primary attention and comes with a high negative emotional load. Furthermore, individuals consider inflation as causal for many other unfavorable outcomes. Again, this finding is important as a basis for our identification of an inflation sentiment shock.

Our focus on a subjective stagflationary model also connects us with recent research on narratives, which are understood as causal models mapping actions to consequences (Eliaz & Spiegler, 2020) or causal accounts for why an economic event occurred (Andre et al., 2023). Most notably, combining survey-based measurement of narratives with experimental treatments, Andre et al. (2023) explore narratives prevalent among US households regarding the surge in inflation during 2021-22 period. While there is substantial heterogeneity, narratives typically center around supply-side factors, thereby neglecting the demand side.

As shocks to intuitive thinking belong to the broader context of belief shocks, our analysis also relates to the recent empirical results in the literature on possible effects of belief shocks on macroeconomic variables. Enders, Kleemann, and Müller (2021) extract belief shocks from nowcast errors by professional forecasters. They show that a favorable (adverse) belief shock regarding future output growth causes a subsequent rise (decline) in economic activity. Earlier contributions to the literature on the effects of belief shocks produced mixed results. Using a DSGE model for estimation, Barsky and Sims (2012) do not find that noise in beliefs about future productivity growth—labeled "animal spirits"—has a sizable impact on macroeconomic variables. This contrasts with Lorenzoni (2009), a theoretical study that takes a somewhat different approach to modeling noise in beliefs. According to the simulations presented there, belief shocks, by affecting aggregate demand, can generate sizable short-run effects when it comes to output, employment, and inflation. Angeletos, Collard, and Dellas (2018) introduce autonomous innovations in higher-order beliefs, a modification that allows waves of optimism or pessimism to produce autonomous deviations from FIRE in consumers' and firms' expectations. Bhandari et al. (2023) consider fluctuations in beliefs that arise from concerns about model misspecification.

Our microfoundation of the the staflationary subjective model of the economy builds upon Bordalo, Gennaioli, and Shleifer (2020), Bordalo et al. (2023), and in particular, Bordalo et al. (2018). These seminal studies introduce associative memory and diagnostic expectations in a tractable way from the fields of cognitive psychology and neuroscience. They show how expectations and probability assessments can be heavily influenced by contextual cues that trigger the appearance of certain memories. Diagnostic expectations are often applied to a single time series, e.g. inflation. They can explain why expectations often put too much weight on the most recent outcomes, which leads to over-extrapolation. In our application, similar to applications in Bordalo et al. (2023), we apply diagnostic expectations across two different series, i.e. diagnostic expectations about output growth, given an inflation cue. We show that diagnostic expectations, in conjunction with associative memory, can explain the stagflationary subjective model of the economy.

Similarly to a number of studies from the literature on adaptive learning in macroeconomics, we introduce our DSE framework in the context of a stochastic difference equation that, in the context of DSGE models, is actually an equilibrium condition. This raises the question whether this very equilibrium condition were to hold if the new expectation framework were applied to the primitive building blocks of the economy, and used in the derivation of the equilibrium conditions. This point has been raised and carefully discussed by Preston (2006, 2008), and is also known as Preston critique. In our case, we distinguish between deductive reasoning about the economy, which agents use to reason towards the equilibrium conditions, and estimation reasoning, which agents use to estimate the expectation terms in the equilibrium condition, based on the law of motion of fundamental state variables. We address the Preston critique by assuming that deductive reasoning is rational, while estimation reasoning is subject to DSE. We further elaborate on this point in Section 5.

3 Empirical Motivation

In this section, we provide motivating evidence for the fact that household expectations may exhibit patterns of intuitive thinking that would be absent for FIRE agents, with a focus on elements of a stagflationary subjective model. For this, we use data on US households' expectations from the Michigan Survey of Consumers (MSC). We focus on how US households perceive the relationship between output and inflation. While respondents in the MSC are broadly representative for the US population, their thinking may be more strongly determined by intuitive reasoning than in the case of, say, professional forecasters. Thus, we are more likely to find patterns associated with intuitive thinking in the MSC rather than, say, the Survey of Professional Forecasters, and we may therefore tilt the evidence in favor of our approach. However, as argued by Coibion and Gorodnichenko (2015b) and Candia et al. (2021), expectations of price setters in firms appear to resemble more those of average households than professional forecasters. Moreover, on average, aggregate consumption decisions are heavily influenced by households rather than sophisticated professionals. We therefore see the MSC as an appropriate source for motivating evidence on the presence of intuitive thinking in expectation formation.

The MSC collects a wide range of responses, including households' one-year-ahead expectations regarding changes in general business conditions, and the price level. We use quarterly data for the entire available horizon of 1960:Q1 until 2022:Q4, giving us 252 observations. Our expectation measures are based on the answers to the following two questions from the MSC.

- INF1: *By what percent do you expect prices to go up, on the average, during the next 12 months?*
- GDP1: *And how about a year from now, do you expect that in the country as a*

whole, business conditions will be better, or worse than they are at present, or just about the same? (Answers: "Better", "Same", "Worse", "Don't Know / NA")

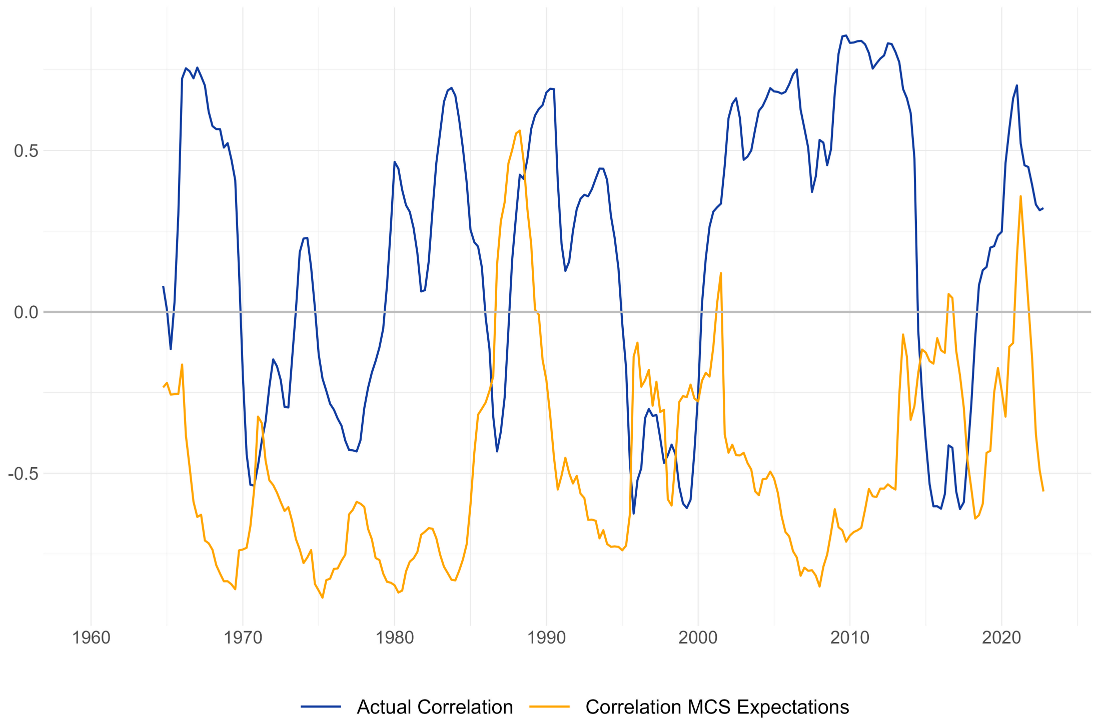
More precisely, for one-year ahead inflation expectations, INF1, we use mean answers – referred to as *Mean*. For one-year ahead GDP expectations, GDP1, we use the difference between the proportion of respondents answering "Better" and the proportion of respondents answering "Worse", plus 100 – referred to as *Relative*.[1]

If, on average, consumers think about the business cycle as being driven primarily by demand factors, we should observe a positive contemporaneous correlation between the INF1 and GDP1 series. By contrast, if they perceive supply factors as the main drivers of business cycles, we should observe a negative correlation. The yellow line in Figure 1 shows the correlation between the two expectation measures for 5-year rolling windows. It is striking that the curve is almost always in the negative domain. This strongly suggests that MSC respondents take a supply-side view. The blue line in Figure 1 shows the correlations of the actual realizations of year-on-year changes for GDP and inflation. Compared to realizations, there is a substantial downward shift in the correlations for expectations. This pattern has been well documented by, among others, Kamdar (2019), Candia et al. (2020), and Bhandari et al. (2023).

It seems difficult to reconcile the correlation patterns for expectations in Figure 1 with FIRE expectations. Clearly expectations always contain errors, but it is hard to see how these errors would lead to a systematic downward shift of the correlation curve for expectations if expectations would always be model-consistent. One explanation for the pattern may be that individuals somehow overreact to supply-side shocks. If so, we would expect the negative correlation to disappear once we control for these supply-side shocks. To explore this, we follow Enders et al. (2021) and use various measures of exogenous technology shocks to "purify" MSC expectations from variation attributed to those shocks. This means regressing the two expectation measures on measures for supply-side shocks to obtain residuals, which we refer to as purified expectation measures. We use the following purification schemes: *(i)* the current value, and 8 lags of TFP using Fernald (2014)'s measures of Business Sector TFP and utilization-adjusted TFP, and IST (investment-specific technology) news shocks from Ben Zeev (2018); *(ii)* the previous list,

[1] The *Mean* measure of INF1 is taken from Table 32 of the MSC, the *Relative* measure for GDP1 from Table 26.

Figure 1: Correlation between Output and Inflation Expectations, and Realizations



NOTE: The correlation for expectations is based on the measures "Relative" in Table 26 of MSC ("GDP1") and "Mean" in Table 32 of MSC ("INF1"). The correlation for realizations is based on Real Domestic Product (GDPC1) and on Consumer Price Index for All Urban Consumers for inflation (CPIAUCSL). The two curves show Pearson correlations over five-year rolling-windows.

agumented by 8 leads for TFP, and 8 lags of oil news shock measures from Känzig (2021)[2]; *(iii)* we add to the previous list the current value and 8 lags and 8 leads of real government consumption expenditures and gross Investment taken from the FRED database. Current values and lags of TFP, IST and oil shocks are obvious controls for supply shocks. We also include TFP *leads* since MSC respondents may actually be able to anticipate future TFP developments. We do not include leads for the IST and oil shocks, however, since the mentioned measures are already news shocks and hence of anticipatory nature. In the last scheme we add 8 lags and leads of government expenditures as a crude control for demand

[2]The oil news supply shock from Känzig (2021) is only available for the subsample of 1975:Q1 – 2022:Q4. When adding this measure, we lose 15 years of observations.

factors.[3] [4] Arguably, TFP, IST, and oil shocks capture a large part of the variation in technology and other supply side factors. The remaining drivers of fluctuations in output and inflation are therefore likely to be associated with the demand side. Thus, if the (excessive) negative correlation between expectation measures for output and inflation were due to (an overreaction to) supply shock, we would expect the correlation of purified output and inflation expectations turning positive. The data, however, tell a different story.

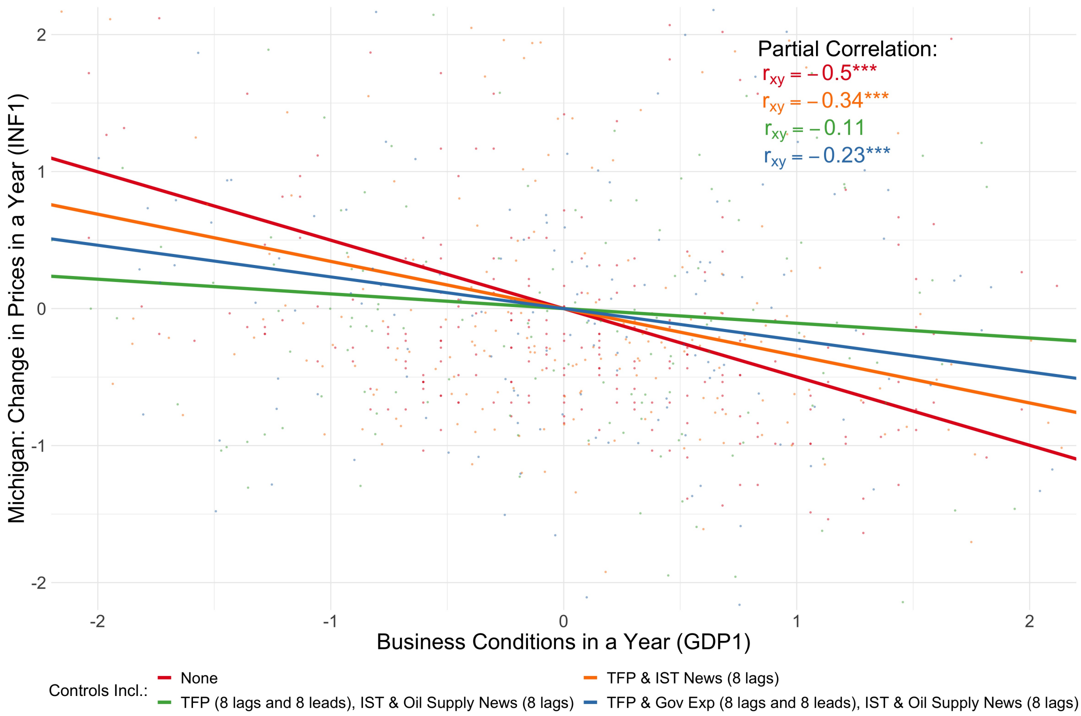
Figure 2 shows regression lines for regressions pertaining to the different purification schemes. It also includes a regression line for the case of no purification. In each case, both variables are standardized. To start with, the red line refers to the case of no purification. Due to standardization, the regression coefficient in this univariate regression is equal to the correlation coefficient and amounts to -0.5 (significant at the 1-percent level). The orange line represents the regression in which both expectation variables are regressed on 8 lags of TFP and IST news shocks. While the line becomes flatter, the correlation still amounts to -0.34 and remains significant at the 1-percent level. As a next step, we add 8 leads of TFP and 8 lags of oil news shocks to the purification. The correlation then amounts to -0.11 and is no longer significant. Recall, however, that the respective purification scheme only includes supply factors. To the degree that demand fluctuations play an important role for macroeconomic fluctuations, we would expect the correlation to be significantly positive rather than zero. When we add 8 lags and leads of government spending as a crude control for demand factors, the correlation becomes again more negative and is again statistically significant at the 1 percent level.

Output and inflation expectations, whether purified or not, show a relatively strong autocorrelation pattern. We therefore also estimate VAR models for these expectations. In particular, we are interested to what degree the negative correlation between the ex-

[3]We do not control for demand shocks here, for several reasons. First monetary and fiscal shocks are considerably harder to identify from the data than supply shocks (Ramey, 2016). See Brennan, Jacobson, Matthes, and Walker (2024) for the difficulties identifying monetary policy shocks using high-frequency data. Moreover, common approaches that use model-based restrictions to estimated demand shocks, such as Gali (1999), would defeat their purpose in the context of our analysis since they would identify non-fundamental fluctuations in expectations, which we are particularly interested in, as demand shocks and thus filter out variation that is the subject of our analysis. See also Lorenzoni (2009) for the fact that expectation shocks concerning the supply side may look like demand shocks, and would thus be estimated as such by common model-based approaches.

[4]We run regressions for many other purification schemes and obtain results very similar to the ones presented below.

Figure 2: Output and inflation expectations with supply-side controls



NOTE: The slopes of the plotted lines are estimated from regressions with the *Mean* of INF1 ("mean") as the left-hand-side and the *Relative* measure of GDP1 as the right-hand side variable. Both variables are purified by regressing them on the current realization, lags and leads of TFP, and the current realization and lags of oil news and IST news shocks, as discussed in the main text. The red line refers to the case with no purification. For each regression, both variables are standardized after purification. Significance levels (p-values): *** $p \leq 0.01$; ** $0.01 < p \leq 0.05$; * $0.05 < p \leq 0.1$.

pectation measures also shows at the level of VAR innovations.[5] In general, we find that the correlation coefficients take on similar values as those shown in Figure 2, and they are always significantly negative.

Overall, the empirical evidence so far suggests that agents' perception of the relationship between output and inflation is not (exclusively) driven by its factual nature but possibly (also) by a subjective intuitive model. This raises the question whether such subjective intuitive models help us to better understand macroeconomic fluctuations. It also raises the related question of how these subjective models can be captured in a tractable and versatile way, such that they can be easily integrated in DSGE models. It is these questions that are our main motivation to develop a tractable framework of DSE. As we will discuss, the model may capture a wider array of elements of intuitive thinking, but our main focus

[5]More precisely, we run VAR-X model for INF1 and GDP1, where the X-controls include the mentioned purification variables. See Appendix A for further details.

in this paper, in terms of applications, are subjective models of the economy.

4 Dual-System Expectations

4.1 Starting Point

In this section, we introduce our formal model of dual-system expectations (DSE). We do this for expectation terms appearing in an equation of the form

$$x_t = \sum_{h=0}^{\infty} A^h B E_t u_{t+h}. \tag{1}$$

Here, x_t is an n-dimensional vector, A represents an $n \times n$ matrix of coefficients, B is an $n \times 1$ matrix of coefficients, and u_t is an exogenous unidimensional state variable that follows the the AR1 process

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{2}$$

with $0 \leq \rho \leq 1$ and ε is white noise with variance σ_ε^2. We refer to u_t as a *fundamental state*, in contrast to non-fundamental "sentiments" introduced below. While our analysis also applies to u_t representing a vector, we prefer to present our model for the scalar case, as this simplifies the exposition.

For the case of DSGE models, equation (1) represents the solution of the Blanchard-Kahn type recursive equilibrium equation

$$x_t = A E_t x_{t+1} + B u_t. \tag{3}$$

In Section 6, we will consider the baseline New-Keynesian model from Chapter 3 in Galí (2015). In this application, x_t will represents the endogenous variables output (growth) and inflation. An example for the fundamental state u_t would be total factor productivity (TFP). Besides DSGE models, (1) also encompasses the case of, among others, consumption and saving decisions, where x_t represents consumption, and u_t represents per-period incomes. Throughout this section, we will simply assume that (1) can be derived from first principles in exactly this form under some consistent overarching expectation framework that entails the application of DSE to (1). We discuss this point in depth in Section 5.

In the FIRE case, agents can effortlessly get perfect model-consistent estimates of the terms $E_t u_{t+h}$ in (1). This assumption is so common that it may even appear odd to use the term "estimates" in this context. In this paper, however, we assume that estimating expectation terms occurs by invoking both rational and intuitive thinking, and the resulting expectations may not be model-consistent. In the absence of model-consistent expectations, the notation $E_t u_{t+h}$ is no longer unambiguous. On the one hand, we may simply read this as a statement that agents form expectations about u_{t+h}. On the other hand, we may read $E_t u_{t+h}$ as a statement that this means forming expectations as model-consistent mathematical expectations of u_{t+h}, in contrast to any other form of expectations about u_{t+h}. To avoid ambiguity, we introduce a new expectation operator, $\mathcal{E}_t$, that takes on the first meaning. Thus, $\mathcal{E}_t$ means that "an expectation is formed". As further discussed in Section 5, once agents have deduced the equilibrium condition of a model, which takes the form of (1), expectation terms get "estimated", and it is at this point when $\mathcal{E}_t$ gets replaced by a specific form of expectation formation, which can be FIRE, DSE, or any other potential form of expectations. An analogy from programming may help to make this more intuitive. When appearing in an equilibrium condition, $\mathcal{E}_t u_{t+h}$ is like the call to a subroutine that contains some code calculating the expected value of u_{t+h} for a particular specification of what "expected" means. The call may thus be for different subroutines, e.g. one containing the code for the FIRE case, another for DSE etc. By contrast, E_t will henceforth refer to model-consistent mathematical expectations as they appear in the FIRE case. With this, more precise, notation, our starting point is actually *not* (1), but its version with the operator $\mathcal{E}_t$, that, in the context of an equilibrium condition, we refer to as *placeholder* expectation operator:

$$x_t = \sum_{h=o}^{\infty} A^h B \mathcal{E}_t u_{t+h}, \tag{4}$$

We read (4) as follows: agents understand that x_t is determined by some form of expectations about future values u_{t+h}. However, to actually determine what this expectation terms would amount to, they – or they cognitive system – needs to insert a specific expectation operator for the placeholder term $\mathcal{E}_t$. This is comparable to the need of a specific subroutine containing some code, in the programming example. Because of this, FIRE and DSE agents can both agree on (4). However, they may end up with different values of x_t as they may invoke different specifications – or (mental) subroutines – for the expectation term.

4.2 Dual-System Expectations as Subjective Estimation Model

Under DSE, agents have a subjective estimation model, stated below, according to which they form a conditional expectation of a "target" variable. At first sight, from the previous subsection, it appears that the target should be u_{t+h}. However, this target would be one-dimensional. In (4), agents form expectations about a future fundamental state in order to obtain a solution for a number of n components of x_t, rather than for a one-dimensional object. Therefore, it is natural that agents form expectations of some n-dimensional vector. This allows our model to include the possibility that thinking about different elements of x_t triggers different contextual cues, which, in turn, affect expectation formation about the different elements in different ways (Bordalo et al., 2023; Bordalo, Gennaioli, & Shleifer, 2020). For instance, one of the elements may be inflation. This comes with different contextual cues than, say, GDP. It is important for our model that it can capture this difference. The natural choice for a n-dimensional target vector is therefore $Bu_{t+h} = (b_1, \ldots, b_i, \ldots, b_n)^T u_t$. It is the $n \times 1$-matrix B that "personalizes" the fundamental state u_{t+h} to each of the components in x_t.[6]

Under DSE, two signals – or "explanatory variables" – are used for each of the n estimation tasks, one for each target element in Bu_{t+h}. One signal speaks to rigorous thinking, the other speaks to intuitive thinking, or, to the "gut". The signal to rigorous thinking is simply the current fundamental state u_t. We assume that this state is directly observed, i.e. no filtering is required to obtain (an estimate of) u_t. We refer to the signal to intuitive thinking as *sentiments*, denoted by the vector $s_t = (s_{t,1}, \ldots, s_{t,i}, \ldots, s_{t,n})^T$. Each element of the sentiment vector is associated with the corresponding element in x_t, and it may be shaped by contextual cues triggered by this correspondence.

The concept of sentiments is an encompassing one and can capture a variety of phenomena. In our applied analysis in Section 6 we are particularly interested in sentiments as a way to represent subjective models of the economy. As an example, consider the "stagflationary" subjective model of the economy. Unexpectedly high increases in gasoline or

[6] Another potential candidate for an n-dimensional object would be $A^h Bu_{t+h}$. However, A^h goes far beyond "personalizing" u_{t+h} to different elements of x_t, as it quintessentially captures the endogenous dynamics of the system. A^h thus comes with a very different quality compared to B, which merely "personalizes". Furthermore, keeping A^h outside of the "estimation object" allows us to explain some interesting phenomena such as a switch in sign in correlation patterns between (DSE) expectations of endogenous variables and their realizations (a pattern that we observe in the data, see Section 6.)

grocery prices may trigger high inflation expectations, as a form of associative thinking (Coibion & Gorodnichenko, 2015b; D'Acunto et al., 2021; D'Acunto & Weber, 2024). These expectations may be excessively high compared to a benchmark FIRE economy in which these price increases would be only comparatively weak signals about inflation. Our model would capture this as an increase in the component of s_t corresponding to inflation, say s_t^π. If inflation is associated with a negative contextual cue, this may in turn trigger associative thinking about other negative events, and hence lead to negative expectations of output (growth), s_t^y. Alternatively, as we will further elaborate on in Section 7, in historical data, low output growth is *diagnostic* for high inflation (expectations), in the sense of Bordalo et al. (2018). This may also trigger a negative spillover from s_t^π to s_t^y. Generally, sentiments reflect elements of intuitive thinking that may be linked to associative memory, context effects, diagnostic expectations, and other elements of intuitive thinking that can be approximately captured by the assumptions about their actual and perceived law of motion, as stated below.

Our first assumption about sentiments concerns their law of notion as (unconsciously) "perceived" by agents. We that this using the superscript DS for the expectation operator:

$$E^{DS} s_{t+1} = s_t. \tag{5}$$

The actual law of motion for sentiments follows the VAR(1) process

$$E s_{t+1} = \Phi s_t + \eta_t, \tag{6}$$

with $\eta_t \in \mathbb{R}^n$, $E\eta_t = 0$, $E[\eta_t \eta_t^T] = \Sigma^{(\eta)}$, and $E[\eta_t \eta_{t-s}^T] = \mathbf{0}$ for all $s > 0$, and $E_t \eta_t u_{t+s} = 0$ for all $s \in \mathbb{Z}$. Furthermore, the eigenvalues of Φ are strictly inside the unit circle. Equation (5) states that agents (unconsciously) feel sentiments as if they were permanent. According to (6), sentiments actually regress to their mean of zero. As an example, consider that agents are currently in a pessimistic mood since, say, gasoline or grocery prices increased substantially. Assume that this leads to negative sentiments. According to the actual low of motion, negative sentiments will gradually fade over time. However, agents are unaware of this and hence do not take this into account when forming expectations. To the degree that sentiments were conscious, agents could correct for their influence since they benefit from forming expectations that are as accurate as possible. This allows them to make decisions that lead to a higher level of satisfaction. (In our model, agents do not derive any anticipatory utility from, say, dreaming about a rosy future; hence there is no positive counter-value of forecast errors induced by sentiments.) However, it is only natural to assume that agents are not aware of being exposed to sentiments, such that they also cannot take into account that sentiments will gradually fade.[7]

Concerning the actual law of motion, sentiments follow a stationary VAR(1) process. Innovations to sentiments are not predictable, but they may be correlated across components of s_t, i.e., $\Sigma^{(\eta)}$ need not be diagonal. This allows us to capture aspects of subjective models of the economy by means of correlation patterns. An important such instance is the stagflationary model. We also assume that sentiments are uncorrelated with fundamental states at all leads and lags. While it may also be of interest to consider that sentiments are affected by fundamental states u_t (or functions thereof), this would make it much harder to identify them in data, as there would then be substantial correlation between fundamental states and sentiments. It would hence be unclear how responses in endogenous variables, like output or inflation, would be separately attributed to fundamental states and to sentiments. We therefore view the stated orthogonality assumption as a natural starting point.

We now characterize the multivariate distributions that are at the core of agents' subjective estimation model that defines DSE. For each i pertaining to a component in x_t, there is a multivariate distribution with three components: (1) a target variable $b_i u_{t+h}$ that is to be predicted; (2) a signal to rational thinking u_t; and (3) a signal to intuitive thinking $s_{t,i}$. We assume that their joint distribution is multivariate normal:

$$(b_i u_{t+h}, u_t, s_{t,i})^T \sim N(\mathbf{0}, \Sigma_{h,i}) \tag{7}$$

with

$$\Sigma_{h,i} = \begin{bmatrix} Var(b_i^2 u_t) & \delta^h \rho^h b_i Var(u_t) & \delta^h Cov(u_{t+h}) \\ \delta^h \rho^h b_i Var(u_t) & Var(u_t) & 0 \\ \delta^h Cov(u_{t+h}, s_{t,i}) & 0 & Var(s_{t,i}) \end{bmatrix} \tag{8}$$

The parameter δ, with $0 \leq \delta \leq 1$, represents cognitive discounting in the sense of Gabaix

[7]See Adam et al. (2024) and Bianchi et al. (2022) for the fact that (even) professional forecasters overweigh the importance of personal information, which our model would also capture as a sentiment effect. Clearly, if professional forecasters were aware of this, they would find it in their interest to correct for it since it would reduce forecast errors.

(2020). Under DSE, agents estimate $\mathcal{E}_t[Bu_{t+h}]$ using the conditional expectation of their subjective model, denoted by $E^{DS}[Bu_{t+h}|u_t, s_t]$. As a shorthand, we also simply write $E_t^{DS}[Bu_{t+h}]$. The expectation operator E^{DS} is the defining part of DSE; it is a standard mathematical expectation operator, *given* (7) and (8). However, these are subjective and generally do not correspond to the actual law of motion of the economic system in equilibrium. It's for this reason that we use the superscript DS. To get a better understanding of DSE, we first consider the case of a unidimensional x_t (and s_t).

4.3 The Case of a Unidimensional x_t

In this subsection we consider the case $n = 1$. To get a first grasp on the subjective model specified by (7) and (8), assume, for the moment, that $\delta = 1$. Using standard results for conditional means for multivariate normal distributions, we get

$$E_t^{DS}[b_1 u_{t+h}] = \rho^h b_1 u_t + \gamma_1 s_{t,1},$$

where $\gamma_1 \equiv Cov(u_{t+h}, s_{t,1})/Var(s_{t,1})$. The subscripts serve as a reminder that we are in the univariate case, which applies to the x_t-component $x_{t,1}$ and s_t-component $s_{t,1}$. The first term on the right-hand side refers to rigorous thinking and reflects the actual law of motion of u_t in (2). The second term reflects a perturbation by intuitive thinking. It perturbs rigorous thinking by reacting to sentiments that, in a FIRE world, would carry no information about the future evolution of the economy. FIRE agents would be able to perfectly control their minds such that intuitive thinking has no influence on expectation formation. This corresponds to the case of $\gamma_1 = 0$ (and $\delta = 1$). In the non-FIRE case, sentiments may become partly, and temporarily, self-fulfilling. The degree of the perturbation is governed by γ_1, it reflects the extend to which agents (unconsciously) react to their "guts". Note that γ_1 does not vary with the time horizon h. This reflects (5), i.e. agents' intuitive thinking assesses the future as if sentiments' somehow come with a permanent predictive power, that does not change with the time horizon h. Technically, we will not make any individual assumptions about $Cov(u_{t+h}, s_{t,1})$ and $Var(s_{t,1})$, but only about their ratio, i.e. γ_1.

We now add δ to the picture. It represents cognitive discounting, which we take from Gabaix (2020). We assume $0 \leq \delta \leq 1$. To understand the role of this parameter, consider how it affects $E_t^{DS}[u_{t+h}]$. We now have

$$E_t^{DS}[b_1 u_{t+h}] = \delta^h \rho^h b_1 u_t + \delta^h \gamma_1 s_{t,1}, \tag{9}$$

since δ multiplicatively affects each covariance term that appears in the conditional expectation expression. In agents' subjective reasoning, for each additional increase in the horizon h, the perceived covariances between the target and the signals u_t and $s_{t,1}$, gets shrunk by an (additional) factor δ. This reflects that, compared to making a forecast about period $t+h$, agents feel less confident about the perceived information in their signal when making a forecast about $t+h+1$ by a factor of δ. In other words, the subjective precision of the signals declines exponentially with horizon. Implicitly, and exactly as in Gabaix (2020), expectations get shrunk towards a default of zero, which is the unconditional steady state value of x_t. The motivation for this assumption in our context is fourfold. First, concerning rigorous thinking, reasoning about the future requires effortful mental simulations (Gabaix, 2020; Gabaix & Laibson, 2017). Cognitively, the required effort tends to increase with the time horizon and therefore expectations about far-away quantities get more and more shrunk towards a default. This applies also to rigorous thinking, which is effortful already in the first place. Second, intuitive thinking, by its very nature, is simple and not very far forward-looking. Therefore, it is natural to apply cognitive discounting also to sentiments. The third motivation is related, but more technical. We assume that agents perceive sentiments as permanent (see (5)). Without cognitive discounting, this may lead to convergence problems for the infinite sum in (4) if we do not want to require that the eigenvalues of A are all strictly inside the unit circle. This may be of interest in some applications. Last, it can be shown that there is an alternative derivation for DSE. With this alternative derivation, δ can be interpreted as the weight of rigorous reasoning in expectation formation, while $1-\delta$ is the weight of intuitive thinking. This version suggests to set $\gamma = (1-\delta)/\delta$ (for $\delta > 0$). Thus, δ becomes the only parameter regulating the mix between rigorous and intuitive thinking. Sentiments then matter if and only if $\delta < 1$.

For the univariate case, under DSE, (4) becomes

$$x_{t,1} = (1-\delta\rho a_1)^{-1} b_1 u_t + \delta a_1 (1-\delta a_1)^{-1} \gamma_1 s_{t,1}, \tag{10}$$

where a_1 and b_1 denote the single elements in A and B, respectively. From this, we already see how sentiments can influence actual outcomes.

Consumption-Saving Example. One example for the univariate case is a simple consumption-saving decision where per-period income follows an AR1 process characterized by (2), such that u_t represents per-period income; and x_t refers to period-t consumption. The parameter b_1 represent a type of annuity factor in this example. In the absence of sentiments, consumption fluctuates only with the innovation term in the income process, dampened by the annuity factor. Sentiments may superimpose their own dynamics, e.g. in the form of waves of optimism or pessimism. The degree to which this can happen is regulated by γ_1. Cognitive discounting, i.e. δ has a dampening effect on fluctuations from both sources.

4.4 The General Case

For the general case, the subjective estimation model manifests itself as a system of n equations, where each equation reads as

$$E_t^{DS}[b_i u_{t+h}] = \delta^h \rho^h b_i u_t + \delta^h \gamma_i s_{t,i} \tag{11}$$

In vector form, this is written as

$$E_t^{DS}[B u_{t+h}] = \delta^h \rho^h B u_t + \delta^h \bar{\Gamma} s_t \tag{12}$$

where $\Gamma \equiv (\gamma_1, \gamma_2, \ldots, \gamma_n)$, and $\bar{\Gamma} = \operatorname{diag} \Gamma$. With this, we can now state the DSE solution for x_t in (4) for the multivariate case.

PROPOSITION .1 *Assume $0 \leq \delta \leq 1$, and that the eigenvalues of ρA and δA both lie strictly inside the unit circle. Then*

$$x_t = (I - \delta\rho A)^{-1} B u_t + \delta A (I - \delta A)^{-1} \bar{\Gamma} s_t, \tag{13}$$

$$E_t^{DS} x_{t+1} = \delta\rho (I - \delta\rho A)^{-1} B u_t + \delta (I - \delta A)^{-1} \bar{\Gamma} s_t. \tag{14}$$

The covariance matrices of x_t and $E_t^{DS}x_{t+1}$ are given by

$$Var(x_t) = (I-\delta\rho A)^{-1}BB^T\left((I-\delta\rho A)^{-1}\right)^T Var(u_t)+\delta^2 A(I-\delta A)^{-1}\bar{\Gamma} Var(s_t)\bar{\Gamma}\left((I-\delta A)^{-1}\right)^T A^T, \quad (15)$$

$$Var\left(E_t^{DS}x_{t+1}\right) = \delta^2\rho^2(I-\delta\rho A)^{-1}BB^T\left((I-\delta\rho A)^{-1}\right)^T Var(u_t)+\delta^2(I-\delta A)^{-1}\bar{\Gamma} Var(s_t)\bar{\Gamma}\left((I-\delta A)^{-1}\right)^T. \quad (16)$$

Proposition .1 sheds light on generic characteristics of an economy with dual-system expectations. Note that the standard case of rational expectations is nested for the special case where $\delta = 1$ and $\gamma = 0$. In this case, agents suppress the voice of their intuitive thinking, or do not "listen to their gut". For the case of $\gamma > 0$, the proposition shows to what degree the solution for the endogenous variable x_t is driven by the fundamental state u_t, and by non-fundamental sentiments. The impact of sentiments depends on both, γ, and δ, as well as the variability of sentiments themselves.

With respect to applications to DSGE models, it is of interest to what degree sentiments affect the autocorrelation structure of x_t, or the persistence of impulse responses. Clearly, sentiments lead to persistent impulse responses in components of x_t to the degree that sentiments are themselves persistent, which is regulated by the autogressive coefficient matrix Φ (see (6)). To the degree that there are persistent patterns in macroeconomic time series that are hard to explain by fundamental factors, sentiments provide a potential non-fundamental explanation for these. Indeed, our estimates of Φ for expectation data associated with output and inflation exhibit substantial persistence.[8]

Proposition .1 also gives a precise answer to the question to which degree sentiments may turn out self-fulfilling. Suppose that, before a certain period τ, sentiments have been constant at zero. At the beginning of period τ, agents unexpectedly wake up with an sentiment of $s_\tau \neq 0$. This happens with no corresponding development in the fundamental u_τ. Equation (14) shows that this affects dual-system expectations of x_{t+1} not by an amount of s_τ but rather by $\delta A(I-\delta A)^{-1}\bar{\Gamma}s_\tau$. The components in this vector may be larger or smaller than s_τ. They may be larger because s_τ affects not only expectations relating to $\tau+1$, but to all time horizons further into the future, although with diminishing

[8]In the DSGE literature, it is common to use habit formation for explaining persistence in macroeconomic variables. For future research, it may be of interest to explore whether sentiments could explain related patterns in the data in the absence of habit formation.

importance, because of cognitive discounting (as well as a contracting effect associated with the matrix A). The effect may be smaller than s_τ when δ is sufficiently small and, more importantly, when the coefficients γ_i in $\bar{\Gamma}$ are small, meaning that agents manage to suppress the influence of intuitive thinking to a large degree. The degree to which sentiments then affect the actual outcome x_τ is further mitigated by the matrix A. This follows simply from the recursive equation $x_t = A\mathcal{E}_t x_{t+1} + Bu_t$. In line with the above discussion, the effect of sentiments on outcomes x_τ can also be smaller or larger than s_τ. In theory, expectations may even become "over-self-fulfilling", meaning that upon impact, the impulse response on x_τ exceeds the magnitude of the impulse s_τ.[9] We will consider impulse responses for a specific application in Section 6.

Before discussing the implications of the results in Proposition .1 for the correlation patterns between the equilibrium values of variables, as well as their expectations, let us clarify in somewhat more detail how sentiments affect these correlation patterns. It follows from (11), (13), and (14) that each sentiment component i is directly associated with, and only with, the same component i in x_t. In other words, say, an inflation sentiment, only directly affects inflation expectations, and then the equilibrium outcome for inflation. It has no direct effect on, say, output expectations and equilibrium output. This manifests itself in the diagonal form of the matrix $\bar{\Gamma}$. This specification keeps the sentiment model tractable and transparent, and also facilitates empirical implementations.

There are, however, two indirect effects of how an inflation sentiment may be linked to output expectations, and realizations. First, since $\Sigma^{(\eta)}$ is generally not diagonal, the sentiments may be correlated with each other. In Section 7 we show how this may arise due to associative memory and diagnostic expectations. This first indirect link between sentiment is purely psychological/cognitive. There is a second channel through which DSE expectations can get correlated due to the presence of sentiments. This is due to the effect that, in (13) and (14), sentiments are transformed with the matrix $(I - \delta A)^{-1}$. In this way, expectations about x_{t+1} become a linear combination of all sentiments. As a further remark, we note that some components in x_t may not naturally trigger intuitive thinking. For these components, it's natural to set the respective sentiments to zero, such that they are then also not correlated with any other sentiments.

[9]This observation may be of interest in the context of phenomena that are otherwise explained by sunspots.

With respect to the discussion in Section 3, Proposition .1 also sheds light on how the correlation patterns of expectations of macroeconomic variables differ from the correlation patterns in their realizations. For the rational-expectation case, in which all variation in x_t and expectations thereof is driven by the fundamental factor, covariances between the elements in x_t and the expectations thereof will always have the same sign. Thus, rational expectation models are inconsistent with the empirical observation from the Section 3. As a result, a change of signs in covariances can only originate from the sentiment parts in the covariance matrices in (15) and (16). In the sentiment part, the only term that differs between these is the appearance of A (and A^T) in (16). To see how covariances may switch signs, denote the covariance matrix of the sentiment part of $E_t^{DS} x_{t+1}$ by V. Then the covariance matrix of the sentiment part of x_t is given by AVA^T. For a sign switch, we need that an off-diagonal element of AVA^T has the opposite sign of of the corresponding off-diagonal element in V. This may happen if one or several off-diagonal elements of A are negative. We will consider an example in Section 6.

A final important observation with respect to Proposition .1 is that if we have data on expectations, we can use (13) to directly identify sentiments, provided we are willing to fix the parameters of the model to specific values. (Alternatively, they could be estimated using, e.g. simulated methods of moments.) We will use this direct procedure in Section 6 where we estimate the intuitive part in (13) and then solve for s_t.

5 Deductive vs. Estimation Reasoning in Equilibrium Models

In this section, we understand (4), or its associated recursive form

$$x_t = A\mathcal{E}_t x_{t+1} + Bu_t, \tag{17}$$

as an *equilibrium condition* of a DSGE model. If we adopt this view, if follows that, in the previous section, we have indeed introduced a new concept of expectation formation directly within the scope of an equilibrium condition. This raises the question whether this very equilibrium condition would still hold if the expectation formation concept had been applied to the primitive elements of the model, and then in all steps used to derive the equilibrium conditions. Thus, it is a legitimate question why we would not already apply DSE to agents' intertemporal utility function, or for firms' value function; or then to the consumption Euler equation or the first-order condition for optimal price setting etc. The relevance of this point has been raised by Preston (2006, 2008) in the context of macroeconomic models of adaptive learning. Our aim in this section is to embed the DSE framework into an overarching and consistent framework of expectation formation that assures that the stated equilibrium conditions are indeed valid when the framework is applied to the primitive building blocks and in all steps towards the derivation of the equilibrium condition.

A main change in perspectives in our approach with respect to the FIRE framework is that we view agents as actively reasoning about the economy, if in a stylized way. In particular, they reason about how to deduce equilibrium relationships from the primitive characteristics of the economy. Once they have deduced their mental representation of equilibrium relationships, they form estimates of the expectation terms in the equilibrium condition, as characterized in the previous section. We argue that these two types of reasonings are different. The first type of reasoning only consists of deducing new relationships from more primitive ones (such as the Phillips Curve from first-order conditions for optimal price setting); we refer to it as *deductive reasoning.* This only involves reasoning in relationships; some entities in these relationships are expectation terms. However, the specifics of these expectations do not matter, at least within bounds, as we shortly discuss. For these expectations, we also use the expectation operator in "placeholder" form $\mathcal{E}_t$. In this case, the idea is to indicate that, in the deduction steps, no specific model of expectation formation is yet used. In the programming analogy, $\mathcal{E}_t$ is the call to a class of subroutines – in our analogy those estimating expectations. However, during the deductive steps, this call is just passed from step to step, without getting referred to a particular subroutine yet. This only happens at the "estimation stage".

When it comes to the determination of the equilibrium value of endogenous variables, reasoning is no longer deductive. Rather, this involves now the application of the law of motion for the fundamental state u_t and forming estimates of quantities associated with future values of the fundamental state, as discussed in the previous section. We refer to the reasoning involved in this stage as *estimation reasoning.*

In a nutshell, we propose an overarching framework of expectation formation that limits the scope of DSE to estimation reasoning. To better understand this idea, and to put it into context, it may help to consider an analogy from econometrics. In empirical work, there is often a main estimation equation. Frequently, however, this equation is not a priori given but rather deduced from more primitive equations that possibly come from a theoretical model, or are derived from a heuristic or naive specification that, when used directly for estimation, would not yield any meaningful results. Deducing the estimation equation from more primitive equations involves algebra, or reasoning steps. In these steps, nothing is estimated. Even if the final estimation results may be substantially biased or perturbed, e.g. due to data problems or an error in the code carrying out the estimation, the derivation of the estimation equation may still be flawless. Thus, a flawlessly deduced estimation equation does not stand in contradiction to flawed estimation results. In analogy to this, there is no contradiction between agents in our model reasoning flawless through the relational structure of the economy to deduce equilibrium conditions, and then intuitive thinking perturbing the resulting estimation.

We now state the above arguments in more formal terms for a simple log-linearized DSGE model that does not include any predetermined variables and for which all dynamic stochastic equations do not exceed first order. For such a model, all the intermediate relationships that are used to deduce equilibrium relationships either take on a recursive form

$$\mathcal{F}_R^l(x_t, z_t, u_t, \mathcal{E}_t x_{t+1}, \mathcal{E}_t z_{t+1}, \mathcal{E}_t u_{t+1}) = 0, \tag{18}$$

or appear in solution form

$$\mathcal{F}_S^l(x_t, z_t, u_t, \mathcal{E}_t x_{t+1}, \mathcal{E}_t z_{t+1}, \mathcal{E}_t u_{t+1}, \ldots, \mathcal{E}_t x_{t+h}, \mathcal{E}_t z_{t+h}, \mathcal{E}_t u_{t+h}, \ldots) = 0. \tag{19}$$

Typically, some conditions are stated in recursive form (e.g. first-order optimality conditions), while others in solution form (utility and intertemporal profits). The vector x_t includes the set of endogenous variables that will appear in the equilibrium relationship, while z_t includes endogenous variables that are "substituted out" in the process of deducing equilibrium relationships (e.g. the nominal interest rate that in a New-Keynesian model is usually substituted out by inserting a Taylor rule). Fundamental states are included by u_t, which, in this section, may also represent a vector. The superscript l indicates that there may be several deduction steps to obtain the equilibrium relationship. In the recursive form, no variables have time indices higher than $t+1$. By contrast, the solution form is obtained by using $\mathcal{F}_R^l$ recursively to substitute for certain components in $\mathcal{E}_t x_{t+1}$ or

$\mathcal{E}_t z_{t+1}$, etc. Due to this, some elements in x_{t+h} or z_{t+h} disappear. However, for notational simplicity we do not make this explicit but rather imagine that there are respective zero coefficients as part of the functional form of $\mathcal{F}^l_S$. For some deduction steps, at least in some models, it is important that the recursive form can be deduced from the solution form, and vice versa. An example is the derivation of the Phillips Curve from first-order conditions for price setting. Thus, in general, two economies, one with FIRE and one with DSE agents, with otherwise identical characteristics, will only have an identical form of (17) if the law of iterated expectation holds for $\mathcal{E}_t$. The observation that the law of iterated expectation is necessary for the derivation of equilibrium conditions in standard form in the case of subjective expectations has been made by Adam and Padula (2011).

The deduction steps to arrive at an equilibrium equation involve starting with some $\mathcal{F}^1_k$, $k = R, S$, which includes the primitive building blocks such as the utility function and the value function of firms, etc. $\mathcal{F}^2_k$ then includes first-order conditions etc. After a sufficient number of deduction steps, an equilibrium relationship is derived and, if the law of iterated expectations holds for $\mathcal{E}_t$, can be stated in solution form as

$$\mathcal{G}(x_t, u_t, \mathcal{E}_t u_{t+1}, \ldots, E_t u_{t+h}, \ldots) = 0. \tag{20}$$

Note again that all steps in the derivation only include deductive reasoning, no estimation reasoning. The essence of estimation reasoning is to obtain a specific solution for x_t that is derived by estimating the terms $E_t u_{t+h}$ (or related terms in vector form) by invoking the law of motion for u_t. The use of this law of motion is specific is indeed the defining element of estimation reasoning.

The following lemma about expectations in deductive reasoning is obvious, but important for a consistent specification of our overarching expectation framework.

LEMMA .1 *Suppose that in all steps from $\mathcal{F}^1_k$, $k = R, S$ to $\mathcal{G}$, a deduction step is valid under $\mathcal{E}_t$ if and only if it is valid under the mathematical expectation operator E_t, then $\mathcal{G}$ takes on the same form for E_t and $\mathcal{E}_t$, differing only in their respective expectation operator. In particular, the law of iterated expectations holds for $\mathcal{E}_t$.*

We refer to deduction steps under $\mathcal{E}_t$ that follow the same logic as under E_t collectively as *rational deductive reasoning.* Importantly, there is no necessity that rational deductive reasoning implies FIRE for estimation reasoning. Referring to the previous analogy from econometric estimations, this is illustrated by empirical studies with correctly deduced estimation equations which still may produce biased estimates due to data problems (or sometimes coding mistakes). In our analysis, we indeed assume an overarching framework with rational deductive reasoning and DSE estimation reasoning. When we want to highlight the overarching framework, we refer to it as DSE for estimation reasoning framework of DSEER. When there is no danger of misunderstanding, however, we prefer to stick to the simpler term of DSE (while also meaning DSEER).

Like any other economic model, DSEER is not a realistic framework. Would it be more realistic if we did not assume rational deductive reasoning? Plausibly, in reality, intuitive thinking would not only perturb estimation reasoning but also deductive reasoning. It is even plausible that the degree to which the two are perturbed by intuitive thinking are positively related. However, from a methodological point of view, it is still valuable to explore what insights into the empirical patterns in macroeconomic series we may get by limiting the deviation from the standard FIRE framework to estimation reasoning. It is certainly natural to proceed step by step and start with a perturbation of estimation reasoning. We view it, however, as a question of major interest to then turn to study perturbations in deductive reasoning. Some interesting steps in this direction have already been taken by Garcìa-Schmidt and Woodford (2019) and Angeletos and Lian (2023).

6 A New Keynesian Economy with Dual-System Expectations

In this section, we apply the DSE framework to the textbook version of the New Keynesian model in Galí (2015), with minor modifications. This model includes the two variables output and inflation. We characterize the equilibrium and then calibrate impulse responses to a shock in inflation sentiments. While our DSGE setting is somewhat minimal and stylized, the analysis of this section illustrates how to apply and operationalize DSE in DSGE models. The impulse responses that we find are relatively persistent and are suggestive for sentiments as a source of persistent fluctuations in larger models.

6.1 A Simple New Keynesian Model with Dual-System Expectations

Following Galí (2015), we consider an economy with a representative household whose objective function in period t is

$$\mathcal{E}_t \sum_{h=0}^{\infty} \beta^h \left(\frac{C_{t+h}^{1-\sigma} - 1}{1-\sigma} - \frac{N_{t+h}^{1+\phi}}{1+\phi} \right) \tag{21}$$

where we use the placeholder expectation operator $\mathcal{E}_t$, β is the discount factor, C_{t+h} denotes a CES index of consumption, σ the coefficient of relative risk aversion, N_{t+h} stands for hours worked, and ϕ is the inverse of the Frisch elasticity of labor supply. Firms operate in monopolistic competition. Their production function is

$$Y_t = A_t N_t^{1-\alpha}, \tag{22}$$

where Y_t denotes output, A_t total factor productivity (TFP), and α determines how marginal productivity—and thus marginal costs—change with hours worked, i.e. N_t.

To apply the DSE framework, the model is log-linearized such that we obtain equilibrium relationships in the form of (17). We deviate from Galí (2015) in a minor way: we consider output deviations from a zero-growth steady state (the "trend") rather than from the natural level. We do so since we find it more plausible that agents' intuitive thinking centers around deviations from trend rather than deviations from the natural level. The latter is a cognitively more demanding concept, partly also because it is unobservable and poses challenges even to econometricians (Orphanides & Norden, 2002). As discussed in the previous section, for the derivation of equilibrium relationships in the form of (17), we assume that agents' deductive reasoning is fully rational. It therefore follows that we obtain the same equilibrium relationships as for the FIRE case, just with the placeholder expectation operator replacing the mathematical operator.

Let y_t denote the log deviation of output from its steady state value. If there is no danger of misunderstanding, we refer to y_t simply as output. Let π_t denote inflation. In log-linear form, the IS equation is given by

$$y_t = \mathcal{E}_t y_{t+1} - \frac{1}{\sigma} \left(i_t - \mathcal{E}_t \pi_{t+1} - \rho_d \right), \tag{23}$$

where ρ_d denotes the discount rate. The New Keynesian Phillips curve is given by

$$\pi_t = \beta \mathcal{E}_t \pi_{t+1} + \kappa y_t - \lambda \frac{1+\phi}{1-\alpha} a_t, \tag{24}$$

where the fundamental state variable a_t denotes the log of TFP, and $\lambda \equiv \frac{(1-\theta)(1-\beta\theta)(1-\alpha)}{\theta(1-\alpha+\alpha\epsilon)}$ and $\kappa \equiv \lambda(\sigma + \frac{\varphi+\alpha}{1-\alpha})$. For the log of TFP, a_t , the law of motion is given by

$$a_t = \rho_a a_{t-1} + \varepsilon_t^a.$$

We include TFP as as an exogenous fundamental state, rather than any other such state, since impulse responses to an inflation sentiment shock, considered below, show similarities to those for a TFP shock, and inclusion of the latter allows us to compare the two. Otherwise, the choice of a TFP shock is completely arbitrary. In particular, this choice has no influence on the shape of impulse responses for an inflation sentiment shock.

We assume the Taylor rule

$$i_t = \rho_d + \phi_\pi \pi_t + \phi_y y_t \tag{25}$$

where the parameters $\phi_\pi \geq 0$ and $\phi_y \geq 0$ specify the reaction of the central bank to inflation and the deviation of output from steady state, respectively.[10] Combining the IS equation, the New Keynesian Phillips curve and the Taylor rule, we obtain the equilibrium relationship for the two endogenous variables output and inflation

$$\begin{pmatrix} y_t \\ \pi_t \end{pmatrix} = A \begin{pmatrix} \mathcal{E}_t y_{t+1} \\ \mathcal{E}_t \pi_{t+1} \end{pmatrix} + B_a a_t \tag{26}$$

with

$$A = \Omega \begin{bmatrix} \sigma & 1-\beta\phi_\pi \\ \sigma\kappa & \kappa + \beta(\sigma+\phi_y) \end{bmatrix}; \quad B_a = \Omega\lambda \frac{1+\phi}{1-\alpha} \begin{pmatrix} \phi_\pi \\ -\sigma-\phi_y \end{pmatrix}; \quad \Omega = [\sigma + \kappa\phi_\pi + \phi_y]^{-1}$$

For both output and inflation, intuitive thinking is perturbed by respective sentiments s_t^y, and s_t^π that we refer to as output and inflation sentiment, respectively. Their units

[10] We do not include any fundamental state variable that exogenously drives fluctuations in the nominal interest rate since it offers no special insights in the context of our DSE model.

are commensurate to the log-linear units of the other model variables. For parsimony, we assume that the γ coefficients in (7) are all equal, i.e. $\gamma_y = \gamma_\pi \equiv \gamma$.[11]

Using Proposition .1, we obtain the DSE solution for the New Keynesian case stated in the next proposition.

PROPOSITION .2 *Suppose that the stability condition*

$$\delta < \phi_\pi + \frac{1-\beta\delta}{\kappa}\phi_y + \frac{\sigma}{\kappa}\left[1 - \delta(1+\beta(1-\delta))\right]. \tag{27}$$

holds, such that both eigenvalues of δA are strictly inside the unit circle. Under dual-system expectations (DSE), the solution for output and inflation for the New-Keynesian model is given by

$$\begin{pmatrix} y_t \\ \pi_t \end{pmatrix} = (I - \delta\rho_a A)^{-1} B_a a_t + \gamma\delta A(I-\delta A)^{-1}\begin{pmatrix} s_t^y \\ s_t^\pi \end{pmatrix}, \tag{28}$$

and expectations are given by

$$E_t^{DS}\begin{pmatrix} y_{t+1} \\ \pi_{t+1} \end{pmatrix} = \delta\rho_a(I - \delta\rho_a A)^{-1} B_a a_t + \gamma\delta(I-\delta A)^{-1}\begin{pmatrix} s_t^y \\ s_t^\pi \end{pmatrix}. \tag{29}$$

The actual law of motion of sentiments is given by

$$\begin{pmatrix} s_t^y \\ s_t^\pi \end{pmatrix} = \Phi\begin{pmatrix} s_{t-1}^y \\ s_{t-1}^\pi \end{pmatrix} + \begin{pmatrix} \eta_t^y \\ \eta_t^\pi \end{pmatrix}, \tag{30}$$

with $(\eta_t^y, \eta_t^\pi) \equiv \eta_t$, $E\eta_t = 0$, $E[\eta_t\eta_t^T] = \Sigma^{(\eta)}$, and $E[\eta_t\eta_{t-s}^T] = \mathbf{0}$ for all $s > 0$, and $E_t\eta_t u_{t+s} = 0$ for all $s \in \mathbb{Z}$. For the perceived law of motion of sentiments, we have $E^{DS}s_{t+1} = s_t$.

6.2 Impulse Responses to Inflation Sentiments

A main reason why we are interested in DSE is that sentiments may act as a non-fundamental driver of fluctuations in macroeconomic variables. We are therefore inter-

[11] In the calibration of impulse responses for an inflation sentiment shock, shown below, γ will actually vanish.

ested in associated impulse response functions. In a first step, this requires elaborating on what can sensibly be understood as an exogenous sentiment shock. In our baseline New-Keynesian model, there are two sentiments. According to (30), they follow a VAR(1)-process in which the innovations for output and inflation sentiments are correlated. As usual with VARs, this raises the question of how to identify fundamental exogenous shocks that have a causal impact on other macroeconomic series.

We address this identification based on robust findings from the recent empirical literature on subjective macroeconomic models and expectations (Andre et al., 2022; Binetti et al., 2024; Hou & Wang, 2024; Stantcheva, 2024). This literature highlights the special role that inflation plays in subjective macroeconomic models. Broadly summarizing, a key finding is that households, and also firms, are highly concerned about inflation (or more generally about price changes). By contrast, unemployment and GDP, as well as further macroeconomic variables, are much less in agents' primary attention field. Moreover, agents' subjective models are characterized by causal pathways running from inflation to output and unemployment, not in the reverse (and also not in both) directions (Hou & Wang, 2024).

Our identification for sentiment shocks is directly based on these empirical findings. We consider the impact of an exogenous shock to *inflation* sentiments, while output sentiments only *react* to inflation sentiments. More precisely, we take the innovations of inflation sentiments, η_t^π, as exogenous. We consider the effect of a shock of size $\Delta\eta_t^\pi$ and determine the response of output sentiments using the conditional mean estimated by a linear regression

$$E[\Delta\eta_t^y|\Delta\eta_t^\pi] = \frac{Cov(\eta_t^y, \eta_t^\pi)}{Var(\eta_t^\pi)}\Delta\eta_t^\pi. \tag{31}$$

To quantify the coefficient on $\Delta\eta_t^\pi$, we use empirical estimates of the innovations η_t^y and η_t^π, which we obtain from running a first-order VAR for empirical measures for sentiments. The latter are derived from output and inflation expectation data from the Michigan Survey of Consumers described in Section 3. Using the measures for inflation and output expectations discussed there, we run the same purification regressions as in Section 3 to filter supply shocks. For the results reported in the main text, we purify for TFP and IST shocks.[12] Ignoring demand shocks, we take the purified series as corresponding to the

[12]Including further supply shock measures has a negligible affect on results.

second term of the right-hand side of (29). We denoted the purified empirical measures for output and inflation expectations by $y^e_{t,t+1}$, and $\tilde{\pi}^e_{t,t+1}$). Based on (29), we use the following equation for identifying sentiments:

$$\begin{pmatrix} \tilde{y}^e_{t,t+1} \\ \tilde{\pi}^e_{t,t+1} \end{pmatrix} = \gamma\delta(I - \delta A)^{-1} \begin{pmatrix} s^y_t \\ s^\pi_t \end{pmatrix} \tag{32}$$

We estimate sentiments by inverting (32), i.e.

$$\begin{pmatrix} s^y_t \\ s^\pi_t \end{pmatrix} = \frac{1}{\gamma\delta}(I - \delta A) \begin{pmatrix} \tilde{y}^e_{t,t+1} \\ \tilde{\pi}^e_{t,t+1} \end{pmatrix} \tag{33}$$

We then use the obtained sentiment series for estimating the VAR (30) to obtain the innovations η^y_t and η^π_t.

It becomes clear from the comparison of (32) and (33) that γ is irrelevant for the calibration of impulse responses. It only affects the scaling of sentiments, and this effect is completely undone in the reaction of output and inflation to sentiments. Suppose that there is a purified output and inflation expectations change by 1 percentage point. Suppose that $\gamma = 0.1$. Thus, agents perceive sentiments as only 10 percent informative. Thus, in (re)constructing the latent sentiment series, these are blown up a factor of 10, since only very little of their variation is reflected in purified expectation data. For the very same reason, when it comes to the reaction of equilibrium output and inflation, only 10 percent of the variation of sentiments affect the variation of output and inflation. Hence, we can set $\gamma = 1$.

Since our purification regressions do not include any demand shocks, the latter are implicitly included in our sentiment and sentiment innovation measures. Unfortunately, typical demand shocks such as monetary and fiscal shocks are considerably harder to identify from the data than supply shocks (Brennan et al., 2024; Ramey, 2016). Moreover, common approaches that use model-based restrictions to estimated demand shocks, such as Gali (1999), would identify our sentiments as demand-shocks to the degree that they have no persistent effect on productivity-related variables. Their identification would thus not be valid if our model represented the true data-generating process. For robustness, we consider estimations that include federal government expenditures in the purification regressions as a crude measure of demand factors.[13] This affects our estimations only in minor ways.

In spite of the difficulties to filter demand factors, we are not particularly concerned about them. To the degree that demand factors are an important determinant in expectation data, we would expect to find a positive correlation between $\tilde{y}^e_{t,t+1}$ and $\pi^e_{t,t+1}$, when purified only for supply shocks, since the remaining shock should be mainly demand shocks. The same applies to the innovations η^y_t and η^y_t. However, for estimations that purify for TFP and IST shocks, the respective correlations amount to -0.34 for our measures of $\tilde{y}^e_{t,t+1}$ and $\pi^e_{t,t+1}$, and to -0.37 for η^y_t and η^y_t. There are two possible conclusions for this. First, demand factors may not affect expectation data to a noticeable degree. Second, demand factors may indeed have a positive effect on the mentioned correlations. This would then mean that additional purification by demand shocks would the, in effect, lead to (even stronger) negative correlations between sentiments, or their innovations. In this sense, with no purification for demand shocks, our results are based on a lower-bound of the correlation between output and inflation sentiments, and on $Cov(\eta^y_t, \eta^\pi_t)$, in particular.

Identifying sentiments and their innovations based on (33) also requires assumptions about δ, as well as about the elements of A (while, as discussed above, γ drops from the calibration). For δ, we follow Gabaix (2020) and set it to 0.8. For A we use standard parameter values. This includes $\sigma = 1$ (log-utility); $\varphi = 5$, such that the the Frisch elasticity for labor supply amounts to 0.2; $\beta = 0.99$ for the discount factor; $\epsilon = 9$ for the elasticity of substitution of consumption good varieties (which implies a steady state markup of 12.5%); $\theta = 0.75$ for the degree of price stickiness (implying an average price duration of 4 quarters); and $\alpha = 0.25$. For the monetary policy rule, we set $\phi_\pi = 1.5$ and $\phi_y = 0.125$.

We now turn back to the VAR estimations for sentiments and their innovations. The starting point for these estimations are the expectation series, purified for TFP and IST shocks.[14] Their correlation amount to -0.34, with significance at the 1 percent level.

[13]We use real government consumption expenditures and gross investment (GCEC1).

[14]For output expectations, we use the *Relative* measure from the MSC. To make this measure commensurate to the log-linear units of output in the New Keynesian model, we take the log of the ratio of *Relative* to its mean, which, in analogy to y_t, we loosely interpret as deviations from steady state expectations. We then purify the remaining series by obtaining residuals from regressions on TFP and IST measures. For inflation, we simply take deviations from the means of the *Mean* measure in the MSC and then purify with the same procedure. It turns out that both the raw level, logs, and purified logs of output expectations all exhibit much higher variation compared to the respective inflation expectation measures. For instance, for

However, this negative correlation need not necessarily come from sentiments themselves. Rather, equation (33) shows that A (or rather the geometric sum $I + \delta A + \delta A^2 + \ldots$) may lead to a negative correlation between sentiments, provided that there is an off-diagonal entry in A. Indeed, when we identify sentiments with the parameter values stated above, We obtain a correlation between sentiments of 0.02 that is not statistically significant from zero. This is due to the (1,2) element in A which reads as $1 - \beta\phi_\pi$, which is negative for $\phi_\pi = 1.5$. This leads to a "shear" movement in the linear transformation associated with $(I - \delta A)^{-1}$ (as well as its inverse) that switches the sign of the correlation when going from sentiments to purified expectations, and vice versa. It turns out that the sign switch disappears when ϕ_π amount to 1.3 or less. In particular, for $\phi_\pi = 1.2$, we obtain a correlation of sentiments of -0.25 which is significant at the 1-percent level. However, such a value for ϕ_π certainly ranges at the lower end of confidence intervals for empirical estimates, at least for the post-Volcker era (see Carvalho, Nechio, and Tristao (2021)).

However, the negative correlation of sentiments is only lost at first sight, when using standard value of ϕ_π. Once we estimate a first-order VAR, the negative correlation reappears in the VAR innovations η_t^y and η_t^π. For the above stated parameter values, we obtain a negative correlation of -0.37, which is significant at the 1-percent level. The reason that the negative correlation of the sentiment *innovations* does no longer materialize at the level of sentiment *states* is that the estimated coefficient matrix of the VAR, $\hat{\Phi}$, is a contraction matrix that shrinks the correlations towards zero.[15] The estimates of the main autoregressive coefficients are 0.69 and 0.83 for output and inflation sentiments, respectively. Thus, both series are relatively persistent. The off-diagonal terms in $\hat{\Phi}$ are both small, such that there are no important "cross-effects" from the inflation sentiment in the previous period to output sentiment in the subsequent period, and vice versa. The

the purified series, the standard deviation of output expectations exceeds those of inflation expectations by a factor of 5. We find it plausible that, to a large degree, this reflects noise rather than anything substantive about output expectations. There are several reasons for this. First, the answers to the output expectation question are categorical and the *Relative* measure is constructed from these. Second, respondents may be less familiar with "business conditions" than they are with prices, and this may lead to more noise. Third, from a more theoretical point of view, we are not aware of any (behavioral) model that could explain why individuals would exhibit a volatility in output expectations that exceeds the actual variation of output by a large factor. Based on this reasoning, we scale the standard deviation of purified output expectations down to the same level as the standard deviation of purified inflation expectations. Around this level, our results are not sensitive to changes in the scaling of the standard deviation of output expectations.

[15]The point estimates are are

$$\hat{\Phi} = \begin{pmatrix} 0.69 & 0.15 \\ 0.06 & 0.83 \end{pmatrix} \tag{34}$$

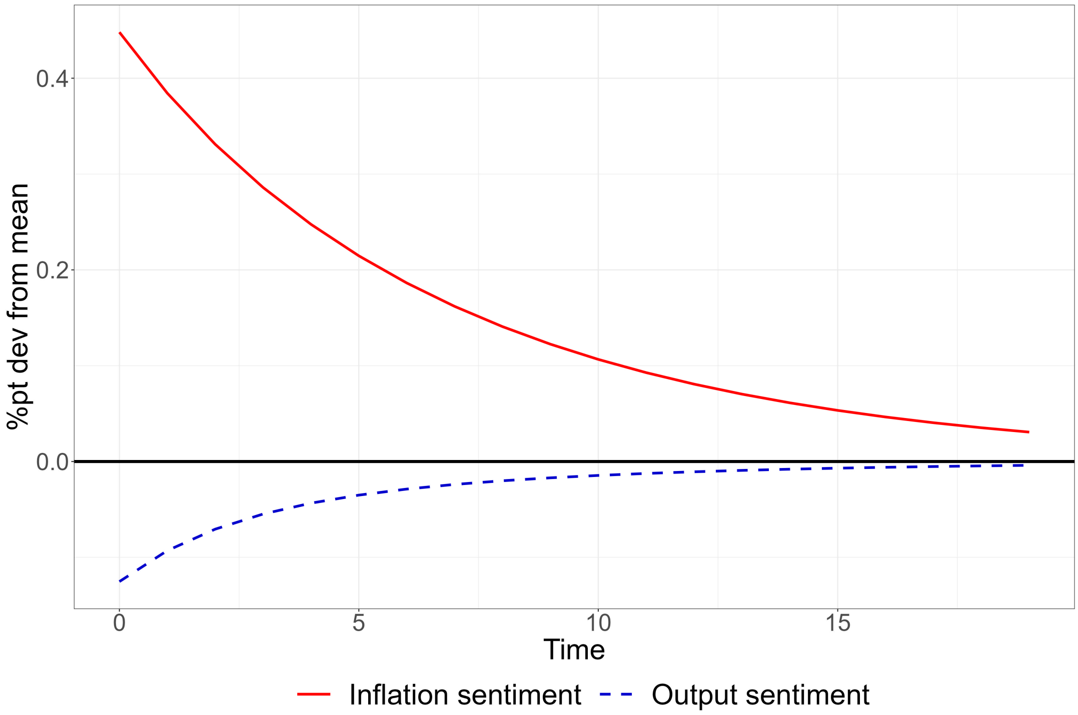
negative association between inflation and output in agents' intuitive thinking manifests itself in the contemporaneous negative correlations of the innovations. In Section 7 we show how this pattern can be explained by diagnostic expectations.

To obtain the impulse responses for an inflation sentiment shock of typical size, we consider a shock $\Delta\eta_t^\pi$ that amounts to one standard deviation of the estimated η_t^π series. Using (31), we obtain the initial reaction of the output sentiment. The impulse responses for sentiments for the subsequent periods are then simply determined by (30). The equilibrium response of output and inflation is then obtained by pre-multiplying the sentiment responses by $\gamma\delta A(I-\delta A)^{-1}$, which follows from (28).

To prevent a potential confusion, let us clarify that we set $a_t \equiv 0$ for the calculation of impulse responses for inflation sentiments. This point may not be fully obvious at first sight, since what agents actually try to achieve is estimating expected values of future fundamental states (see Section 4). Since we only explicitly included a_t as a fundamental state, it may appear that TFP somehow may "color" the impulse responses for sentiment shocks. This is however not the case. We would obtain exactly the same impulse response functions for an inflation sentiment shock, regardless of what type of fundamental shock is included in the model. The fundamental shock would indeed affect the output if the matrix B_a (or any other version of B) were to appear in the sentiment term on the right-hand side of (28). However, any such term is actually absent, and since neither A nor any other parameter is linked to a_t (or any other fundamental shock), there is no factor that would imprint the impulse response to an inflation sentiment shock a shape that resembles impulse responses for TFP (or other fundamental) shocks.

Impulse responses for sentiments are shown in Figure 3. The solid red line shows inflation sentiments, the dashed blue line shows the reaction of output sentiments. The time units are quarters. The initial impulse of inflation sentiment is set equal to a standard deviation of inflation sentiment *innovations*, i.e. η_t^π. This amounts to 0.0039, which can be interpreted as leading to an increase of s_t^π from its mean to 0.39 percentage points above the mean. The initial reaction of output amounts to -0.0026, or a reduction from the mean by .26 percentage points. We thus see the negative correlation between inflation and output sentiment innovations. As we discuss in the next Section, this may reflect an underlying subjective model about the economy according to which higher inflation is associated with lower output. For now, we stay agnostic about the underlying reason. We

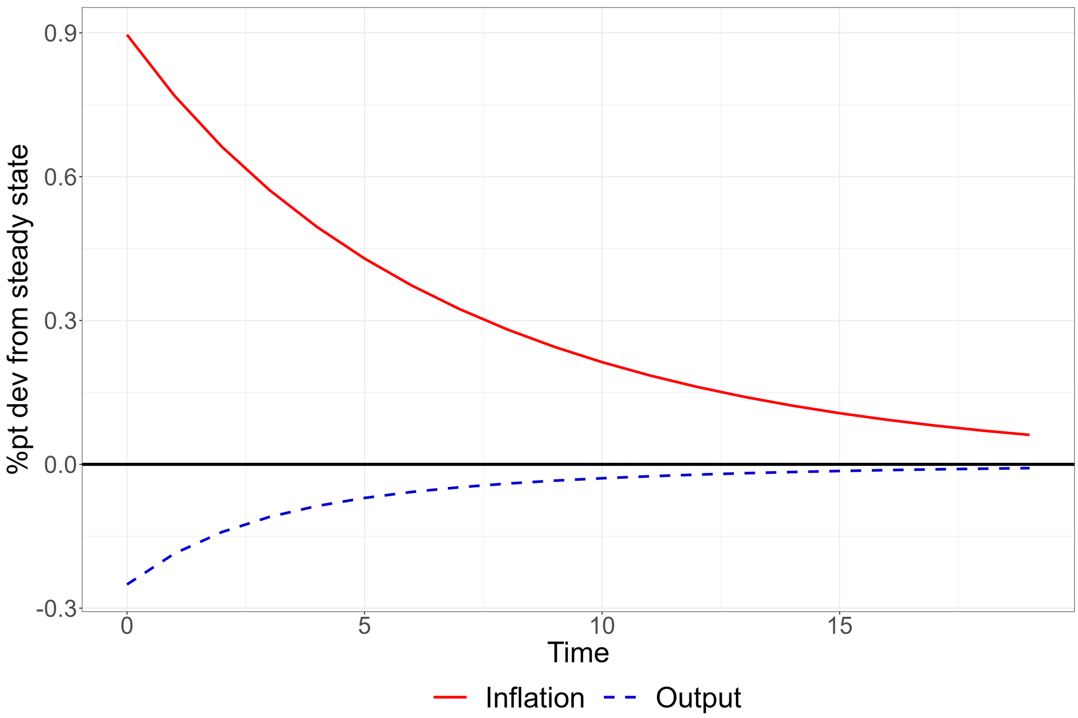
Figure 3: Impulse response for sentiments



take the negative association simply as emerging from expectation data, where our DSE model, and the structure of the New Keynesian model in which DSEs are embedded, have been used to identify sentiment components.

What we ultimately are interested in is how a sentiment impulse translate into impulse responses for the endogenous variables output and inflation. The results are shown in Figure 4. The impulse responses for inflation are shown by the solid red line, the output reaction is depicted by the dashed blue line. The time units are quarters. The units for output and inflation are percentage point deviations from a zero inflation/growth steady state. Upon impact, inflation increases by 0.63 percentage points, while output decreases by 1.03 percentage points. After one year, inflation is still at 0.43 percentage points, and output at -0.41. The cumulative effect over the first four quarters is 2.11 percentage points for inflation and -2.72 percentage points for output.

Figure 4: Impulse response for sentiments



7 Microfounding the Negative Association between Output and Inflation Sentiments

The analysis in Section 6 has shown that the negative association between output and inflation sentiments is particularly evident at the level of innovations in the (actual) law of motion of sentiments, i.e. η_t^π and η_t^y. In this Section, we present a simple model of associative memory and diagnostic expectations that can explain this pattern. Our aim is twofold. First, we show how the concept of sentiments can be linked to specific models of intuitive cognition. Second, we show how models of associative memory and diagnostic expectations can be used to gain insight into how individuals' intuitive thinking perceives relationships between different macroeconomic variables.

Following Bordalo et al. (2023), we assume that intuitive cognition centers around hypothesis formation based on agents' memory database. As in Bordalo et al. (2023), we assume that the entries in this database take on binary values. This means that, at the individual level, entries about inflation and output are represented as "high inflation" vs. "other levels of inflation", or "low output growth" vs. "other levels of output growth. The reason why it is natural to assume that individuals are particularly alert to "high" inflation (vs. "other" levels) is the empirical finding that inflation appears as *the* main source of worry among economic variables (Binetti et al., 2024; Hou & Wang, 2024; Stantcheva, 2024). In turn, an indicator for low output also appears natural, as this indicates another worrisome economic outcome and there is robust empirical evidence that negative events or news get more attention than positive ones (D'Acunto et al., 2021; Sias, Starks, & Turtle, 2023). Importantly, however, our main conclusion is robust to adding further binary indicator for low inflation and high output growth (vs. other levels).

In a nutshell, at the individual level, our model works as follows. Agents' observe a binary version of η_t^π, which acts as a cue about inflation. Agents' intuitive thinking then considers hypotheses about output growth and evaluates them in light of the inflation cue. In particular, agents accept the hypothesis that comes with the highest "diagnostic value", as defined below, which in turn, determines intuitive output expectations. The main purpose of our model is to show that, under plausible assumptions, this process implies a negative correlation between inflation and output sentiment innovations in a straightforward way.

Our empirical measures of inflation and output expectations come from the MSC, where we use aggregate data (in the form of means or shares, see Section 3). Clearly, aggregate expectations do not appear binary.[16] At the end of this section, we therefore show how the means of individual intuitive expectations take on continuous values when the individual susceptibility for what is perceived as "high" inflation is heterogeneous and continuously distributed. This element of heterogeneity is sufficient to perfectly mask the binary nature of intuitive expectations at the individual level. Importantly, the aggregate measures of intuitive expectations preserve any negative correlation between intuitive inflation and

[16]At least in raw form, expectations do also not appear binary a the individual level. There are, however, several obstacles to uncover a potential binary structure even if it were part of the true data-generating process. First, the expectation data, as they appear in a survey, include both rational and intuitive components of expectations. The rational part is naturally continuous. The New Keynesian framework that we use to estimate sentiment components is also continuous, such that the resulting components are again continuous. This would also hold at an individual level. Second, the binary values of inflation and output expectations may be heterogeneous. Since the panel dimension of the MSC is somewhat weak, identifying a binary structure of individual intuitive expectations is not straightforward. Third, output expectations are asked with a categorical question format ("better", "same", "worse"), which adds further difficulties for getting deeper insights into individual intuitive expectation components. Finally, the diagnostic expectation dimension is most likely not the only one contributing to intuitive expectations. There are other contributions that, in sum, act as if adding noise, again making it difficult to discover an underlying binary structure. We therefore proceed in the opposite direction and show that a continuous distribution of sentiment innovation is obtained at the aggregate level (even) if individual intuitive expectations are binary, but there is heterogeneity associated with the expectation formation process (see main text).

output expectations that are present at the individual level.

There is a final point we would like to discuss before proceeding with a formal model. With respect to the New Keynesian model in the previous section, one may wonder why we provide a microfoundation of intuitive expectations at the level of sentiment textitinnovations η_t^π and η_t^y, rather than at the level of the sentiment *states* s_t^π and s_t^y. The reason is that, from a conceptual point of view, our DSE model can capture a wider array of intuitive expectation formation processes if we specify sentiments in VAR form. The correlation structure of sentiments at the level of the states, s_t^π and s_t^y, depends on the innovations, and the dynamics of the VAR, captured by Φ, as well as on the matrix A. In our case, we estimate Φ from expectation data, and A is determined according to standard calibrations for a New Keynesian model. With this "identification", the negative correlation between intuitive inflation and output expectations appears in sentiment *innovations*. At the level of the states, it gradually disappears due to the properties of the matrix Φ which shrinks the influence of the innovations. This is due to the fact that the VAR is stationary, and also due to an element of "shearing" transformation associated with the matrix Φ. For our parameter choices, the overall negative correlation between inflation and output expectations (including both intuitive and rational components) is due to the matrix A (see the discussion in Subsection 6.2). Note however, that the only way that such a negative correlation can appear, even after filtering for supply shocks, is the presence of any non-fundamental terms at all, however they are correlated. There needs to be a term that the linear transformation represented by A can act upon. In a FIRE model, sentiments would be identical to zero and hence there would be no such term that A can transform into negatively correlated expectations.[17] From a more cognitive perspective, the innovations η_t^π and η_t^y capture an immediate response to cues from the external environment. The VAR process captures that these innovations are then integrated into (further) internal "mental states" – in our case captured by the VAR process – and it is the integrated states that determine intuitive expectations – or sentiments – as a whole.[18]

[17]As discussed in Subsection 6.2, the negative correlation between intuitive inflation and output expectations would also materialize at the level of sentiment states if the inflation coefficient in the Taylor rule, $\phi_p i$, were sufficiently low, such that the (1,2)-element of A would be positive, or still negative but sufficiently close to zero.

[18]See Wachter and Kahana (2024) for a fully-fledged dynamic model of memory and associated mental states.

7.1 Associative Memory and Diagnostic Expectations at the Individual Level

Let $\eta_t^{\pi,b} \in \{\underline{\eta}^{\pi},\ \overline{\eta}^{\pi}\}$, $\underline{\eta} < \overline{\eta}$, denote the binary version of η_t^{π}. This is understood as an indicator for whether agents got exposed to cues associated with high inflation ($\eta_t^{\pi,b} = \overline{\eta}$), or not ($\eta_t^{\pi,b} = \underline{\eta}$). Examples of such cues include salient changes in grocery, gasoline, and other prices (Coibion & Gorodnichenko, 2015b; D'Acunto et al., 2021; D'Acunto & Weber, 2024), or whether there has been a spread of narratives that inflation may increase (Andre, Roth, Haaland, & Wohlfart, 2021; Shiller, 2019). Importantly, under FIRE, the predictive power of such events for inflation and output would be low. Formally, we consider the limit case where it approaches zero, such that the respective cues only matter in the domain of intuitive thinking, or "speak to the gut". In this subsection, we assume that all agents are identical, such that they all react to the cue in the same way. We also assume that $\eta_t^{\pi,b}$ directly represents the innovations to inflation sentiments (which are thus identical to the reaction to the cue). In Subsection 7.2, we consider the case of heterogeneity in susceptibility to cues; there, we will need to distinguish between cues and $\eta_t^{\pi,b}$, as the latter become a function of the former. Nevertheless, we will continue using the term (binary) "inflation sentiment innovations" for $\eta_t^{\pi,b}$ and (binary) "output sentiment innovations for $\eta_t^{y,b}$. Strictly speaking, this represents an abuse of terminology since it is η_t^{π} and η_t^{y} that are defined as innovations associated with the VAR process for sentiments, which are continuous. In this sense, $\eta_t^{\pi,b}$ and $\eta_t^{y,b}$ (the latter to be introduced below) are not innovations in a proper sense. However, in the context of this analysis, we find this terminology convenient and there is little danger of confusion. We will clarify the dependence between the binary and proper continuous version of sentiment innovations in Subsection 4.4. In this Subsection, we ignore the issue and treat $\eta_t^{\pi,b}$ and $\eta_t^{y,b}$ as if they were proper innovations to sentiment states that follow a VAR.

Agents' memory consists of a database containing the history of actual inflation and output in binary indicator form. Denote the binary version of inflation by $\pi_t^b \in \{\underline{\pi},\ \overline{\pi}\}$, $\underline{\pi} < \overline{\pi}$, and the binary version of output by $y_t^b \in \{\underline{y},\ \overline{y}\}$, $\underline{y} < \overline{y}$. Being exposed to $\eta_t^{\pi,b}$, agents' form hypotheses about the (binary) output level that is cognitively associated with $\eta_t^{\pi,b}$, according to their memory database. We assume that agents accept the hypothesis that appears with the highest "diagnostic value" in their memory database.

Based on Bordalo et al. (2023, 2018), we define the diagnostic value of a hypothesis H

Table 1: Diagnostic assessment of hair color

	Red hair	Other hair color
Irish	10%	90%
Rest of the World	1%	99%
Diagnostic value	10	0.91

NOTE: Taken from Bordalo et al. (2018).

as follows. Let $V \in \{v, -v\}$ and $W \in \{w, -w\}$ denote two binary features (or "variables") in agents' memory database. Here, $-v$ is understood as "not v", and similarly for $-w$. A hypothesis $H(W|v)$, about the value of W, given a realization of V, either concerns the outcome $W = w$, or $W = -w$. The diagnostic value of a hypothesis $H(W = w|V = v)$ is defined as the following likelihood ratio:

$$DV(H(W = w|V = v)) = \frac{f(W = w|V = v)}{f(W = w, V = -v)}, \tag{35}$$

where f denotes the relative frequency of the respective outcomes in agents' memory database. The diagnostic value thus indicates the likelihood of w in the presence of v, relative to the likelihood in the presence of $-v$. Intuitively, the diagnostic value of w is high for v if w is substantially more common under v than under $-v$. However, in absolute terms, $f(W = w|V = v)$ may still be rather low. It is in such cases where judgment based on diagnostic value can lead to severe biases.[19]

We assume that the agents' intuitive reasoning adopts the following "diagnostic" decision rule about alternative hypothesis about W, given $V = v$:

$$\textit{Accept } W = w \;\; \textit{if and only if } DV(H(W = w|V = v)) > DV(H(W = -w|V = v)) \tag{36}$$

The psychological rationale behind (36) comes from associative memory (Bordalo et al., 2023). An outcome with a high diagnostic value is easily remembered. When thinking about W, w easily comes to mind when cued by v and is thus highly "available" for assessing its relative frequency. The latter is thus typically overestimated, especially when it relates to a relatively rare outcome.

A nice example, taken from Bordalo et al. (2018) and shown in Table 1, illustrates how

[19] A high $DV(H(W = w|V = v))$ also relates to the representativeness heuristic (Bordalo et al., 2023, 2018). Using the respective terminology, w is representative for v exactly when the diagnostic value, as defined above, is high.

deciding for or against a hypothesis based on diagnostic value works. Here, V refers to a population group, either Irish, or from the rest of the world; W refers to hair color. With the prospect of meeting an Irish person, an individual may form the two hypotheses $H(red|Irish)$, i.e. "the Irish person has red hair, vs. $H(other|Irish)$, i.e. the Irish person has hair of another color". If the individual follows (36), he or she considers the diagnostic values in the bottom row of Table 1. $H(red|Irish)$ comes with a much higher diagnostic value than $H(other|Irish)$, so the individual accepts $H(red|Irish)$ and hence expects the Irish person to have red hair. We refer to this expectation formation as diagnostic expectations.[20] As a shortcut, we also write that red hair is "diagnostic" for the Irish.[21]

We now turn back to the case of output and inflation. We first show that, under the assumption that low output is diagnostic for high inflation (and vice versa), innovations to inflation and output sentiments are indeed negatively correlated. We then show that the assumed diagnostic patterns are indeed present in historical output and inflation data in the US and that this pattern is very robust.

To complete the model of the relationship between $\eta_t^{\pi,b}$ and $\eta_t^{y,b}$, we still need to specify the last term, which represents the binary version of η_t^y. We assume that it is determined according to (36), i.e. by the value of y_t^b that leads to the highest diagnostic value:

$$\eta_t^{y,b} = \arg \max_{y_t^b \in \{\underline{y}, \overline{y}\}} DV(y_t^b | \pi_t^b = \eta_t^y), \tag{37}$$

We further assume that

$$DV(y_t^b = \underline{y}|\overline{\pi}) > DV(y_t^b = \overline{y}|\overline{\pi}) \text{ and } DV(y_t^b = \overline{y}|\underline{\pi}) > DV(y_t^b = \underline{y}|\overline{\pi}). \tag{38}$$

This means that, in the memory database, low output growth is diagnostic for high inflation, and "other" output growth is diagnostic for "other" inflation levels. The rationale for this assumption is that this is what what we observe as a robust pattern in the data, as discussed below.

[20] All concepts discussed above are consistent with Bordalo et al. (2023), however, their main application concerns a more specific form of expectations in a setting of autoregressive series. They, and the subsequent literature, mostly use the term "diagnostic expectations" for this more specific case.

[21] Refering to the representativeness heuristc, the statement that red hair is "representative" for the Irish has exactly the same meaning. Moreover, we may also see red hair as a *stereotype* for the Irish (Bordalo, Coffman, Gennaioli, & Shleifer, 2016).

Table 2: Diagnostic expectation about output growth

	Low output growth	Other output growth outcomes
High inflation	43%	57%
Other inflation outcomes	19%	81%
Diagnostic value	2.24	0.71

Assumption (38) implies a perfect negative correlation between $\eta_t^{\pi,b}$ and $\eta_t^{y,b}$. To see this, note that there are only two possible outcomes in the joint distribution of $(\eta_t^{\pi,b}, \eta_t^{y,b})$, either $(\overline{\pi}, \underline{y})$, or $(\underline{\pi}, \overline{y})$. Assuming that both outcomes occur with a strictly positive probability, it follows that $\underline{\pi} < E\pi < \overline{\pi}$, and $\underline{y} < Ey < \overline{y}$, since expectations always lie between the minimum and maximum value, and since we have assumed that the respective minimum values are strictly smaller than the maximum values. Hence, the covariance between $\eta_t^{\pi,b}$ and $\eta_t^{y,b}$ must be negative. Moreover, since there are only two possible outcomes, and since inflation sentiment innovations are high if and only if output sentiment innovations are low, the correlation must amount to -1. We state this result in following proposition.

PROPOSITION .3 *Assume that (38) holds, i.e. low output growth is diagnostic for high inflation, and vice versa. Then* $cor(\eta_t^{\pi,b}, \eta_t^{y,b}) = -1$.

Clearly, this result strongly depends on the two assumptions that: (i) associative memory and intuitive thinking (approximately) rely on binary features; (ii) low output growth is indeed diagnostic for high inflation. The first assumption is discussed in Bordalo et al. (2023). Concerning the second assumption, Table 2 shows a diagnostic table for binary inflation and output growth data for the US, akin to Table 1. Inflation is measured based on quarterly data for the CPI (CPIAUCSL). Output growth is calculated from quarterly values the real GDP series (GDPC1). A high inflation outcome is defined as one that falls into the top quartile of the historical distribution. Similarly, a low outcome for output growth is one that falls into the bottom quartile of the distribution. The bottom row Table 2 shows that it is indeed low output growth that comes with the highest diagnostic value for high inflation. For the case of other inflation outcomes, the diagnostic ratios are the inverses of those shown in Table 2, such that other output growth outcome is most diagnostic in that case. Thus, Table 2 is fully consistent with (38).[22] To our knowledge,

[22]We have checked the robustness of this pattern to changes in cutoff levels for the definition of "high" and "low", and also to a finer categorization that includes indicators for "tail" outcomes, such as for low inflation or high output growth.

this paper is the first pointing out the diagnostic nature of low output for high inflation, and linking this observation to empirical inflation and output expectations.

7.2 Associative Memory and Diagnostic Expectations at the Aggregate Level

Our empirical measures of η_t^π and η_t^y do not appear binary but continuous. They are derived from aggregate measures such as means or shares of individual answers to MSC survey questions (see Section 3). This raises the question how binary sentiment innovations at the individual level can be compatible with continuous innovations at the aggregate level. In this Subsection, we introduce a single element of heterogeneity and show that this is sufficient for completely masking the binary nature of sentiment innovations at the individual level. In particular, we assume that cues are common to all individuals. However, the susceptibility for associating them with "high" inflation is heterogeneous and continuously distributed.[23]

With heterogeneous susceptibility, we need to introduce new notation to make some concepts explicit that were implicit in the binary version with no heterogeneity. Let $\kappa_t \in \mathbb{R}$ denote the inflation cue in period t. We assume that this cue is drawn from a continuous *iid* random variable with distribution function F_κ. Define $\eta_{t,i}^\pi \in \{\underline{\pi}, \overline{\pi}\}$ and $\eta_{t,1}^y \in \{\underline{y}, \overline{y}\}$ as in the previous subsection, except that they are now agent-specific. Let $\vartheta_i \in \mathbb{R}$ denote a threshold of agent i such that $\eta_{t,i}^\pi = \overline{\pi}$ if and only if $\kappa_t > \vartheta_i$. This is how we model heterogeneity in the susceptibility to inflation cues. Note that the cue κ_t is common across agents. The thresholds ϑ_i are drawn from a continuous distribution with distribution function F_ϑ. We continue to assume that (37) and (38) apply.

What we are interested in here is the distribution of the mean of $\eta_{t,i}^\pi$, and $\eta_{t,i}^y$ over agents i. Consider the case that $\kappa_t = \kappa'$, for some $\kappa' \in \mathbb{R}$. Then, for all agents with $\vartheta_i < \kappa'$, we have $\kappa_t > \vartheta_i$ and hence $\eta_{t,i}^\pi = \overline{\pi}$. Thus, there is a share $F_\vartheta(\kappa')$ of agents with high

[23] To be clear, we do not mean that measures of individual sentiment innovations were truly binary even if we had suitable data to estimate them well. In reality, there are many other possible elements that influence sentiment innovations apart from diagnostic expectations, all of them sharing the property that they would have no predictive power about equilibrium outcomes in a FIRE world. As a sum of many independent influences is often best captured by a normal distribution, we would expect overall sentiment innovations at the individual level to show pronounced elements of normality. However, if the assumptions made in the previous subsection have any merit, and we believe they do, we would expect sentiment innovations to also contain a characteristically binary component.

sentiment innovations. Similarly, the share of agents with $\eta^{\pi}_{t,i} = \underline{\pi}$ is $1 - F_{\vartheta}(\kappa')$. Aggregate inflation sentiment innovations are therefore given by $\eta^{\pi}_{t} = F_{\vartheta}(\kappa')\overline{\pi} + (1 - F_{\vartheta}(\kappa'))\underline{\pi}$. The type of the distribution of η^{π}_{t} does not depend in any way on $\underline{\pi}$ and $\overline{\pi}$. To get insights into the general form of this distribution, we may therefore set $\underline{\pi} = 0$ and $\overline{\pi} = 1$. With this, we have

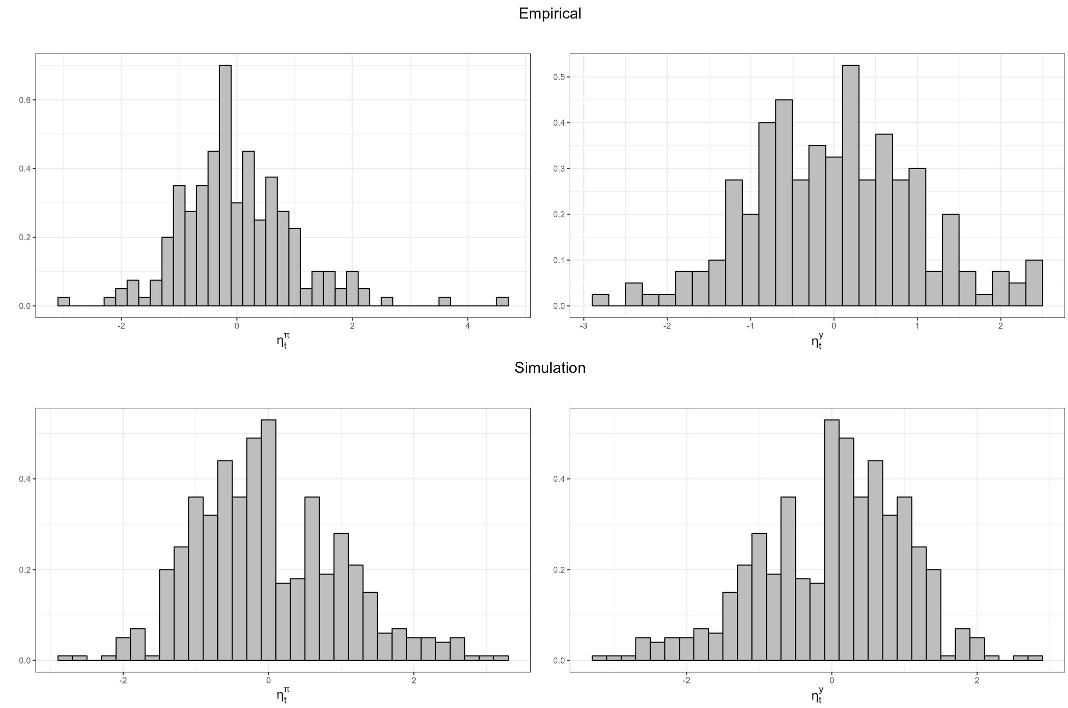
$$\eta^{\pi}_{t} = F_{\vartheta}(\kappa_t). \tag{39}$$

In other words, the random variable η^{π}_{t}, which represents aggregate inflation sentiment innovations, is a function of the random variable κ_t, with the functional relationship given by the distribution function of ϑ_i.

To gain some insight into the implications of (39), consider the case that $F_{\vartheta} = F_{\kappa}$, i.e. the thresholds and cues share the same distribution. At this point, F may represent any continuous distribution. Note that the equality of the distribution functions does not mean equality of outcomes. In our application, the latter would even be meaningless since ϑ_i is not actually understood as a random variable but as a distribution of thresholds over agents. By contrast, κ_t is understood as a random variable in the usual sense. With equality of the two distributions, we may now write (39) as $\eta^{\pi}_{t} = F_{\kappa}(\kappa_t)$. To this, the probability integral transform applies: η^{π}_{t} is equal to the value of the distribution function of a random variable with the very same distribution function. It follows that η^{π}_{t} is distributed uniformly on $[0, 1]$. It is easy to show that if inflation expectations take on the values $\underline{\pi}$ and $\overline{\pi}$, η^{π}_{t} is uniformly distributed over $[\underline{\pi}, \overline{\pi}]$. Concerning output sentiment innovations, the same logic applies. In particular, we have $\eta^{y}_{t,i} = \underline{y}$ if and only if $\eta^{\pi}_{t,i} = \overline{\pi}$. Moreover, Proposition .3 continues to apply at the aggregate level, such that the correlation between $\eta^{\pi}_{t,i}$ and $\eta^{y}_{t,i}$ is again -1.

When estimating sentiment innovations using the procedure explained in Subsection 6.1, their distribution is not uniform but comes closer to a truncated normal distribution. This raises the question whether we can get such a shape under reasonable assumptions about F_{κ} and F_{ϑ}. Since the cues and thresholds may be the result of many contributing factors that can reasonably be assumed to be (at least somewhat) independent of each other, a natural benchmark is the normal distribution for both cues and thresholds. Thus, let us consider $\vartheta \sim N(\mu_{\vartheta}, \sigma^2_{\vartheta})$, and $\kappa \sim N(\mu_{\kappa}, \sigma^2_{\kappa})$. In (39), the argument of F_{ϑ} is κ_t. Due to the functional form of the normal density, we have $F_{\vartheta}(\kappa_t) \equiv \Phi((\kappa_t - \mu_{\vartheta})/\sigma_{\vartheta})$, where Φ denotes the standard normal distribution function. The argument of Φ can be rewritten

Figure 5: Empirical and simulated distribution of aggregate sentiment innovations



as $(\mu_\kappa - \mu_\vartheta)/\sigma_\vartheta + \sigma_\kappa/\sigma_\vartheta\, Z$, where $Z = (\kappa_t - \mu_\kappa)/\sigma_\kappa$ represents a standard normal random variable. Equation (39) then becomes

$$\eta_t^\pi = \Phi\left(\frac{\mu_\kappa - \mu_\vartheta}{\sigma_\vartheta} + \frac{\sigma_\kappa}{\sigma_\vartheta} Z\right). \tag{40}$$

If κ_t and ϑ share the same normal distribution, then the argument of Φ reduces to Z and the probability transform applies again, such that η_t^π is uniformly distributed. For $\mu_\kappa = \mu_\theta$, but $\sigma_\kappa < \sigma_\vartheta$, the argument of Φ is smaller than Z. As a result, in the left and right tails, the values of the standard-normal density $\phi\left(\sigma_\kappa/\sigma_\vartheta Z\right)$ are smaller compared to $\phi(Z)$. Therefore, values around 0 and around 1 of $\Phi\left(\sigma_\kappa/\sigma_\vartheta Z\right)$ are less frequent compared to $\Phi(Z)$. As a result, the distribution of η_t^π takes on a shape that resembles a truncated normal distribution. If means are still equal but the order of variances is reversed, values close to zero and close to one become more likely than the rest. Furthermore, if $\mu_\kappa > \mu_\vartheta$, and variances are equal, the mass of the distribution of η_t^π shifts to the right, such that values around zero are least likely, and values around 1 most likely. The reverse happens for $\mu_\kappa < \mu_\vartheta$. If all parameters are unequal between the distribution of μ_κ and μ_ϑ, then a mixture of the discussed patterns emerges.

The standardized empirical distributions of aggregate sentiment innovations are shown in the upper panel of Figure 5, for inflation on the left, and output on the right. Note that there are only 171 observations, such that the shape of the distributions is also substantially influenced by randomness from sampling. Nevertheless, the distributions come relatively close to a truncated normal distribution. To match this empirical distribution, we thus need to assume $\sigma_\kappa < \sigma_\vartheta$. Suppose that cues κ_t emerge from surprising price increases in grocery or other shopping items (D'Acunto et al., 2021; D'Acunto & Weber, 2024). Since there are many such items, and since at least some price increases occur relatively frequently (and may bear little weight for aggregate inflation), it is plausible that the perceived variance of cues is relatively low, compared to the spread of individual thresholds.[24] If so, then $\sigma_\kappa < \sigma_\vartheta$ is a reasonable assumption indeed. The lower panel of Figure 5 shows simulated distributions for the sentiment innovations. These are obtained for $\underline{\pi} = 0$, $\overline{\pi} = 1$, $\mu_\kappa = \mu_\vartheta = 0.5$, $\sigma_\vartheta = 0.6$, $\sigma_\kappa = 0.2$, and then standardized. While we do not attempt here to estimate these parameters from the empirical distribution of sentiment innovations, it becomes clear that the binary nature of diagnostic expectations, discussed in the previous subsection, does not pose binding limitations on matching the empirical distribution of sentiment innovations when assuming that the susceptibility to cues is heterogeneous.

8 Conclusion

In this paper, we have presented a dual-system framework of expectation formation. According to this framework, estimation of future state variables are based on a subjective estimation model. There are two categories of signals, or explanatory variables, in this model. The first category consists of signals that speak to the rational mind. The second category of signals consists of elements of intuitive thinking that perturb the expectation formation of the rational mind. We show that this model can capture subjective models of the economy, in particular, a stagflationary model that associates higher inflation with lower output growth, independent of any supply shock. We also calibrate impulse responses to an inflation sentiment shock and show that it closely resembles a TFP shock, while no actual TFP shock is involved. The shape of the impulse response is caused purely

[24] This may also help to explain why, during phases of low inflation, household expectations are excessively high compared to actual realizations (D'Acunto and Weber (2024)).

by agents' subjective model.

While in the current paper the perturbations of intuitive thinking in expectations are assumed to be independent of the state of the economy, it is plausible that in practice they can be connected with current economic events. For instance, it would be natural to consider that perturbations in expectations can be triggered by hyped media reporting (such as when a technical recession is portrayed as a slump) or by the use of decisive language by central banks (like a pledge to "do whatever it takes"). To what extent would such connections amplify the impact of the very shocks that underlie the hyped reporting or the use of decisive language? What would be the implications for the conduct of monetary policy? Our framework is a flexible tool within which these and related questions can be addressed. At this point, we leave a detailed exploration of these aspects for future research.

References

Adam, K., Kuang, P., & Xie, S. (2024). *Overconfidence in private information explains biases in professional forecasts.* (Working Paper 28836)

Adam, K., & Padula, M. (2011). Inflation dynamics and subjective expectations in the united states. *Economic Inquiry*, *49*(1), 13–25.

Andre, P., Haaland, I., Roth, C., & Wohlfahrt, J. (2023). *Narratives about the macroeconomy* (Working Paper No. 10535). CESifo.

Andre, P., Pizzinelli, C., Roth, C., & Wohlfahrt, J. (2022). Subjective models of the macroeconomy: Evidence from experts and representative samples. , *89*, 2958-2991.

Andre, P., Roth, C., Haaland, I., & Wohlfart, J. (2021). Inflation narratives.

Angeletos, G.-M., Collard, F., & Dellas, H. (2018). Quantifying confidence. , *86*(5), 1689-1726.

Angeletos, G.-M., & Lian, C. (2023). Dampening general equilibrium: Incomplete information and bounded rationality. In *Handbook of economic expectations* (p. 613-645). Elsevier.

Augenblick, N., Lazarus, E., & Thaler, M. (2024). *Overinference from weak signals and underinference from strong signals.* (Working Paper)

Barsky, R. B., & Sims, E. R. (2012, June). Information, animal spirits, and the meaning of innovations in consumer confidence. *American Economic Review*, *102*(4), 1343-77.

Ben Zeev, N. (2018). What can we learn about news shocks from the late 1990s and early 2000s boom-bust period? *Journal of Economic Dynamics and Control*, *87*, 94-105.

Bhandari, A., Borovička, J., & Ho, P. (2023). *Survey data and subjective beliefs in business cycle models.* (Federal Reserve Bank of Richmond Working Paper)

Bianchi, F., Ludvigson, S. C., & Ma, S. (2022). Belief distortions and macroeconomic fluctuations. *American Economic Review*, *112*(7), 2269-2315.

Binetti, A., Nuzzi, F., & Stantcheva, S. (2024). *People's understanding of inflation* (Tech. Rep.). National Bureau of Economic Research, Working Paper 32497.

Blanchard, O. J., & Kahn, C. M. (1980). The solution of linear difference models under rational expectations. , *48*(5), 1305-1311.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. , *131*(4), 1753–1794.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2023). Memory and probability. , *138*(1), 265-311.

Bordalo, P., Gennaioli, N., Ma, Y., & Shleifer, A. (2020). Overreaction in macroeconomic expectations. , *110*(9), 2748-2782.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2018). Diagnostic expectations and credit cycles. *The Journal of Finance*, *73*(1), 199–227.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2020). Memory, attention, and choice. , *135*(3), 1399–1442.

Brennan, C. M., Jacobson, M., Matthes, C., & Walker, T. B. (2024). Monetary policy shocks: Data or methods?

Candia, B., Coibion, O., & Gorodnichenko, Y. (2020). *Communication and the beliefs of economic agents.* (NBER Working Paper 27800)

Candia, B., Coibion, O., & Gorodnichenko, Y. (2021). *The inflation expectations of u.s. firms: Evidence from a new survey.* (NBER Working Paper 28836)

Carvalho, C., Nechio, F., & Tristao, T. (2021). Taylor rule estimation by ols. *Journal of Monetary Economics*, *124*, 140–154.

Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? , *120*(1), 116-159.

Coibion, O., & Gorodnichenko, Y. (2015a). Information rigidity and the expectations formation process: A simple framework and new facts. , *105*(8), 2644-2678.

Coibion, O., & Gorodnichenko, Y. (2015b). Is the phillips curve alive and well after all? inflation expectations and the missing disinflation. *American Economic Journal: Macroeconomics*, *7*(1), 197-232.

Dräger, L., Lamla, M. J., & Pfajfar, D. (2016). Are survey expectations theory-consistent? the role of central bank communication and news. , *85*, 84-111.

D'Acunto, F., Malmendier, U., Ospina, J., & Weber, M. (2021). Exposure to grocery prices and inflation expectations. *Journal of Political Economy*, *129*(5), 1615–1639.

D'Acunto, F., & Weber, M. (2024). Why survey-based subjective expectations are meaningful and important. *Annual Review of Economics*, *16*.

Eliaz, K., & Spiegler, R. (2020). A model of competing narratives. *American Economic Review*, *110*(12), 3786-3816.

Enders, Z., Kleemann, M., & Müller, G. J. (2021). Growth expectations, undue optimism, and short-run fluctuations. *The Review of Economics and Statistics*, *103*(5), 905-921.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223-241.

Fernald, J. (2014). A quarterly, utilization-adjusted series on total factor productivity..

Gabaix, X. (2020). A behavioral new keynesian model. *American Economic Review*, *110*(8), 2271-2327.

Gabaix, X., & Laibson, D. (2017). *Myopia and discounting.*

Gali, J. (1999). Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? *American Economic Review*, *89*(1), 249–271.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle.* Princeton University Press.

Garcìa-Schmidt, M., & Woodford, M. (2019). Are low interest rates deflationary? a paradox of perfect-foresight analysis. *American Economic Review*, *109*(1), 86–120.

Gratton, C., Sun, H., & Petersen, S. E. (2018). Control networks and hubs. *Psychophysiology*, *55*(3), e13032.

Hou, C., & Wang, T. (2024). *Uncovering subjective models from survey expectations* (Tech. Rep.).

Ilut, C., & Valchev, R. (2023). Economic agents as imperfect problem solvers. , *138*(1), 313–362.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections. *The American Economic Review*, *95*(1), 161–182.

Kahneman, D. (2011). *Thinking, fast and slow.* Random House.

Kamdar, R. (2019). The inattentive consumer: Sentiment and expectations. (Working Paper)

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

Kieren, P., & Weber, M. (2024). Expectation formation under uninformative signals. *Management Science*.

Lorenzoni, G. (2009). A theory of demand shocks. , *99*(5), 2050-2084.

Montiel Olea, J. L., & Plagborg-Møller, M. (2021). Local projection inference is simpler and more robust than you think. *Econometrica*, *89*(4), 1789-1823.

Orphanides, A., & Norden, S. v. (2002). The unreliability of output-gap estimates in real time. *Review of economics and statistics*, *84*(4), 569–583.

Preston, B. (2006). Adaptive learning, forecast-based instrument rules and monetary policy. *Journal of Monetary Economics*, *53*(3), 507–535.

Preston, B. (2008). Adaptive learning and the use of forecasts in monetary policy. *Journal of Economic Dynamics and Control*, *32*(11), 3661–3681.

Ramey, V. (2016). Chapter 2 - macroeconomic shocks and their propagation. In J. B. Taylor & H. Uhlig (Eds.), (Vol. 2, p. 71-162). Elsevier.

Shiller, R. J. (2019). *Narrative economics.* Princeton University Press.

Sias, R., Starks, L. T., & Turtle, H. J. (2023). The negativity bias and perceived return distributions: Evidence from a pandemic. *Journal of Financial Economics*, *147*(3), 627–657.

Sporns, O., & Betzel, R. F. (2016). Modular brain networks. *Annual review of psychology*, *67*(1), 613–640.

Stanovich, K. E., & West, R. F. (1991). 24. individual differences in reasoning: Implications for the rationality debate?

Stantcheva, S. (2024). *Why do we dislike inflation?* (Working Paper No. 32300). National Bureau of Economic Research.

Wachter, J. A., & Kahana, M. J. (2024). A retrieved-context theory of financial decisions. *The Quarterly Journal of Economics*, *139*(2), 1095–1147.

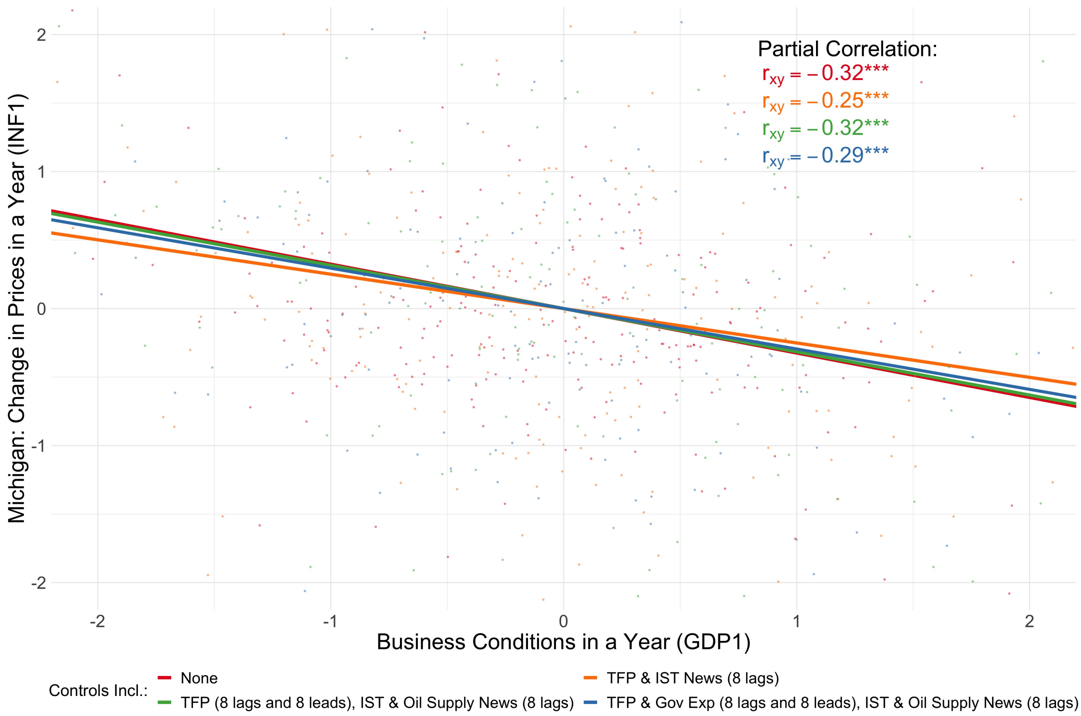
Appendix A: Further Empirical Evidence on Intuitive Thinking

The empirical evidence so far suggests that agents' perception of the relationship between output and inflation is not (exclusively) driven by its factual nature but rather (also) by an intuitive model. This raises a question that is of particular importance in this paper: Do such intuitive models help us to better understand macroeconomic fluctuations? Adapting this question to a standard macroeconomic framework leads us to ask whether *innovations* or *shocks* to intuitive models trigger impulse responses in key macroeconomic variables. As a further step in this direction, it is therefore of interest to consider a VAR estimation for the MSC series and to investigate whether traces of the same "intuitive model" that appeared in Section 3 also appear when examining VAR innovations. For technical reasons, we do not directly estimate a VAR for the purified series used in Section 3. Rather we run VAR-X specifications with the two MSC expectations as the VAR variables and the purification variables as the exogenous X-controls.[25] Below, we report results where the VAR includes one autoregressive lag for MSC expectations. We use the VAR-X estimations to extract residuals and then run regressions of the residuals for inflation expectations on the residuals for output expectations. Figure A1 shows the results for the regressions for the VAR-X residuals, or innovations. All correlations are now significantly negative at the 1-percent level, and the different purification schemes play almost no role for the value of the correlation.

Since the VAR-X innovations are correlated, it is standard practice to ask whether there is a meaningful scheme that identifies independent "structural" expectation shocks. We examine a parsimonious linear scheme. Denote the VAR-X innovations for output and inflation expectations by η_t^y and η_t^π, respectively. Denote the structural shocks that are to be identified by ε_t^y and ε_t^π, respectively. Identification then relies on finding a 2-by-2 matrix H such that $\eta_t = H\varepsilon_t$, where η_t and ε_t are vectors collecting the output and

[25] If we directly used the purified MSC series from the previous regressions, there would emerge a correlation between VAR innovations and contemporaneous or lead values of the supply side controls (such as TFP) since the (RHS) lags of the VAR variables have not been purified with literally the same contemporaneous or lead values as the (LHS) contemporaneous VAR variables. This correlation would defy the purpose of purification and lead to biases in the local projection estimates presented below.

Figure A1: Downward-Sloping NKPC in VAR(1) of MSC Forecast Errors



NOTE: The slopes of the plotted lines are based on estimated residuals from a VAR-X model for INF1 and GDP1, where the X-controls include the purification variables as explained in the main text and indicated in the legend. The red line refers to the case with no purification. Significance levels (p-values): *** $p \leq 0.01$; ** $0.01 < p \leq 0.05$; * $0.05 < p \leq 0.1$.

inflation expectation components, respectively. Writing this equation explicitly, we have

$$\begin{aligned} \eta_t^y &= h_{yy}\varepsilon_t^y + h_{y\pi}\varepsilon_t^\pi \\ \eta_t^\pi &= h_{\pi y}\varepsilon_t^y + h_{\pi\pi}\varepsilon_t^\pi \end{aligned} \tag{A1}$$

where the subscripts of the coefficients h_{ij} characterize their positions in H. Obviously, it is natural to assume that $h_{yy} > 0$ and $h_{\pi\pi} > 0$. Therefore, for the covariance between the two innovations in η_t to be negative, one or both of the off-diagonal elements $h_{y\pi}$ and $h_{\pi y}$ must be negative. The correlation is guaranteed to be negative if both off-diagonal elements are negative. In a nutshell, this could reflect an "all good/bad in one" heuristic. In such an intuitive "stagflationary" model, a high innovation to output expectation ("good") is associated with a low innovation to inflation expectation ("good"). The higher the innovation to output expectation, the lower the innovation to inflation. Hence, $h_{\pi y}$ should be negative. Similarly, a high innovation in inflation expectation ("bad") is associated with a low innovation in output expectation ("bad"), and so the same logic applies. Our identification scheme now sets

$$h_{y\pi} = \tau h_{\pi y}, \tag{A2}$$

with $\tau > 0$ exogenously fixed to a value not too far from 1. This means that, within associative memory, the cross-effect from being mentally triggered ("cued") by an observation about output to innovations for inflation expectations is not too different from the cross-effect where the two domains—output and inflation—are reversed. With this assumption, H and thus the series for the structural shocks ε_t is identified. It turns out that the value of H depends very little on τ for values between 0.75 and 1.[26]

Having identified ε_t, we estimate impulse responses by running local projections (Jordà (2005), Montiel Olea and Plagborg-Møller (2021)).[27] We do so again for all purification schemes described in Section 3. The LHS variables are contemporaneous values and leads of GDP and inflation. Both are subject to the same purification schemes as MSC expectations. More precisely, we use purified versions of contemporaneous values and leads of cyclical real US GDP and leads of quarterly US CPI growth als LHS variables.[28] To explore the robustness of the estimations, we run two different specifications of local projections, given by

$$y_{t+h} = \beta_0^{(h)} + \beta_1^{(h)} \widehat{\varepsilon}_t + \beta_2^{(h)} y_{t-1} + v_t^{(h)}, \tag{A3}$$

and

$$y_{t+h} - y_{t-1} = \beta_0^{(h)} + \sum_{k=0}^{K} \beta_{k+1}^{(h)} \widehat{\varepsilon}_{t-k} + \beta_{K+1}^{(h)} (y_{t-1} - y_{t-2}) + v_t^{(h)} \tag{A4}$$

where $h \geq 0$ indicates the lead horizon, y_{t+h} stands for the h-lead of a purified macroeconomic variable (output or inflation), and $\hat{\varepsilon}_t$ stands for the impulse variable, i.e. the estimated shocks to intuitive expectations. The first version derives from the analysis in Montiel Olea and Plagborg-Møller (2021), while the second follows the specifications used

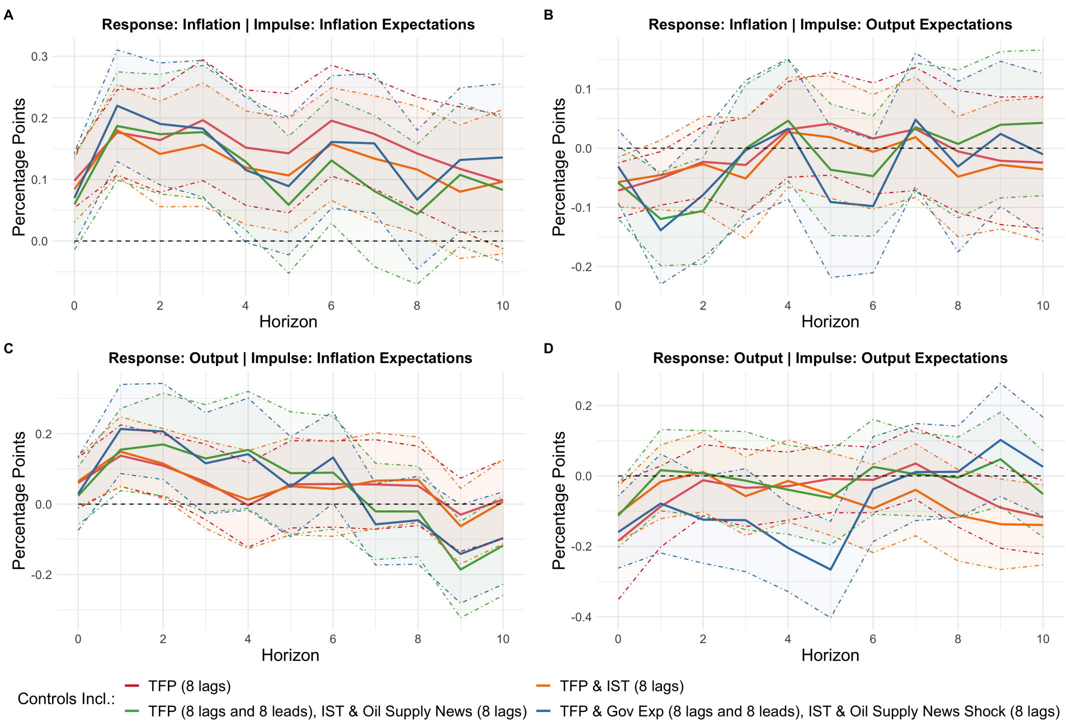
[26] With (A2), we can solve the equation $\eta_t = H\varepsilon_t$ for the elements of H, assuming unit standard deviations for ε_t. This yields three non-linear equations for the elements of H. We solve these equations and only consider solutions for which $h_{yy} > 0$, $h_{\pi\pi} > 0$, and $h_{\pi y} < 0$. For $\tau \in [0.8, 1]$, there always exist solutions meeting these requirements, and they are unique. For $\tau > 1$, no such solutions exist. The elements of H vary with τ in a very minor way. All reported results are for the case $\tau = 1$. We solve (A2) for η_t series that are not (re)standardized.

[27] In the current context, we prefer local projections to more model-driven estimates since we intentionally want to "let the data speak" as far as possible.

[28] We use the series *GDPC1* and *CPIAUCSL* from the FRED database. The cyclical portion of the variation in real US GDP is estimated using Hodrick-Prescott filtering of the logarithm of *GDPC1* ($\lambda = 1600$). Alternative procedures, such as band-pass filtering at frequencies corresponding to 6 – 32 quarters (Baxter-King or Christiano-Fitzgerald) do not materially change the results.

by Enders et al. (2021). In the latter case, we choose $K = 8$ lags for the expectation shocks.

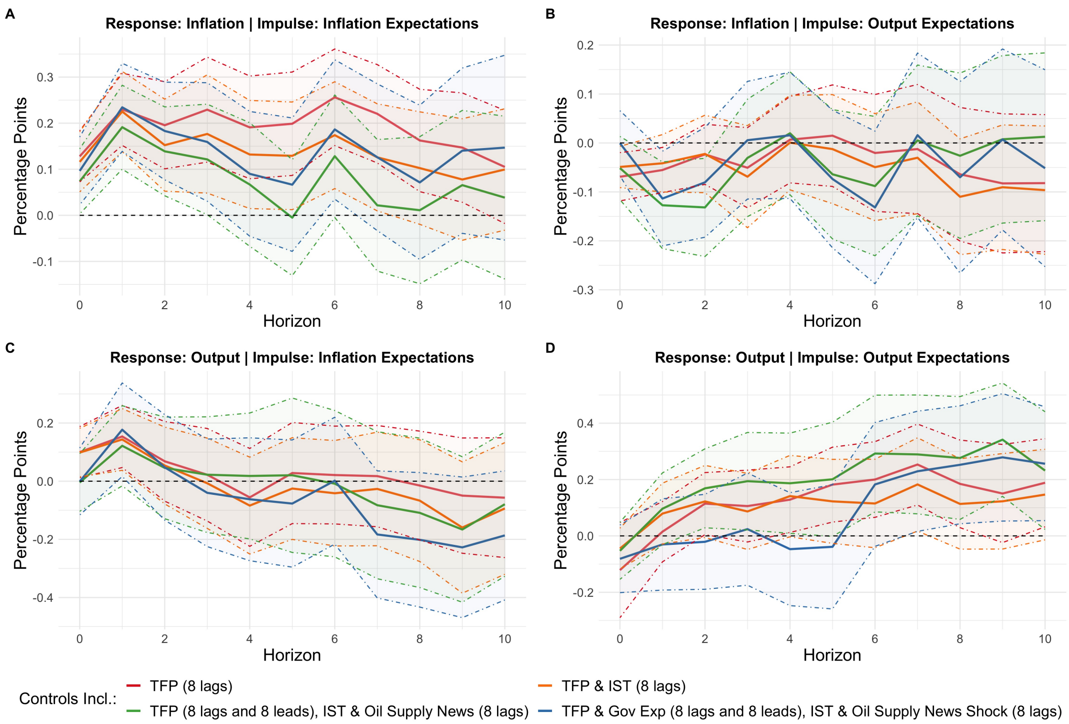
Figure A2: Linear projection estimates of structural impulse responses for specification (A3)



NOTE: The figure plots impulse responses for the linear projection specification (A3). The confidence bands indicate 90% significance regions. The magnitude of the initial impulse is one standard deviation of the corresponding SVAR residual. As an example, a value of 0.1 on the vertical axis means that a one standard deviation shock in, say, (purified) inflation expectations is associated with a 10% standard deviation higher value of, say, (purified) inflation. The number of observations in the regression for the contemporaneous impact are 243, 200, 140, and 140, respectively.

Figure A2 plots the estimated impulse responses for the specification (A3). What stands out is the pattern in Panel A: for all purification schemes, the estimate for the short-run impact of a shock to intuitive inflation expectations is significantly positive. Depending on the purification, the effects of a shock of one standard deviation of the expectation shock amounts to between 0.15 and 0.2 standard deviations of purified inflation. The patterns shown in the other panels are less clear-cut. Panel C shows a tendency for output first reacting positively to intuitive inflation expectation shocks. Output then reverts back to zero, with some specifications also showing more negative effects after 7 quarters. Panels B and D show the reaction of inflation and output to an intuitive output expectation shock. Panel B suggests that the reaction of inflation is first negative and then turns back to values close to zero. Panel D suggests that output is first negatively affected and then turns back to zero. This effect is somewhat counterintuitive; however, it is not very robust. Turning to Figure A3, we see that specification (A4) yields mostly positive estimates for the effect of intuitive output expectation shocks on output. For some purifications, these are significant at a horizon between 6 and 10 quarters. For all other cases, the patterns in Figure A3 are very similar to Figure A2. In particular, the relatively strong effect of intuitive inflation expectations on inflation in Panel A is very similar across the two specifications.[29] In sum, the estimated effects provide suggestive evidence that shocks to intuitive expectations may indeed affect actual macroeconomic outcomes.

Figure A3: Linear projection estimates of structural impulse responses for specification (A4)



NOTE: The figure plots impulse responses for the linear projection specification (A4). The confidence bands indicate 90% significance regions. The magnitude of the initial impulse is one standard deviation of the corresponding SVAR residual. As an example, a value of 0.1 on the vertical axis means that a one standard deviation shock in, say, (purified) inflation expectations is associated with a 10% standard deviation higher value of, say, (purified) inflation. The number of observations in the regression for the contemporaneous impact are 235, 192, 132, and 132, respectively.

[29] We also run many other specifications, using different structural VARs for the identification of structural shocks, different specifications for the linear projections, and different expectation measures from the MSC.