The Macroeconomic Effects of Inflation Expectations: The Distribution Matters*

Guido Ascari†
University of Pavia
De Nederlandsche Bank, CEPR

Paolo Bonomolo‡
De Nederlandsche Bank
European Central Bank

Alessandro Celani§
Örebro University

April 29, 2025

Abstract

We investigate the macroeconomic effects of shocks to the distribution of short-term inflation expectations augmenting a parsimonious monetary policy Bayesian VAR with heterogeneous expectations from the Michigan Survey. A first surprising result is that (the Michigan survey) inflation expectations do not seem to be much influenced by macroeconomic developments, while the opposite is not true. Moreover, a comprehensive density impulse response function analysis shows that it matters to take into account the whole expectation distribution. First, it matters because considering only the first and second moment of the distribution leads to an underestimation of the macroeconomic effects of expectation shocks. Second, mean and dispersion shocks are stagflationary. Third, left-tail perturbations account for the largest effect of expectation shocks on macroeconomic fluctuations. It follows that central bank communication should focus on the tails: reducing the noise/dispersion might be more effective than anchoring the mean.

Keywords: Inflation, Expectations, Distribution, Monetary Policy.
JEL classification: E3, E52, C32.

---

*We wish to thank Klaus Adam, Fabio Canova, Efrem Castelnuovo, Domenico Giannone, Yurij Gorodnichenko, Fadi Hassan, Leonardo Melosi, Francesca Monti, Ricardo Reis and Giovanni Ricco for their comments and seminar participants at the Bank of Italy, the Dutch Central Bank, the National University of Singapore, the Swiss Central Bank, and conference participants at the 17th Conference on Computational and Financial Econometrics 2023, the 2nd BdF-CEPR-HEC-PSE Paris Conference on the Macroeconomics of Expectations, Paris 2024, the EABCN Conference: New Challenges in Monetary Economics, Mannheim, 2024, the 2024 European Seminar on Bayesian Econometrics, the 2024 Workshop of the Australasian Macroeconomic Society, the Barcelona School of Economics Summer Forum 2024, the 2024 IAAE annual conference in Thessaloniki, the 2024 Sailing the Macro workshop, the 1st CAM-Risk Conferencee in Pavia. The views expressed are those of the authors and do not necessarily reflect official positions of De Nederlandsche Bank or the European Central Bank.

†Guido Ascari, De Nederlandsche Bank, Frederiksplein, 1017 XL, Amsterdam, The Nederlands and Department of Economics and Management, University of Pavia, Via San Felice 5, 27100 Pavia, Italy, Email: g.ascari@dnb.nl, guido.ascari@unipv.it.

‡Paolo Bonomolo, De Nederlandsche Bank, Frederiksplein, 1017 XL, Amsterdam, The Nederlands and European Central Bank, 60640 Frankfurt am Main, Germany. Email: paolo.bonomolo@gmail.com.

§Alessandro Celani, Örebro University School of Business, Fakultetsgatan 1, 702 81, Örebro, Sweden. Email: alessandro.celani@oru.se.

1 Introduction

What is the macroeconomic impact of a shock that changes the short-term inflation expectation distribution?

The literature on inflation expectations is at the core of macroeconomics, as inflation expectations play a crucial role in theoretical models and policy discussions. Recently, this literature thrived with important theoretical contributions and empirical analysis, thanks to the increasing availability of surveys. The large recent literature on inflation expectations takes mainly a microeconomic perspective, focusing on understanding the formation and determinants of inflation expectations.

While inflation expectations are obviously endogenous, one should not discard the possibility that they can also move exogenously or independently from macroeconomic developments. In their survey, D'Acunto et al. (2022) conclude that agents' expectations are biased and volatile in the time series, which suggests that inflation expectations might be subject to exogenous shocks. Moreover, this idea seems embedded in the way policy makers think about expectations in their speeches. These often feature discussions about the possibility of inflation expectations getting out of control, as if policy has to respond to counteract exogenous shifts in inflation expectations. This would not make sense in an environment of rational, or more general well-behaved, inflation expectation formation process.

Therefore, here we take a macroeconomic perspective and revert the common question about how expectations depend on the economic outlook to: how do exogenous shifts in expectations affect macroeconomic outcomes? The literature on surveys indeed showed that exogenous variations in inflation expectations do affect agents' economic consumption and investment decisions, and that expectations about future inflation are associated to worse expected macroeconomic outcome (e.g., Coibion et al., 2019, 2023; Weber et al., 2022). Ascari et al. (2023) show the theoretical and empirical relevance of exogenous variations of inflation expectations. Specifically, a shock that increases the average short-term inflation expectation has negative macroeconomic effects, increasing inflation and decreasing output.

However, we know that the rich information set contained in survey expectations is not limited to the consensus forecast. Whether of households, firms, or professional forecasters, we observe a pervasive cross-sectional heterogeneity. Possible explanations include substantial inattentiveness to the temporal evolution of macroeconomic aggregates, sensitivity to salient prices or personal experiences, and a variety of different biases (e.g., D'Acunto et al., 2022). This dispersion in beliefs, or disagreement (Mankiw et al., 2003), exhibits substantial time variation. Reis (2022) convincingly argues that there is important information content contained in the higher order moments of the inflation expectation distribution.

Hence, we investigate the macroeconomic effects of various shocks to the short-term inflation expectation distribution, using a Bayesian VAR (BVAR) to study the joint evolution of a number of macroeconomic aggregates and the short-term inflation expectation distribution. In order to have a sufficiently large number of survey respondents to meaningfully talk about higher moments and a long sample size, we use the Michigan Survey of Consumers (MSC).

From a methodological standpoint, we explore three distinct approximations of the large cross-sections of survey data: a discrete approximation, a continuous kernel-based density, and a parsimonious Gaussian distribution. The discrete method leverages the raw survey responses, offering an intuitive estimator of the probability mass function. The continuous kernel approximation smooths irregularities in the data and introduces continuity, accounting for the tendency of respondents to report rounded values. Finally, the Gaussian approximation provides a simplified baseline, assessing whether capturing the full heterogeneity of expectations is relevant or not for macroeconomic analysis. The discrete and continuous version are inherently high-dimensional and subject to constraints typical of probability distributions. To address these challenges, we propose a piecewise constant decomposition that aggregates probabilities over intervals, reducing dimensionality and enabling their direct use in macroeconometric models. This approach balances interpretability and computational efficiency, ensuring that the heterogeneity driving macroeconomic fluctuations is accurately captured without reliance on overly complex or restrictive assumptions. This allows us to map a broad range of heterogeneity in the expectation formation process, thus, improving from studies employing just the consensus, and contributing to the emerging literature concerned with the joint estimation of a time series of densities and macroeconomic aggregates (Chang et al., 2024; Meeks and Monti, 2023). Moreover, note that exogenous perturbation of distributional datasets requires modification in the way experiments are carried out. Another methodological contribution we make is to show how to conduct dynamic analysis when an exogenous movement of the distribution is of interest. In particular, given an initial distribution of interest obtained perturbing the steady-state distribution, we compute the implied combination of exogenous distributional shocks. Since the number of exogenous shocks to the distribution is higher that the time series involved, this task requires a structural scenario analysis in the spirit of Antolín-Díaz et al. (2021).

We use this setup to study the cross-feedback effects between expectations and the macroeconomy. Our identification assumptions are based on the natural discrepancy between the date on which the survey takes place and the release of the macroeconomic variables. A first notable, and somewhat surprising result, is that movements in expectations are relevant to explain the macroeconomic variables, but not vice versa. This result provides a strong rationale to study the main question of our paper stated at the beginning of this Introduction.

We therefore investigate how several exogenous changes in the inflation expectation distribution affect output and inflation, through a comprehensive density impulse response function analysis.

First, we show that shocks that increase the mean and the variance of the distribution are stagflationary.[1] Moreover, our analysis shows that incorporating the whole cross-sectional heterogeneity is important, as models based on just mean and variance would underestimate the macroeconomic effects of inflation expectation shocks to these moments. Regarding the possible transmission mechanism, we include the consumer sentiment variable from the MSC. This variable drops after a positive shock to the first and second moment of the distribution, signalling a bad expected future outcome by consumers, whose consequent behavior could trigger the effects on inflation and output.

Second, we look at various shocks that change the skewness and the kurtosis of the distribution. Unlike location, dispersion is a more general concept and there are many ways to perturb

[1]This is consistent with the result in Ascari et al. (2023), using different US data.

the shape of the distribution. With our approximation, we are able to generate movements related to specific parts of the distribution. We first look at symmetric shocks, showing that tails shock – that is, perturbation where more mass is distributed to the tails – are recessionary. Then, we investigate asymmetry to identify whether one part of the distribution is more relevant in inducing these effects. Our findings show that the left tail of the distribution is the driver of the negative effects on the macroeconomic variables that we found in the symmetric case. On the contrary, movements in the right tail of the distribution have negligible (and positive) effects. Coherently, consumer sentiment drastically drops after a left tail shock, while it does not react in case of a right tail shock.

These results call for an investigation about who are the households populating the left tail of the inflation expectation distribution. The MSC is a comprehensive survey, collecting inflation expectations, but also a variety of other information. It turns out that the probability of being in the left tail is positively related to have a college degree and to the level of income, suggesting that our findings are not due to poorly educated households with no spending ability.

Third, we use our framework to investigate the role of inflation expectations in shaping the dynamics of output and inflation during the last three US recessions: the 2001 episode that followed the collapse of the dot-com bubble, the 2008 Great Financial Crisis and the recent Covid-19 pandemic.

Finally, we discuss some possible implications of our results for central bank communication. Specifically, which part of the distribution should central bank communication target, assuming that the central bank could change the shape of inflation distribution through effective communication? Should communication just focus on shifting the mean of the distribution towards the 2% target? We devise a combination of mean and dispersion shocks such that the distribution on impact is anchored at 2%, with different degrees of cross-sectional heterogeneity. As long as the dispersion around the mean is relatively low, shifting the mean towards 2% has positive macroeconomic effects. However, the results change if shifting the mean is accompanied by an increase in dispersion. Then, economic conditions deteriorate substantially. Our results, therefore, suggest that, with respect to short-term inflation expectations, communication should focus on decreasing dispersion, and particularly moving mass from the left tail, where pessimistic expectations lie, rather than on moving the mean towards the 2% target.[2]

Literature. The literature on inflation expectations is vast, and there are excellent very recent surveys (i.e., Weber et al., 2022; D'Acunto et al., 2022; Binder and Ryngaert, 2024), such that there is no much point in summarizing it here. Moreover, as said, most of the literature focuses on the effects of various factors (e.g., macroeconomic developments, beliefs, personal experience, etc.) in shaping inflation expectations, while here we rather invert the causation asking whether and how exogenous changes in inflation expectations have an impact on macroeconomic variables.

Surprisingly, there are few papers in the literature investigating this question. Clark and Davig (2011) is an early work that looks at short-term inflation expectations in a VAR analysis, but it does not focus on the macroeconomic effects. Ascari et al. (2023) study the effects of a

[2]Interestingly, this echoes the discussion in Blinder et al. (2008) about the importance of central bank communication to reduce noise and thus raise the signal-to-noise ratio.

shock to the mean of short-term inflation expectations both theoretically and empirically using data from the Survey of Professional Forecasters in the US. They find that such a shock is stagflationary, as we do. Barrett and Adams (2022) analyzes the same topic with a focus on the empirical identification of such a shock, reaching opposite empirical conclusions. Neri (2023) asks what are the macroeconomic effects of changes in long-term inflation expectations in an empirical VAR on Euro area data. Given the focus on the long-term inflation expectations, Neri (2023) connects to a large literature studying the effects of 'de-anchoring'.

Contrary to the papers above, our paper focuses on the whole inflation expectation distribution, and not just on the consensus forecast. Reis (2022) was an inspiring work for our analysis. He shows that, while inflation expectations seemed anchor in 1971 by looking at the cross-sectional mean of professional forecasters' or households' expectations, looking at the whole distribution would have revealed a very different picture providing timely sign of a shift in inflation expectations. Reis (2022) provides two important takeaways for our work: (i) movements in the whole distribution of inflation expectations provide important signals about future inflation; (ii) there is a lot of information in the tails of the distribution. Motivated by these two points, our analysis reinforces and expand the Reis (2022) conclusion beyond the effects on inflation: inflation expectation distribution matters for macroeconomic dynamics.

Meeks and Monti (2023) is particularly related to our work. They use functional principal component regression to fit an augmented Phillips Curve, where the distribution of short-term inflation expectations appears on the right-hand side. They find statistical relevance of considering the heterogeneity of expectations for inflation dynamics. Chang et al. (2024) propose a State-Space model, where the dynamics of macroeconomic aggregates and log densities form the state-transition equation. The infinite-dimensional log densities are discretized through cubic splines over a certain interval.

Methodologically, our work is mostly related to the emerging literature on the joint estimate of macroeconomic aggregates and distributional data derived from individual attributes. Statisticians are increasingly interested in analyzing samples of random objects that do not belong to vector spaces, such as univariate probability measures (Kokoszka et al., 2019; Chen et al., 2021; Zhang et al., 2022; Petersen et al., 2022; Zhu and Müller, 2023). On a broader level, one can conceive densities as a specific class of functions featuring non-negative and integration constraints. The more general statistical examination of functional data resulted in a large literature, comprehensively treated by Ramsay and Silverman (2005) and Horváth and Kokoszka (2012), among others. In economics, functional data emerge in the context of yield curves (Diebold and Li, 2006; Inoue and Rossi, 2019). Methods derived from functional data analysis could also be used to deal with probability measures, as in Meeks and Monti (2023) or Chang et al. (2024) quoted above. In these cases, proper densities are restored from the estimated functions by ex-post renormalizations, which usually generate negligible approximation errors.

Finally, to some extent, our work is related to an extensive recent literature on 'expectational shocks', where exogenous changes in expectations are assumed to drive economic fluctuations, through belief, sentiment, confidence or news shocks.[3]

[3]This literature investigates, both empirically and theoretically, how news about future TFP (e.g., Beaudry

In what follows, Section 2 presents the data, Section 3 the methodology, Section 4 the results, and Section 5 concludes.

2 Data

We estimate the model on monthly US data, using a number of macroeconomic variables and microdata from the MSC. The macroeconomic variables are: Real Oil Price (OP), Consumer Price Index (CPI), Industrial Production (IP), and Nominal Interest Rate (IR). The period considered is from January 1983 to December 2019, for a total of $T = 422$ observations.[4]

From the MSC, we use microdata on inflation expectations, and consumer sentiment. Each consumer is asked to respond about the future expected inflation. In this study we focus on expectations over the next 12 months, i.e., one year ahead inflation expectations.[5] The consumer sentiment is constructed in the MSC by aggregating five questions related to the financial and business conditions of the households, as well as short and long-term perspectives on the economy. Individual sentiments are then pooled together to create the Michigan Consumer Sentiment (SENT), which serves as a barometer for expectations about the economy.[6]

Finally, we use two measures of uncertainty: financial uncertainty (FU) and macroeconomic uncertainty (MU) form Jurado et al. (2015).

3 Methodology

In this section we first describe how we summarize the large amount of information contained in the survey by means of distributions. Then, we introduce our model that combines this information together with macroeconomic time series. Finally, we show how to simulate the effects of exogenous variations in the distribution of inflation expectations through density impulse response functions.

3.1 Distributional Approximation

We consider three different distributional approximations of the large cross-sections of the expectations. In what follows, π denotes the general support of the expectations distribution, and $\pi_{i,t}$ the observed individual inflation expectation of individual i at time t, where $i = 1, ..., N_t$ and N_t is the number of individuals.

1. The first approximation complies with the original format of the Michigan survey, where inflation expectations are reported as integer percentage values. The most intuitive way

and Portier, 2006, 2014; Barsky and Sims, 2011, 2012) or exogenous waves of optimism and pessimism due to sentiment (e.g., Benhabib et al., 2015; Angeletos et al., 2018) affect business cycle fluctuations.

[4] We start in 1983 to avoid the potential effects on expectations of the change in the monetary policy regime during the Volcker disinflation period. The data can be downloaded from the St. Louis FRED database with the following ID: WTISPLC, INDPRO, CPIAUCSL, FEDFUNDS, UMCSENT. We stop in 2019 to avoid the Covid period.

[5] The questions refer to "prices in general" or "inflation", without specifying a particular measure. The survey is a short rotating panel with average number of respondent of 566 (min: 480, max: 1459). A summary of the main features of the survey data is available in Appendix E.

[6] A detailed description of the index construction is available at https://data.sca.isr.umich.edu/fetchdoc.php?docid=24770 .

to represent this dataset is simply by recording the proportion of respondents who report specific inflation values. This approach utilizes the "raw data" directly, without additional transformation. We denote this distributional approximation with the upperscript D (for discrete) and summarize it as follows:

$$f_t^D(\pi = i) = p_{i,t}, \quad p_{i,t} = \frac{1}{N_t}\sum_{j=1}^{N_t} \mathbb{1}\left[\pi_{j,t} = i\right] \tag{1}$$

where i is restricted in the interval $\{-50, ..., 50\}$, $\mathbb{1}[\cdot]$ is the indicator function that takes value of 1 when the argument is true, and $\mathbf{p}_t = [p_{-50,t}, \ldots, p_{50,t}]'$ stores the time t probability masses for each i. The discrete approximation uses empirical probabilities, which serve as a valid nonparametric estimator for a discrete Probability Mass Function (PMF) and converge consistently to the true distribution as $N_t \to +\infty$. Nevertheless, in finite samples it may exhibit high variance. Discrete Kernel Density (DKD) estimators use smoothing techniques to remedy this deficiency, mitigating the effects of small sample sizes while retaining the discrete nature of the data. These estimators generalize empirical probabilities, which can be recovered as a special case when the bandwidth approaches 0. To formally test whether this limiting case is supported by the data, we implement a DKD estimation procedure using the `R` package `np` (Hayfield and Racine, 2008).[7] No significant differences are observed between the two approaches; therefore, we proceed in the paper relying on the simpler and intuitive empirical probabilities.

2. Although the discrete assumption preserves the survey original structure, it neglects a crucial aspect of the expectation formation process: rounding. It is a well-known fact that consumers tend to report round numbers for simplicity, cognitive ease or bias, and/or to express approximate guesses facing lack of information or uncertainty (Binder, 2017). When consumers report an expectation of 5, for instance, they likely have in mind something around 5, encompassing a range of real-valued neighboring numbers. Thus, treating the underlying expectations distribution as continuous is an alternative reasonable assumption. To account for this, we find it convenient to work with densities that are output of a continuous Kernel estimator, which allows us to remain agnostic about the specific functional form. We label this approximation with the upperscript K (for Kernel) and represent it as:

$$f_t^K(\pi) = \frac{1}{N_t \mathcal{B}_t}\sum_{j=1}^{N_t} \phi\left(\frac{\pi - \pi_{j,t}}{\mathcal{B}_t}\right) \tag{2}$$

where $\phi(\cdot)$ is the standard Gaussian kernel, and $\mathcal{B}_t$ is a bandwidth selected at each t using Silverman's rule of thumb. The kernel approach can be viewed as a cross-sectional filter that smooths individual expectations over the cross-sectional domain. Sometimes the discrete dataset shows jagged edges that are more likely to be noisy rather than rational patterns (like the variability in the composition of survey participants). Then the Kernel

[7] In particular, for each t we fit the Wang and Van Ryzin (1981) Kernel to the Inflation Expectations dataset using a cross-validation strategy to select the optimal bandwidth. Figure D.1 in Appendix D compares the average distributions over the temporal dimension derived from the empirical probabilities and the above mentioned kernel method.

smoother helps to impose structure by reducing the overall cross-sectional variability. For this reason, the continuous Kernel is our preferred approximation; and will serve as the benchmark for all the experiments throughout the paper.[8]

3. The final approximation we consider is based on a parsimonious Gaussian distribution, which serves as a baseline for assessing whether capturing the full heterogeneity is important or not. Indeed, it may be that only a handful of indicators are sufficient to capture the distributional shifts that are relevant to macroeconomic dynamics. We want to stress that our goal here is not to achieve the best distributional fit for the expectations dataset, but rather to identify the aspects of heterogeneity that drive macroeconomic fluctuations. Under this approximation, identified with the upperscript G, we have:

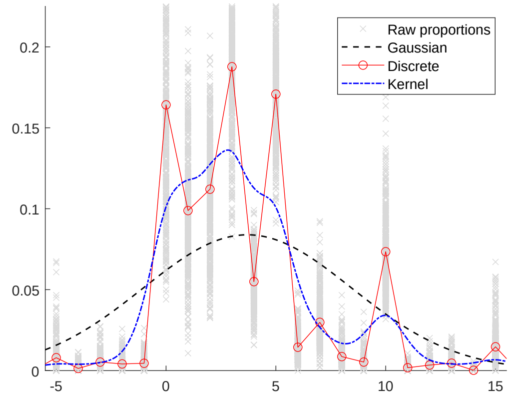
$$f_t^G(\pi|\mathbf{w}_t^G) = \mathcal{N}(\mu_t, \sigma_t^2) \tag{3}$$

where $\mathbf{w}_t^G = [\mu_t, \sigma_t^2]'$ stores the parameters characterizing of this last approximation.

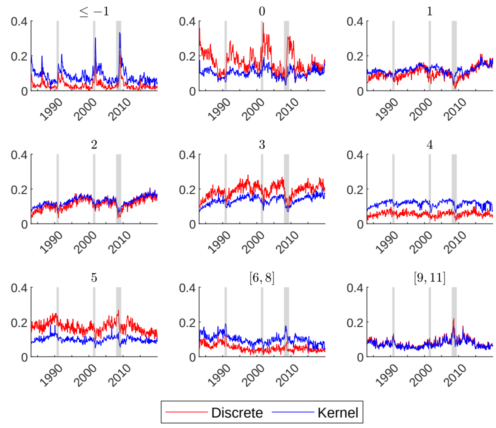
Panel (a) of Figure 1 shows the raw proportions $\mathbf{p}_t$ for each t (i.e., the grey crosses), along with the average distributions implied by the three different approximations: $\mathrm{E}_t(f_t^D)$, $\mathrm{E}_t(f_t^K)$, and $\mathrm{E}_t(f_t^G)$. The pervasive heterogeneity observed suggests that a flexible distribution is indeed essential. Inflation expectations show clear asymmetry, with a long right tail and multiple modes. The majority of the probability mass is concentrated between 0 and 5, which we refer to as the bulk of the distribution. Numbers such as 0, 5, and 10, have higher mass compared to the closest neighbors, indeed suggesting that consumers tend to report "round numbers". The Gaussian distribution is clearly misspecified, as it fails to capture the evident asymmetry and multimodality in the data. Notably, it assigns excessive weight to values below 0, where the probability is actually low.

Aggregating probabilities. We face two main challenges in incorporating the discrete f_t^D and continuous f_t^K approximations in a macroeconometric model. The first one is related to their very high-dimensional nature. Indeed, the flexibility gained by not imposing parametric forms comes at the cost of losing the ability to summarize the distribution with a handful of parameters, as for the Gaussian case. Second, as probabilities, they do not live in a vector space (being nonnegative and with a constrained sum/integral), thus making the application of typical vector space methods not applicable. We tackle the first and the second issue by complementing f_t^D and f_t^K with their respective cumulative probabilities/distribution functions over intervals of interest. The number and width of these intervals are chosen with the objectives of: (i) substantially reducing the dimensionality of these distributions, (ii) minimizing the likelihood of constraint violations, thus enabling the use of an unrestricted vector-valued model. In particular, let $\mathcal{Q} = \{q_0, q_1, ..., q_M\}$, with $q_0 = -51$ and $q_M < 50$, be a set of integers that partition the

[8] We want to stress that the two kernel estimations conducted (this continuous one and the discrete kernel mentioned earlier) are conceptually distinct and serve different purposes. The discrete kernel estimation is intended to assess whether the empirical probabilities can be effectively used to estimate efficiently the underlying PMF of the data, a conclusion supported by the results. In contrast, the continuous kernel estimation serves to smooth the irregularities in the data. While this could be achieved using various continuous distributions, the Gaussian kernel was chosen for its flexibility.



(a)



(b)

Figure 1: **Inflation Expectations Dataset.** Panel (a): Inflation expectations dataset and the distributional approximations considered in this study, from the cross-sectional perspective. Raw proportions (gray crosses), mean of the discrete approximation (red circles), of the continuous Kernel approximation (blue dashed-dotted line), and of the Gaussian approximation (black-dashed line). Panel (b): time series evolution of the 9 aggregated proportions of the discrete (red lines) and the Kernel (blue lines) approximation.

inflation support. Note that $M \leq 101$, given that the maximum number of intervals is obtained when we consider all the integers between -51 and 50.

For both the discrete and kernel approximation, we propose to work with aggregate probabilities over the M intervals:

$$w_{m,t}^D = \sum_{j \in (q_{m-1}, q_m]} p_{j,t}, \quad w_{m,t}^K = \int_{q_{m-1}}^{q_m} f_t^K(x)dx \tag{4}$$

where $\mathbf{w}_t^D = [w_{1,t}^D, ..., w_{M,t}^D]'$, and $\mathbf{w}_t^K = [w_{1,t}^K, ..., w_{M,t}^K]'$ collect the M dimensional, time t aggregate probabilities among all the intervals. The rationale is simple: for each category, we assume that the cross-sectional variability of the distribution can be well captured by a single factor, calculated as the sum/integral of the individual probabilities. While alternative methods for factor extraction can be explored, simple aggregation enhances interpretability. For example, by setting $q_1 = -1$, $w_{1,t}^D$ represents the evolution of the percentage of individuals reporting negative inflation values.[9]

This aggregation scheme naturally suggests a piecewise constant decomposition for both the probability masses f_t^D and densities f_t^K, with jumps occurring at the knots $q_m \in \mathcal{Q}$. To note this, consider that for the discrete case, equation (4) can be written compactly as:

$$\mathbf{w}_t^D = \tilde{\mathbf{H}}_D \mathbf{p}_t$$

[9]To make a parallel with the literature, this metric is equivalent to the indicator used by Allayioti et al. (2023) in a Smooth Transition VAR environment to identify periods characterized by (using their terminology) "fear of deflation".

where $\tilde{\mathbf{H}}_D$ is an $(M \times 101)$ selection matrix given by:

$$\tilde{\mathbf{H}}_D = \begin{bmatrix} \boldsymbol{\iota}'_{q_1-q_0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\iota}'_{q_2-q_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{0} & & \boldsymbol{\iota}'_{q_M-q_{M-1}} \end{bmatrix}$$

with $\boldsymbol{\iota}_N$ being a N dimensional vector of ones. To get a reasonable approximation of the vector $\mathbf{p}_t$, we use the Moore-Penrose generalized inverse of $\tilde{\mathbf{H}}_D$, that we indicate as $\mathbf{H}_D$ to keep the notation simple:

$$\mathbf{p}_t \approx \mathbf{H}_D \mathbf{w}_t^D. \tag{5}$$

In this particular case, each entry of the vector $\mathbf{w}_t^D$ is approximated as:

$$p_{j,t} \approx \frac{w_{m,t}^D}{q_m - q_{m-1}}, \quad j \in (q_{m-1}, q_m]$$

as shown in Appendix A. It is worth noting that when $M = 101$, $\mathbf{H}_D$ collapses to an identity matrix, meaning no approximation occurs. As M decreases, the approximation error increases and the dimensionality reduction becomes substantial. In this case, the bias-variance trade-off crucially hinges on the knots' position, which we discuss next.[10]

An important feature of the basis function system implied by (5) is the efficiency, stemming from the fact that $\mathbf{H}'_D\mathbf{H}_D$ is a diagonal matrix. This ensures the resulting coefficients are independent, thus avoiding multicollinearity, and simplifying matrix operations compared to non-orthogonal bases. The primary advantage of the piecewise constant approximation, however, lies in its enhanced interpretability. Specifically, $w_{j,t}^D$ directly represents the proportion of respondents reporting an inflation value within the interval $(q_{j-1}, q_M]$ during period t. This feature is critical, as a key contribution of this paper is the introduction of a framework to analyze how changes in a distribution affect macroeconomic variables. When we design distributional shifts for structural scenario analysis, we use infinite/high dimensional distributions, which must be translated back into M aggregate categories. This task becomes trivial with the proposed decomposition, as it avoids the need for additional processing in the form of optimization to obtain loadings from nonlinear basis functions. By contrast, more complex basis systems, such as the piecewise cubic basis with constraints proposed by Chang et al. (2024), require nonlinear optimization steps due to nontrivial mapping. It is important to stress that the goal of this paper is not to indentify the best functional form to fit the densities – for which the basis system of Chang et al. (2024) is undoubtely preferred – but rather to capture the heterogeneity driving macroeconomic fluctuations. For this, we prioritize interpretability

[10] A similar argument can be made for f_t^K, by noticing that in practice we do not work with this infinite dimensional object but rather with its high dimensional vectorized counterpart $\mathbf{f}_t^D$, obtained by evaluating f_t^K at a dense grid $[-50, -50+\delta, \dots, 50-\delta, 50]'$, with $\delta = 1/100$ for this paper. Specifically, the Kernel-based equivalent of (5) is given by

$$\mathbf{f}_t^K \approx \mathbf{H}_K \mathbf{w}_t^K \tag{6}$$

where $\mathbf{H}_K$ is a $[101/\delta \times M]$ selection matrix constructed in a similar manner to $\mathbf{H}_D$.

Table 1: **Intervals selection procedure**

	Empirical quantiles: q_m, $m = 1, \ldots, M$									
	-.75	**0**	**1**	**2**	**3**	**4**	**5**	**7**	**10**	**15**
$\mathbf{M = 21}$	1	3	2	3	4	1	4	1	1	1
$\mathbf{M = 19}$	1	2	3	2	4	1	3	1	1	1
$\mathbf{M = 17}$	1	2	2	2	3	1	3	1	1	1
$\mathbf{M = 15}$	1	2	2	1	3	1	3	0	1	1
$\mathbf{M = 13}$	1	1	2	1	3	1	2	0	1	1
$\mathbf{M = 11}$	1	1	1	2	2	0	2	1	0	1
$\mathbf{M = 9}$	1	1	1	1	1	1	1	1	0	1

Notes: This table shows the results of the iterative process of selecting interval boundaries q_m, for different values of M. The procedure starts with $M = 21$ and reduces M incrementally, recalculating the empirical quantiles at each step. Each column represents all the empirical quantiles found in the procedure. Each row corresponds to a specific M, the number of intervals considered at that step. Entries represent the frequency with which each quantile appears as an interval boundary.

over predictive accuracy, adopting an intuitive decomposition. While this method might not be optimal for forecasting future outcomes of the inflation expectation distribution,[11] more complex basis functions – though potentially better for forecasting – come at the expense of direct interpretability, which is central to our analysis.

Interval selection. There are no established guidelines for selecting the number of intervals or the placement of knots in spline-based methods. When higher-order polynomials are used, the exact position of the knots becomes less significant because continuity is enforced at their intersections. In our case with stepwise approximation, intervals selection is important as it determines the values at which discontinuity occurs. Our approach follows a simple rationale: intervals and their boundaries are determined jointly, to ensure that each category contains, on average, a significant portion of the total probability. This helps avoiding redundant intervals with negligible representation. The calculation is based on the raw, discrete dataset, and the resulting intervals are subsequently applied to the kernel-smoothed version of the data.

Let $\tau_1 = 0.025$, $\tau_M = 0.975$ $\tau_m = \tau_{m-1} + \frac{\tau_M - \tau_1}{M-1}$, $m = 2, ..., M$ be an equally spaced sequence on the unitary interval. We start from $M = 21$ and compute the median values (over the temporal dimension) of the τ_m-th empirical quantile of $[\pi_{1,t}, \ldots, \pi_{N_t,t}]'$. The empirical quantiles of the raw dataset form stepwise functions, as they are derived from inflation expectations sorted in increasing order. With $M = 21$ we observe that many values satisfy the same quantiles, indicating that the width of the intervals $\tau_m - \tau_{m-1}$ (which is roughly 0.047 in this case) is not large enough to encompass significant probabilities. To address this, we progressively reduce M to $19, 17, 15, \ldots,$[12] repeating the calculations at each step. The procedure terminates when all the empirical quantiles feature distinct values, ensuring that each interval contains a non-

[11] Kokoszka et al. (2019) tackle the problem of forecasting dynamic densities with a variety of methods, both parametric and non.

[12] This method of constructing the intervals, combined with the choice of considering only odd values for M ensures that the central quantile $q_{(M+1)/2}$ always correspond to the median, a key value we aim to control.

negligible probability mass. The criteria is satisfied for the first time at $M = 9$ (with average probability mass $\tau_m - \tau_{m-1} \approx 0.12$), with distinct empirical quantiles. Table 1 summarizes the results of this procedure, showing, for each M considered, the frequency with which each value appears. The final sequence is adjusted further by rounding the first element ($q_1 = -1$), and by replacing the last two knots with $q_{M-1} = 8$, and $q_M = 11$. This adjustment is motivated by the unique characteristics of the survey, where the value 10 holds significant importance, acting as a distinct mode in the long right tail. With this variation, the M-th interval is centered at 10, allowing it to be explicitly controlled. The interval selection procedure ends with $M = 9$ intervals separated by knots:[13]

$$\mathcal{Q} = \{-51, -1, 0, 1, 2, 3, 4, 5, 8, 11\} \tag{7}$$

The procedure yields satisfactory results, as the last ($M + 1$-th) interval $(11, 50]$ is sufficiently wide to ensure that the sum of the first M proportions is systematically lower than 1.[14] To illustrate the results, Panel (b) of Figure 1 displays the time series evolution of the $M = 9$ aggregated proportions of the discrete approximation $w^D_{m,t}$ (red lines), and the Kernel approximation (blue lines).[15]

3.2 The Time Series Model

The output of the distributional approximation is simply represented by the vectorized version of the parameters among the three approximations considered: $\mathbf{w}^D_t$, $\mathbf{w}^K_t$ (both $M = 9$ dimensional), and $\mathbf{w}^G_t$ (two-dimensional). In other words, the nine time series of the parameters $w^D_{m,t}$ and $w^K_{m,t}$ for $m = 1, 2, ...9$ – shown in Panel (b) of Figure 1 – approximate the time series dynamics of the discrete and kernel approximations, respectively, as the two time series of the parameters $\mathbf{w}^G_t = [\mu_t, \sigma^2_t]'$ approximate the dynamics of the Gaussian approximation. These are the time series that enter our time series BVAR model. For $a \in \{D, K, G\}$, the time series of parameters $\mathbf{w}^a_t$ and the index of consumer sentiment form the expectation block $\mathbf{e}_t = [\mathbf{w}^a_t, \quad s_t]'$, whose dimension, $N_e \times 1$, depends on the type of approximation considered. All the remaining variables (MU, FU, OP, CPI, IP, IR) form the macroeconomic block $\mathbf{x}_t$ of dimension $N_x \times 1$. We form the $N = N_e + N_x$ dimensional vector of endogenous variables $\mathbf{y}_t$ by vertically stacking the two vectors for each t: $\mathbf{y}_t = [\mathbf{e}'_t, \ \mathbf{x}'_t]'$.

Let $\boldsymbol{\mu}_t$ be the long run component of $\mathbf{y}_t$, we assume that the deviations $(\mathbf{y}_t - \boldsymbol{\mu}_t)$ have stationary dynamics and null unconditional expectations. We model these deviations as a stable VAR:

$$\boldsymbol{\Phi}(L)(\mathbf{y}_t - \boldsymbol{\mu}_t) = \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{8}$$

[13]The same interval boundaries are used for the Kernel approximation, with just one difference: we adjust them by subtracting 0.5 from each q_m value, effectively placing integer values at the midpoint of each interval. This centering is crucial for the Kernel approximation, as it ensures that the aggregated probabilities remains accurately aligned with the interval centers. In contrast, this adjustment does not affect the discrete approximation since the probability mass outside the integers is always zero.

[14]In the resulting aggregate probabilities $\mathbf{w}^D_t$ and $\mathbf{w}^K_t$ there is only one instance where the probabilities touch the lower bound of 0. Since the violation of the inequality constraint is so rare, there is no compelling need to modify the likelihood to account for this negligible possibility.

[15]Figure D.2 in Appendix D shows the correlation matrix among $\mathbf{w}^D_t$.

where $\mathbf{\Phi}(L) = \mathbf{I}_N - \mathbf{\Phi}_1 L - ... - \mathbf{\Phi}_P L^P$ is the usual polynomial in the lag operator L, and $\mathbf{\Phi} = [\mathbf{\Phi}_1, ..., \mathbf{\Phi}_P]$ are reduced form lagged parameters. Finally, $\boldsymbol{\epsilon}_t$ is the N dimensional innovation process with positive definite covariance matrix $\mathbf{\Sigma}$. Our model is equivalent to the VAR in deviations from its steady states of Villani (2009), with the difference that in this specification some of the steady states are allowed to change deterministically over time to take care of non-stationarity.[16] Under this specification, the long-run dynamics can be concisely represented as $\boldsymbol{\mu}_t = \mathbf{M}_t \boldsymbol{\theta}$, where the rows of $\mathbf{M}_t$ contain the variable specific deterministic basis functions, and $\boldsymbol{\theta}$ is the vector of parameters to estimate.

Priors and hyperpriors. Bayesian inference requires a prior distribution on the reduced form parameters $\boldsymbol{\theta}$, $\boldsymbol{\phi} = \text{vec}(\mathbf{\Phi}')$, and $\mathbf{\Sigma}$. We assume that the long-run coefficients are apriori independent and Normally distributed: $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}_0, \mathbf{V}_\theta)$, where $\boldsymbol{\theta}_0$ and $\mathbf{V}_\theta = \text{diag}(v_{1,\theta}, ..., v_{K,\theta})$ are, respectively, the estimated coefficients and 100 times their asymptotic variance resulting from N separate OLS long-run fit. The VAR coefficients are a priori independent across equations, and Normally distributed: $\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\phi)$. For each coefficient $\mathbf{\Phi}_{p,ij}$, reflecting the p-th lagged effect of variable j on equation i, we set its prior variance according to the convenient Minnesota structure:

$$\text{V}(\mathbf{\Phi}_{p,ij}) = \begin{cases} \frac{\lambda_1}{p^{\lambda_p}} & i = j \\ \frac{\lambda_1 \lambda_2 \sigma_i^2}{p^{\lambda_p} \sigma_j^2} & i \neq j \end{cases}$$

where σ_i^2 denotes the sample variance of the residuals from an AR(4) model for variable i, λ_1, λ_2, and λ_p are hyperparameters. Given the high dimensional setting, we estimate λ_1 as in Chan (2021), though we depart from this approach by assuming an inverse-Gamma hyperprior: $\lambda_1 \sim \mathcal{IG}(\alpha_1, \alpha_2)$. The prior setting concludes by assuming the usual Inverse-Wishart for $\mathbf{\Sigma}$: $\mathbf{\Sigma} \sim \mathcal{IW}(\mathbf{\Sigma}_0, \nu)$, where $\mathbf{\Sigma}_0$ and ν are, respectively, the prior scale matrix and prior degrees of freedom. We discuss the hyperparameters setting and the implementation details of the Gibbs sampling steps in Appendix B.

Identification. We impose a set of minimal identification assumptions based on the natural discrepancies between the date the survey is conducted, and the date in which macroeconomic variables are released. The Michigan survey usually takes place between the third and the fourth week of each month, whereas all the macroeconomic variables of the current month are released weeks later. Hence, when consumers form their expectations, and answer the survey questions, they don't know which is the current value of the aggregate variables. As a consequence, they can't be influenced by them during the current month. This feature of the data allows us to impose that all the macroeconomic variables have a zero restricted impact on the expectation equations. However, we don't impose this assumption for the oil price and uncertainty, which

[16] The MU, FU, and expectation block variables are stationary, so we assume a constant trend for these series. In contrast, all macroeconomic variables exhibit non-stationary behavior, necessitating time-varying trends. Specifically, we model CPI, IP, and IR with a quadratic polynomial. However, the long-run behavior of OP exhibits a more complex pattern, and a simple quadratic trend does not yield a stable gap. To enhance the flexibility of the quadratic polynomial, we extend it to a quadratic spline with a single knot positioned at the midpoint of the time domain. This modification allows the overall trend to accommodate up to two changes in the sign of the first derivative, compared to a single one for the simpler quadratic polynomial. The selected specification results in relatively stable dynamics: we observe that fewer than 40% of draws from the posterior leads to explosive behaviors.

Table 2: **Identification Assumptions.**

	Unc and OP shocks	**Expectation shocks**	**Macro shocks**
Unc and OP	+	0	0
Expectation block	/	+	0
Core Macro	/	/	+

Notes. Entries in this table show the restrictions imposed: + positive sign, 0 no effect, / no restriction. Rows represent the variables and column the shocks. Inflation expectation shocks and other fundamental macroeconomic shocks are block identified. The Unc and OP are affected on impact only by their own shock. The macro block shock does not affect on impact any of the other variables.

we treat as the most exogenous variables in the system. It is reasonable to assume that oil prices might instantaneously impact inflation expectations and sentiment because consumers are likely to observe the petrol price - highly correlated with oil prices - at the petrol station while filling their car tanks, and because the literature suggests that the price at the pump is a very salient price for consumers. The identification assumptions are illustrated in Table 2.

3.3 Density Impulse Response Function Analysis

In this paper, we want to investigate the macroeconomic effects of shocks to the distribution of inflation expectations, approximated by the M aggregated proportions. Thus, we are going to perturb this distribution, using the identified shocks to inflation expectations, and then study the response of the macroeconomic block through a Density Impulse Response Function (DIRF) approach. Although each component of $\mathbf{w}_t^a$ enters the BVAR individually, it arises from the same, constrained object. Therefore, when perturbing the distribution, it is essential to devise movements that: i) respect the inherent constraints, and ii) represent economically meaningful scenarios.

This Section explains how we tackle these challenges in constructing such a density impulse, that is, a shock to the distribution. Let us use as an example what we would call later a "mean shock", that is a perturbation where the location of the inflation expectation distribution increases from the baseline by, say, 1%. In a non-distributional framework, this corresponds simply to a 1% increase in consensus (mean) expectations. In the distributional setting, however, all proportions shift such that the modified distribution maintains the baseline shape, but shifts to the new location.

Moreover, in addition to the specific movement, we need to define the composition of shocks that generate these movements in the aggregated proportions. Recall that given our identification restrictions, the shocks to expectations are block identified, so we use a targeted subset of shocks. These correspond, at each time t, to N_e reduced form shocks: the shocks to the M parameters of distribution approximation plus the shock in the equation of sentiment. This reflects our view that sentiment pertains to the expectation formation process for consumers. It is part of the expectation block, as it is part of the MSC, because it directly affects consumer's expectations reflecting the consumer's perceived stance of the economy. Thus, our structural

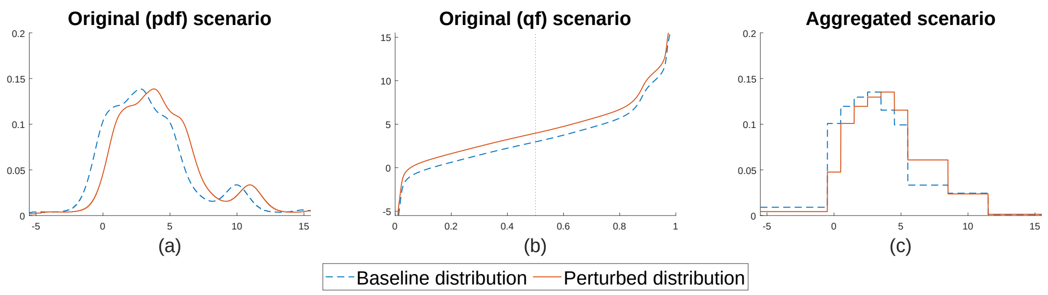


Figure 2: **Example of Density Impulse construction: location shock.** In this figure we show the workflow to construct the experiment of the positive location shocks under the Kernel distributional approximation. Panel (a) shows the pdf function; Panel (b) shows the quantile function, Panel (c) shows the aggregated proportions (or probabilities). Each panel shows two curves: the baseline one (dashed-blue line) and the perturbed one (solid-orange line).

shock to the expectation block is a linear combination of these N_e reduced form shocks. However, there is an infinite number of linear combination of N_e shocks that satisfy the particular movement in the aggregated M proportions. This feature of "constrained" future movements and shocks pave the way for a structural scenario analysis in the spirit of Antolín-Díaz et al. (2021). However, given our distributional framework, we extend the standard methodology to what could be interpreted as a "density structural scenario analysis", since we find the posterior distribution of the N_e shocks realizations that satisfy the restrictions imposed by the perturbation of the distribution, imposing all the other non-expectation shocks to be zero.Intuitively, the infinite solutions of our problem are subject to a probability distribution that reflect the correlations in the data. This procedure, which is explained in details in Appendix C, is very general and allows us to experiment with a variety of economically meaningful shocks to the distribution of inflation expectations. We can analyze variations in the common statistics like the various moments, and we can even study the effects of exogenous marginal movements of specific sections of the distribution.

To construct our density impulse, that is, the perturbation of the distribution, we find it convenient to use as baseline the average distributions, i.e., either $\mathrm{E}_t(f_t^D)$ or $\mathrm{E}_t(f_t^K)$ shown in Panel (a) of Figure 1. We then apply shocks to these distributions, compute the perturbed distributions, and re-aggregate them to align with the proportions used in the BVAR. However, perturbing a distribution is inherently challenging, as linear combinations of distributions are not valid distributions satisfying the integration constraint. To avoid this problem, we exploit the correspondent quantile function (qf hereafter), that can be more easily manipulated, because linear combinations of qfs yield valid qfs, which can then be mapped back to a valid probability density function pdf.[17] Figure 2 visualizes graphically the steps of the procedure for the location shock in the Kernel case, which are summarized as follows.[18] First, we start with the baseline pdf (dashed-blue line in Panel (a)) – that will be equal to $\mathrm{E}_t(f_t^K)$ in our DIRF experiments – which is aggregated as the dashed-blue line in Panel (c). As said, rather than perturbing

[17]Notably, the "shock" itself must satisfy the monotonicity constraint to combine properly with the baseline qf and produce a valid perturbed qf. The location shock corresponds to the quantile function of a Dirac delta distribution, characterized by an infinite point mass at 1.

[18]For the sake of illustration, Figure C.1 and Figure C.2 in Appendix C shows the same procedure for the case of the location shock for discrete approximation and for the case of a positive variance shock, respectively.

directly this distribution, we make use of an intermediate step by exploiting the qf. Hence, second, we recover the correspondent qf for the baseline distribution (dashed-blue line in Panel (b)). Third, we perturb the qf (solid-orange line in Panel (b)). Fourth, from the qf we can uniquely recover the correspondent pdf (solid-orange line in Panel (a)). Finally, we aggregate the perturbed pdf, as shown by solid-orange line in Panel (c).

4 Results

Steady states. Figure 3 displays the estimated steady-state distribution of inflation expectations. The wide range of the answers (roughly from -5 to 20 percent) and the multimodality of the distribution highlight the importance of the heterogeneity of beliefs. The median of the distribution is about 3.6%.[19]

The remainder of this Section presents the main results. Section 4.1 studies the cross-influence between the expectation and the macro blocks of the model, providing indirect empirical support to our identification strategy. Section 4.2 looks at the relative importance of shocks to the two blocks in explaining the volatility of the variables in the BVAR. Section 4.3 contains the main results of the paper, that is, the analysis, through density impulse response functions, of the effects, on the macroeconomic variables, of different kind of shocks to the inflation expectation distribution: mean, variance, kurtosis and tail shocks. Section 4.4 investigates the role of shocks to expectations during the three recessions caused, respectively, by the dot-com bubble, the GFC and the Covid-19 outbreak (c). Finally, Section 4.5 investigates the effects of communication shocks.

4.1 Cross-influence between blocks of variables

Before presenting our main findings, we investigate the cross-influences between two block of variables, focusing on the lagged coefficients in the BVAR. Although lagged coefficients lack structural information, they inform us about the relative predictive power of the lagged values of the two respective blocks of variables. Importantly, contrary to the dynamic analysis of the following sections, lagged coefficients are independent of the identification assumption described above in Table 2, providing a valuable context to support (or challenge) the identification.

In a frequentist setting, one potential method to explore this relationship is to conduct Granger-causality tests on parameter blocks. Here, we adopt a probabilistic perspective, leveraging on our model's hierarchical structure to estimate the optimal shrinkage hyperparameters for sub-blocks of the model.[20] To implement this, we adjust the framework discussed before, which relies on a pair of shrinkages that do not consider the block structure of our distributional setting, by assigning distinct shrinkages to parameters associated with three portions of the system:

[19] Figure D.3 in Appendix D shows the estimated trends and gaps for the three macroeconomic variables in the BVAR.

[20] This approach is comparable to that of Chang et al. (2024), with a notable distinction: they select shrinkages among blocks that maximize the Marginal Likelihood (ML) over a grid of candidates. Our prior setup is non-conjugate and more flexible, though it does not yield analytical results for the ML.

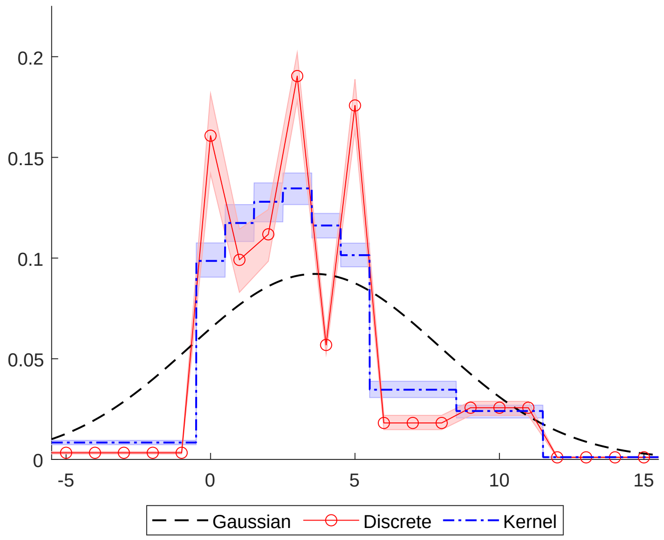


Figure 3: **SS-BVAR Steady States Distributions.** Posterior Steady State distributions of the three approximations implied by the SS-BVAR. Discrete approximation (red circles), continuous Kernel approximation (blue dashed-dotted line), and Normal approximation (dashed black line). The red and blue shaded areas delimit the 95% credible bands of the first two approximations.

- those related to the lagged effects of the macroeconomic variables on the expectation equations: $\lambda^{(m\to e)}$ for $\mathbf{\Phi}_p^{(m\to e)}$ (each block of size $N_e \times 3$),
- those related to the lagged effects of the expectation variables on the macroeconomic equations: $\lambda^{(e\to m)}$ for $\mathbf{\Phi}_p^{(e\to m)}$ (each block of size $3 \times N_e$),
- all other elements, including the autoregressive blocks and cross-block interactions (e.g., macro to/from exogenous and expectation to/from exogenous), which are not central to this analysis: $\lambda^{(\text{rest})}$ for $\mathbf{\Phi}_p^{(\text{rest})}$.

Notably, the blocks $\mathbf{\Phi}_p^{(m\to e)}$, and $\mathbf{\Phi}_p^{(e\to m)}$ contain an equal number of coefficients, making them directly comparable. Each of the three shrinkages has an inverse Gamma prior, with shape and scale parameters $\alpha_1 = \alpha_2 = 0.001$, that peaks around 0.001, and is rather non-informative. Figure 4 shows the comparison between the prior and posterior distributions for the two shrinkages of interest. The posterior median of $\lambda^{(e\to m)}$ is approximately 13 times greater than that of $\lambda^{(m\to e)}$, with minimal distributional overlap. Furthermore, the posterior distribution of $\lambda^{(m\to e)}$ lies in a peripheral area of the prior. Probably, under an even flatter prior, its location would decrease further, thus magnifying this result. In contrast, the absolute magnitude of $\lambda^{(e\to m)}$ is not negligible considering the size of the system, making the parameters of this block to be potentially influencing.

Independently of the identification assumptions, therefore, macroeconomic variables are less relevant for expectations than expectations are for macroeconomic variables. This result suggests that the expectation block is useful as a lagged predictor of the macroeconomic block, but not vice versa. In other words, expectations dynamics influence macroeconomic dynamics, while the reverse is not true. This evidence supports the recursive identification scheme described

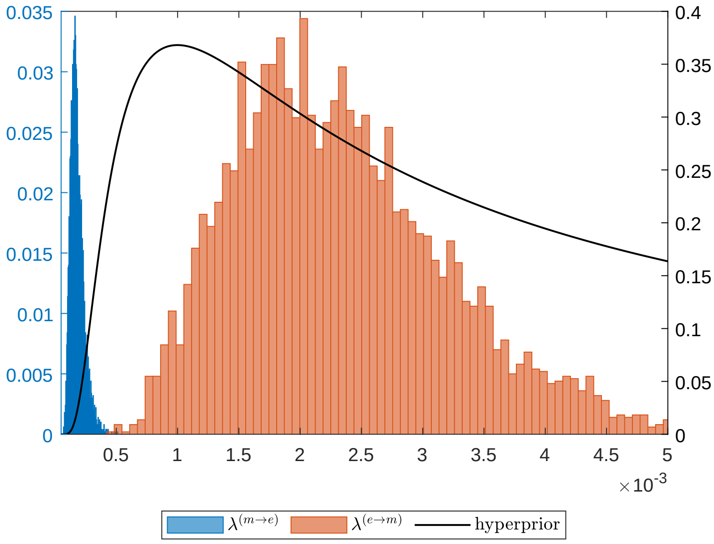


Figure 4: **Expectations are predictors of macroeconomic variables and not vice versa.** Hyperprior (black line, right axis scale) and posterior distribution of the two shrinkage parameters: $\lambda^{(m \to e)}$ (blue bins, left axis scale), and $\lambda^{(e \to m)}$ (orange bins, left axis scale).

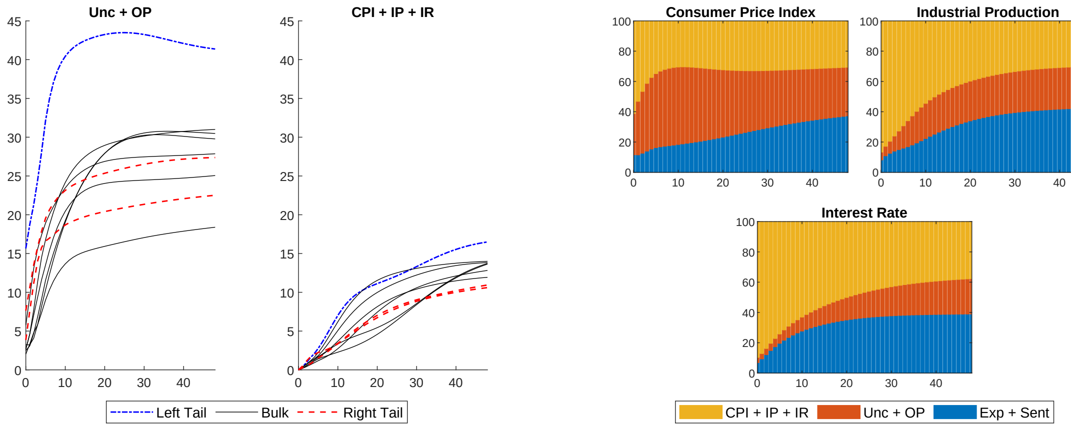
earlier, based on the timing of the information set. Finally, let us stress that this is a satellite exercise, while in the subsequent analysis we will assume the shrinkage setting described in the prior section (with a common shrinkage parameter).

4.2 The important role of exogenous variations in inflation expectations

Here we show that the analysis of the forecast error variance decomposition (FEVD) of the two blocks of variables of interest points to a similar message as in the last section: expectations matters relatively more for macroeconomic variables than vice versa.

Expectations are barely affected by macroeconomic shocks. In Figure 5 we report the forecast error variance decomposition (FEVD) of the two blocks of variables of interest. The right plot in Panel (a) shows the quota of FEVD – at different horizons on the x-axis – for each of the 9 (Kernel based) expectations explained by the three macroeconomic variables. To facilitate interpretation, the FEVDs are depicted using different colors based on their location. The lines start at zero given the assumption that macroeconomic variables do not affect expectations on impact. However, what stands out is the relatively low magnitude of the decompositions even for long horizons, reflecting a quasi-exogeneity of expectations with respect to the macroeconomic aggregates. This is a very surprising result given the fundamental tenant in macroeconomic models that inflation expectations are fully endogenous to macroeconomic developments. Instead, our result suggests that consumers in the MSC devote little attention to these macroeconomic variables in forming their inflation expectations. This is somewhat consistent with the recent large literature on survey expectations that showed that inflation expectations are far from rational, subject to various biases, and affected more by personal experiences or few salient prices rather than by the behavior of macroeconomic aggregates.

An additional interesting feature is the ranking that emerges by analyzing the single FEVD.



(a) (b)

Figure 5: **Shock to inflation expectations are important drivers of core macroeconomic variables, but not vice versa**. Panel (a): Contribution of the two exogenous variables (first subplot) and the three macroeconomic variables (second subplot) to the FEVD of the nine expectations series at different horizons (x-axis). Different colors indicate the location in the domain of the proportions. Left tail (≤ 0, blue dashed-dotted line), bulk ($\in [1, 5]$, black lines), right tail (≥ 5, red dashed lines). Panel (b): Contributions of the three different identified blocks of indicators to the FEVDs of the macroeconomic variables at different horizons (x-axis).

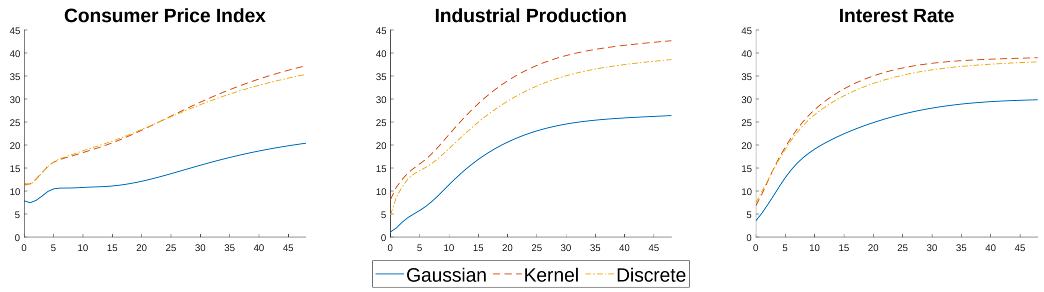


Figure 6: **A simple Gaussian approximation leads to an underestimation of the effects of shocks to inflation expectations.** Comparison among the different approximations of the contribution of the expectation block to the FEVD of the macroeconomic variables.

Although the responsiveness to macroeconomic fluctuations is generally low, it differs across proportions. In particular, the proportions on the right tail (red dashed lines) are the least affected, with values around 10% at the longest horizon, while the ones on the left tail (blue dashed-dotted lines) are the most affected. The more central quantiles (black solid lines) are somewhat in the middle. This pattern suggests that consumers' expectation formation process might be heterogeneous. The plot on the left of Panel (a) shows the quota of FEVD for each of the 9 (Kernel based) expectations explained by the exogenous variables in our model, that is, the uncertainty measures and the oil prices. First, shocks to the block of these exogenous variables affects expectations much more than shocks to the macroeconomic variable block. Second, results are again heterogeneous across proportions, with the left tail being particularly responsive to shocks to these exogenous variables.

Exogenous variations in expectations affect macroeconomic variables. Panel (b) of Figure 5 shows the FEVD for the macroeconomic variables. We partition the FEVD in three main contributions to streamline the interpretation: the expectations block (blue area), the exogenous variables block, i.e., uncertainty and real oil price, (orange area), and the macroeconomic block (yellow area). The result is the opposite to the one in Panel (a): exogenous shocks to the expectation block are an important determinant of variations in macroeconomic variables. Moreover, their relevance increases with the horizon. The respective share of the variance of the two year forecast error is roughly 25% percent for CPI, 36% percent for Industrial Production and 36% percent for the interest rate. Figure 6 shows that it is important to take into account the whole distribution of inflation expectations. It shows the FEVD for the three different approximations considered. Considering only the first two moments of the distribution (Gaussian approximation) would result in a sizeable underestimation of the importance of shocks to inflation expectations for the FEVD of macroeconomic variables, with respect to the other two proposed distributional approximations (Discrete and Kernel) that consider nine intervals over defined proportions. Importantly, these results are robust to our identification assumptions.[21]

Together Figure 4 and Figure 5 uncover the same consistent main message: expectations matter for macroeconomic variables, but not vice versa. Figure 5 suggests that the importance of inflation expectations goes beyond their role in the propagation mechanism of macroeconomic shocks, because there is an exogenous source of variation of expectations that directly affects macroeconomic variables. Therefore, it is important to understand the reaction of macroeconomic variables to shocks to inflation expectations. Moreover, Figure 6 shows that the whole distribution matters, and not just the first two moments, as mainly considered in the literature so far. Therefore, it is important to consider shocks to the whole distributions to understand how macroeconomic variables reacts to exogenous changes in inflation expectations. This is what we move next, by performing a battery of density impulse response functions, using the methodology explained in Section 3.3.

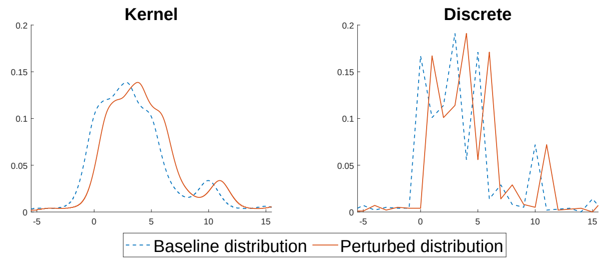
4.3 Dynamic implications of shocks to inflation expectations: the distribution matters

This Section contains the main findings of this paper. We perturb the distribution of inflation expectations, using expectation shocks and study the dynamic implications on all the variables in the model. We conduct several experiments analyzing the marginal effects of a change in the mean, in the dispersion and in some selected portions of the distribution.

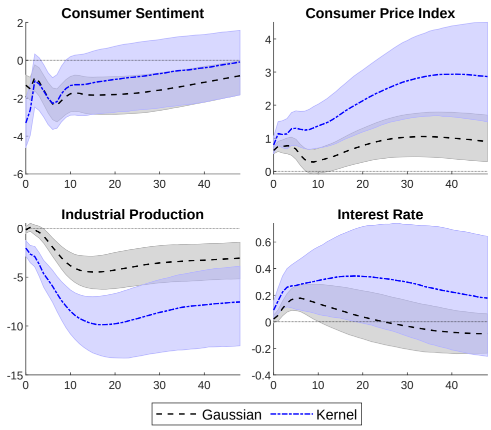
4.3.1 Location shock

The first experiment we run is the distributional counterpart of the IRFs one would study if the model contains a synthetic measure of expectations, like the mean. We ask the following question: what is the macroeconomic effect of a shock that increases inflation expectations by 1%? A positive location shock moves the quantile function upward without altering its shape, causing a rightward shift in the pdf, as shown by the dashed blue and orange lines in Figure 7, Panel (a). We generate a scenario where the distribution of inflation expectations translates to

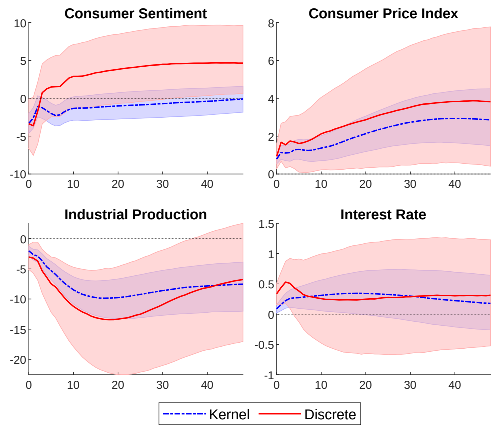
[21] Figure D.7 and D.8 in Appendix D.1 show that these results hold when we assume that interest rate are exogenous to expectations, or that all the macroeconomic variables are exogenous to expectations, respectively.



(a)



(b)



(c)

Figure 7: **A positive location shock has stagflationary effects.** The figure shows the macroeconomic effects of a shock that increases the location of the inflation expectations distribution of 1% on impact. Panel (a): Baseline and perturbed distributions on impact for the discrete and Kernel approximations. Panel (b): comparison between the Gaussian (black dashed lines), and the Kernel (blue dashed-dotted lines) approximations. Panel (c): comparison between the Kernel and the discrete (red continuous lines) approximation. Shaded areas delimit the 68% credible bands.

the right by 1% keeping its shape unchanged. In the Gaussian case, the experiment is simple since it results in a positive change in the mean with no variation in the variance.

We evaluate the results among the three approximations by first comparing the Gaussian and Kernel results (Panel (b)), and then the Kernel and discrete results (Panel (c)). We start by interpreting the macroeconomic effects of such a shock in the Gaussian case, presented with black dashed lines in Panel (b).

First, an exogenous increase in the mean of inflation expectations is stagflationary, inducing a rise in inflation and a decrease in industrial production. This result is consistent with Ascari et al. (2023) who obtain similar evidence in BVAR on aggregate variables, using different both dataset and identification assumptions. Moreover, the negative response of sentiment suggests that the effects on the economy work through the association of higher inflation with a pessimistic perception about the economy. This is in line with robust evidence from surveys pointing to households consistently having a supply-side view of inflation, such that they become more pessimistic about the economic outlook when their inflation expectations increases (see, e.g., Kamdar, 2019; Candia et al., 2020; Binder, 2020; Binder and Makridis, 2022). Moreover, Coibion et al. (2023) exploit a randomized controlled trial approach on Dutch households data to provide causal evidence from inflation expectations to actual spending. Households who have exogenously higher inflation expectations become much more pessimistic about the state of the economy and their future income growth, and they sharply reduce their spending on durable goods.

Second, blue dashed-dotted lines present the results of the Kernel-based distributional approximation. Both the parsimonious and the rich specifications agree on the fact that an exogenous increase in inflation triggers stagflationary effects. Even though the two models exhibit similar qualitative responses, however, there is a statistically significant and marked difference in the quantitative responses of industrial production and the consumer price index. Specifically, the parsimonious specification largely underestimates the magnitude of the recession induced by the shock. As a response to the same perturbation, the effect on industrial production in the richer model is more than two times bigger, with a long-lasting effect. Therefore, the distribution matters in this exercise, because not taking into account the whole distribution would lead to an underestimation of the effects of a shock to the mean of the inflation expectations.

Finally, in Panel (c) we compare the two rich distributional approximations, namely the Kernel and the Discrete (red solid lines). The responses of the macroeconomic variables are very close for the two rich approximations, but the credible bands are very wide for the Discrete approximation. This is intuitive given the inherent characteristics of the kernel approximation. Being the latter a cross-sectional smoother, it constrains the distributional variation of the discrete case. Hence, it is natural that the two produce responses that are comparable in median, with wider (possibly noisier) bands for the discrete case.

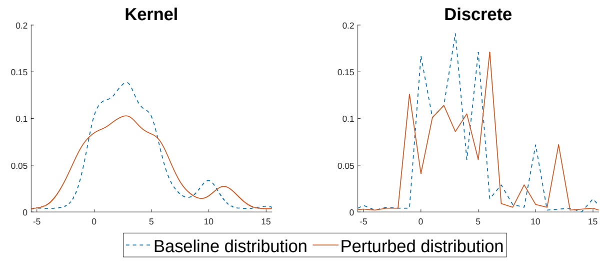
4.3.2 Dispersion shocks

The mean shifting shock informs us about the effect of a hypothetical scenario in which all the households exogenously believe inflation will be higher within a year. Another relevant scenario is represented by an exogenous variation of the disagreement, without a change in the consensus. What are the macroeconomic effects of a shock that increases the dispersion of the distribution, leaving the median unchanged?[22]

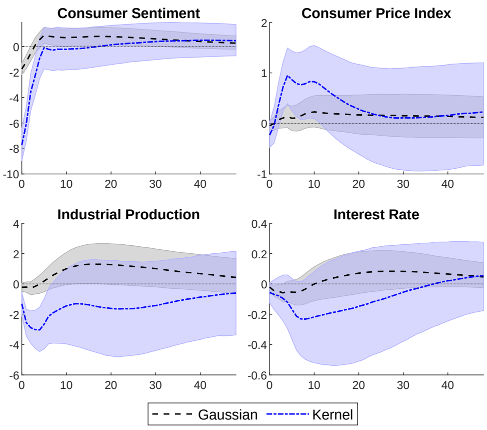
Dispersion is a broad concept, as it involves any movement that modifies the shape of a distribution. In a Gaussian context, it is easy to pin down the concept of dispersion, as there is only one parameter that governs it. With more flexible (but still parametric) distributions, multiple parameters influence dispersion, paving the way for a number of different scenarios. In our framework, any movement that squeezes the distribution, moving probability mass from the bulk to the tails, while keeping the median stable, represents an increase in dispersion.

We study dispersion shocks in two stages. First, mimicking the location shock in a dispersion context, we introduce a "dispersion perturbation" that can be applied to all three approximations so as to compare the resulting macroeconomic effects. Being the Gaussian approximation

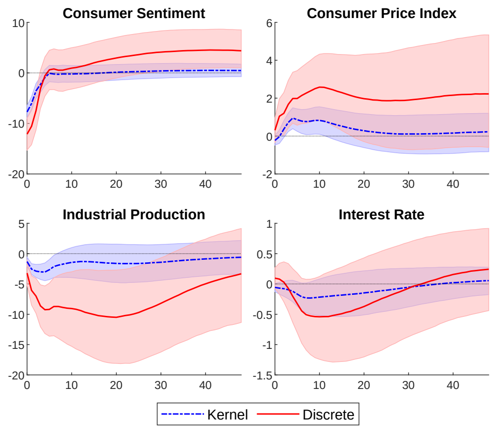
[22]These changes could also be interpreted as expectation uncertainty. While we are aware that a more precise measure of uncertainty would be the standard deviation taken from the probability distribution of a single household, the MSC does not provide those. Moreover, when available in surveys, these data have several rounding and time-inconsistency issues making estimations of uncertainty not reliable. Hence, very often the literature assumes that an increase in the dispersion of individual expectations across households proxies an increase in the expectation uncertainty of the households.



(a)



(b)



(c)

Figure 8: **A positive standard deviation shock has stagflationary effects.** The figure shows the macroeconomic effects of a positive standard deviation shock. Panel (a): Baseline and perturbed distributions on impact for the discrete and Kernel approximations. Panel (b): comparison between the Gaussian (black dashed lines), and the Kernel (blue dashed-dotted lines) approximations. Panel (c): comparison between the Kernel and the discrete (red solid lines) approximation. Shaded areas delimit the 68% credible bands.

the most restrictive approximation, we will derive this common shock from the Gaussian case, and then apply it to the other two approximations. In the second stage, we will further explore dispersion with the flexible Kernel approximation, considering kurtosis and tail shock. Finally in the next subsection, we will also consider asymmetric perturbations.

Standard deviation shock. As for all our perturbations, we go through the qf, as explained above in Figure 2. In the case of a unit increase in the standard deviation, the perturbing qf is the one of a standard Normal, i.e., $N(0,1)$. By adding this qf to the baseline Gaussian qf, we obtain the qf of a Gaussian with the same mean, and a variance increased by one (see Figure C.2 in Appendix C).[23] We highlight two key aspects of this approach. First, due to the Kernel and the discrete approximation asymmetry, this approach keeps a stable median while altering the distribution mean, so inevitably our concept of "consensus invariance" refers to the median. Second, our approach focuses on fixing the standard deviation of the perturbing distribution, and not on the perturbed one outcome. Thus, in both the flexible approximations, it is not

[23]This general concept applies to any pair of random variables whose distribution is unimodal and parameterized with location and scale. If X_1 and X_2 are independent Normal random variables with means μ_1, μ_2 and standard deviations σ_1, σ_2, then the sum of their qfs gives a Normal qf with mean $\mu_3 = \mu_1 + \mu_2$, and standard deviation $\sigma_3 = \sigma_1 + \sigma_2$.

guaranteed that the standard deviation increases just by one. The same shock might impact differently the standard deviation and higher moments.[24]

Panel (a) of Figure 8 shows the baseline and perturbed distributions on impact after the standard deviation shock under the the Kernel and the discrete approximations. Panel (b) and (c) show the effects of the standard deviation shock on macroeconomic variables across the three approximations. The blue dashed-dotted line in Panel (b) shows that a positive dispersion shock causes a fall in industrial production and an increase in CPI. Hence, as the mean shock, a positive dispersion shock has also stagflationary effects on the economy, channeled by a sharp decrease in sentiment. However, this result holds true for the rich specifications – both Kernel and Discrete, see Panel (c) – that take into account the whole distribution, while a simple Gaussian approximation would fail to detect the stagflationary impact of a dispersion shock. In the Gaussian case, industrial production does not respond significantly on impact and then actually increases along the impulse response function, CPI barely moves, and sentiment shows a small and very transient drop (see the black dashed line in Panel (b)). The results suggest that the pass-through effect of a standard deviation shock from expectations to real economy, via sentiment, can be detected only if the entire distribution is accounted for. From a methodological point of view, the main messages from Figure 8 echoes the ones from Figure 7. First, the distribution of inflation expectations matters. Considering the entire distribution enables to detect the stagflationary effects of a change in disagreement, that is, a change in the standard deviation of inflation expectations. The findings indicate a potential pass-through effect from the standard deviation of expectations to the real economy, mediated by sentiment, which becomes evident only when the full distribution is considered. Second, while the responses of macroeconomic variables are similar for the two detailed approximations – Kernel and Discrete – the credible bands are significantly wider for the Discrete approximation. As explained above, this is due to the nature of the approximation, and thus, from now onward, we will solely show the results for the – less noisy – Kernel approximation.

Additional Dispersion Shocks. While in the Gaussian case, due to its parsimonious features, the range of possible dispersion investigations ends here, the other two flexible approximations do allow for the examination of various dispersion shocks. However, there are numerous ways to perturb a distribution, so it is necessary to identify dispersion variations of particular interest. We are here focusing on perturbations including "kurtosis" and "tails" behaviors, thus moving beyond the straightforward Gaussian scaling.[25]

- *Kurtosis shock*. Here we consider an increase in dispersion that is not a pure scaling effect (as before), but rather places more mass on the tail regions of the distribution. In order to generate this movement, we employ the qf of a centered Generalized Normal (GN) distribution with shape $\beta_{GN} = 0.5$, and scale σ_{GN} calculated indirectly to ensure that the

[24] Although this is not directly controlled, in the experiment we observe a standard deviation increase of 0.924 (0.94) for the discrete (Kernel), making the comparison reasonable. Another approach could be to perform a grid search to identify the standard deviation of the perturbing qf needed to induce a standard deviation increase of exactly one.

[25] To avoid confusion (given that now the standard deviation is a function of different parameters) we label the previous experiment about standard deviation as "Scaling", to emphasize the origin of dispersion, as opposed to the Kurtosis.

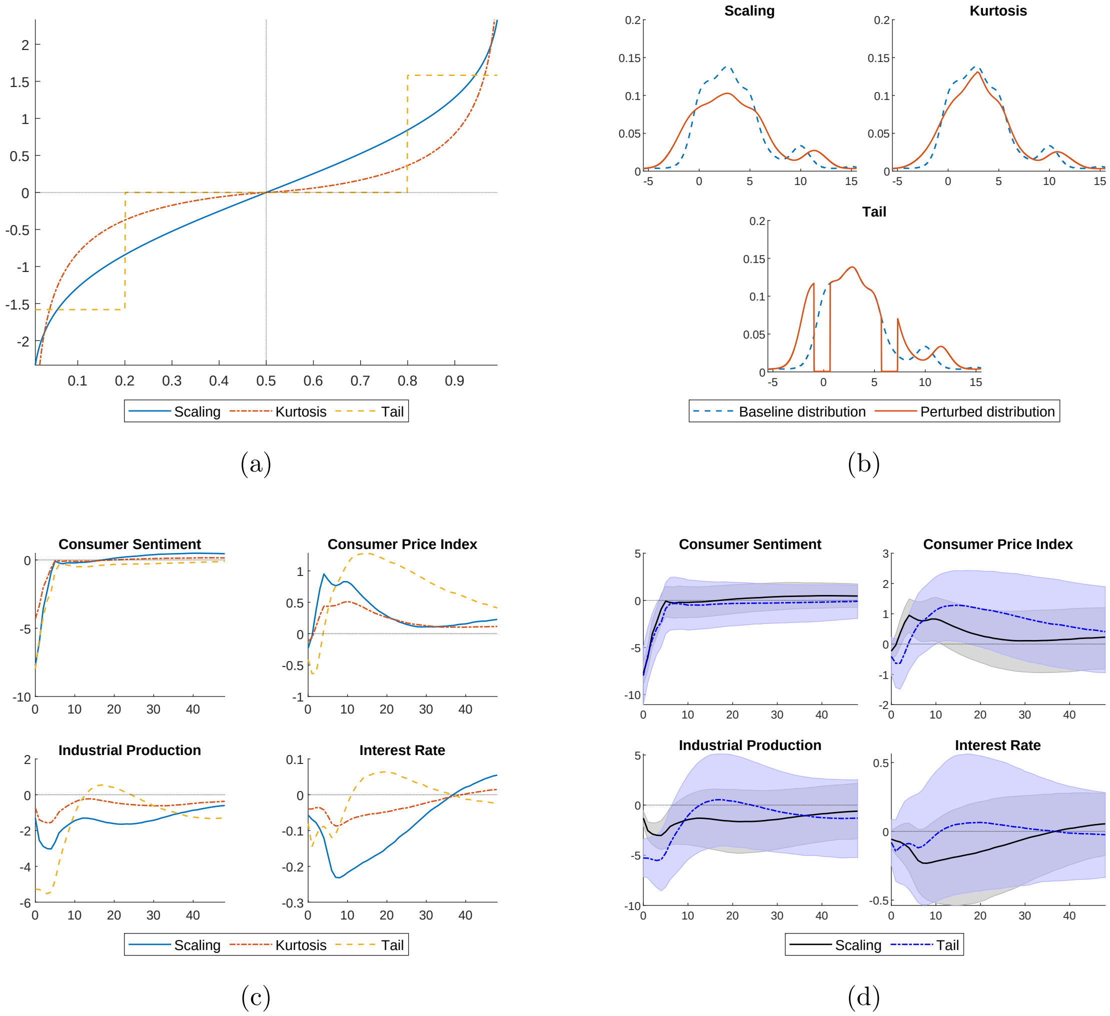


Figure 9: **Positive dispersion shocks: Tail shocks are recessionary.** The figure shows the macroeconomic effects of three dispersion shocks: Scaling, Kurtosis, and Tail. Panel (a): three quantile functions used to induce the increase in dispersion. Panel (b): contemporaneous effects of the three different perturbations on the inflation expectations distribution. Panel (c): Median IRFs of the macroeconomic variables after the three shocks. Panel (d): Median IRFs and 68% credible bands of the macroeconomic variables after two selected shocks: the Scaling (blue dashed-dotted lines), and the Tail (black lines).

resulting standard deviation is equal to 1, as in the previous standard deviation shock.[26] We label this shock as "Kurtosis". The red dashed-dotted line in Panel (a) of Figure 9 shows the perturbing qf used to induce this dispersion shock, in comparison with the pure scaling effect considered before (blue solid line), and the top-right graph in Panel (b) shows the corresponding perturbation of the inflation expectation distribution (Kernel approximation).

- *Tail shock.* While previous shocks have influenced the entire expectations domain with varying intensities, we now consider an exogenous shock that affects only a subset of the range of individuals' beliefs. Specifically, we investigate shifts in the two tails, defined by the values of the domain associated with the first and last 20% of aggregate probability (around 0.63 for the left and 5.7 for the right). The way we move them follows the rational of the location shock: the left (right) tail moves to the left (right) without altering its shape, leaving the bulk unaffected (that in this case is the central 60% of aggregated probability). As for the others, the shift is calibrated such that the standard deviation of the perturbing qf is 1. This shock does not affect the median, and it is symmetric. We label this shock as "Tail". The yellow dashed line in Panel (a) of Figure 9 displays the perturbing qf used to induce this dispersion shock, and the bottom graph in Panel (b) shows the corresponding perturbation of the inflation expectation distribution (Kernel approximation).[27]

We now investigate the macroeconomic effects of these shocks, and compare them with the results of the previous "standard deviation shock". Panel (c) of Figure 9 displays the median IRFs for all the three experiments - Scaling, Kurtosis and Tails - for sentiment and the macro variables. The graphs show that the IRFs after a Kurtosis shock exhibit a very similar shape as the one for the scaling shock, but the effects are smaller. This is not very surprising if one compares the Scaling and Kurtosis perturbations of the inflation expectations distribution in Panel (b). Both the two perturbed distributions deplete the bulk and distribute the mass on the other parts of the distribution – as evident also from the qf in Panel (a). The two shocks differs in the way they distribute the mass from the bulk to the other part of distribution. However, given the way we constructed the two shocks, the main difference regards the effect on the bulk: the Scaling perturbation affects much more the bulk than Kurtosis. That's explain why the Scaling shock has similar but larger effects.

The Tails shock is instead very different in nature. It leaves the bulk of the distribution substantially unaffected and simply stretches out the two tails of the distribution. Looking at Panel (c), the Tail shock exhibits significantly stronger effects on both industrial production and prices. Specifically, a Tail shock is not stagflationary on impact, but rather recessionary, because both industrial production and prices decrease. However, Panel (d) compares the IRFs of the Scaling and the Tail shocks, – shaded area delimit the 68% credible bands – shows that

[26] The GN distribution includes the Normal for $\beta_{GN} = 2$, and the Laplace for $\beta_{GN=1}$. In general, it features a positive excess of kurtosis for any $\beta_{GN} \in (0, 2)$.

[27] We want to stress that such an experiment is impossible to conduct with other parametric approximations, thereby confirming the validity of our "agnostic approach". While the Kurtosis shock could be examined parametrically by approximating the inflation expectations distribution either with a GN or a T distribution, these method will not accommodate such a tail type of shock.

the price level response is not significantly different from zero.

From these experiments, we learn that dispersion shocks negatively impact industrial production, consistent with a negative response of sentiment. A dispersion shock operates in two ways: it reduces the probability mass in the bulk and redistributes it across the distribution. Both features matters. The Scaling shock demonstrates that depleting the bulk has stagflationary effects: industrial production decreases while the price level rises. In contrast, the Tail shock reveals that shifting more mass to the distribution tails also has pronounced negative effects on industrial production, but it causes the price level to decrease, at least initially and for a few months.

4.3.3 Asymmetric dispersion shocks

The previous subsection shows that Tail shocks can be recessionary on impact, because they induce a negative response of the price level. Does this effect depend on one of the two tails? It is natural to conjecture that the left tail should be responsible for the negative effect on the CPI, unless tails generate puzzling effects – as, for instance, an increase in inflation expectations by households populating the right tail implies lower inflation. More generally, this Section investigates whether the effects of dispersion shocks are driven by the redistribution of probability mass from the bulk to a specific part of the distribution, namely the right or the left part.

To investigate asymmetric effects, we must decompose the contribution of left and right components of dispersion shocks analyzed so far. Given their symmetric nature, each dispersion shock's perturbing qf can be seen as a combination of two parts. To notice that, let $\tilde{q}(\tau)_s$ represent a symmetric perturbing qf defined on $\tau \in (0, 1)$. It can be represented as the sum of a left perturbing qf ($\tilde{q}(\tau)_- = \tilde{q}(\tau)_s \mathbb{1}(\tau < 0)$), and a right perturbing qf ($\tilde{q}(\tau)_+ = \tilde{q}(\tau)_s \mathbb{1}(\tau > 0)$), so that $\tilde{q}(\tau)_s = \tilde{q}(\tau)_- + \tilde{q}(\tau)_+$. Analyzing $\tilde{q}(\tau)_-$ and $\tilde{q}(\tau)_+$ separately (that are not symmetric by definition) provides insight into the marginal contribution of the two sides of the distribution in the overall symmetric effect. Panel (a) and (b) of Figure 10 visualizes the contemporaneous perturbation applied to the distribution in the case of Scaling and Tail shocks, respectively, for the cases of left tail, right tail and symmetry (as in the previous subsection). Note how these are the asymmetric cases of the experiments above. In the case of a Scaling shock, the probability mass is taken from the bulk and redistributed uniformly. However, this happens either just for the left part of the distribution (the top-left graph in Panel (a), i.e., $\tilde{q}(\tau)_-$) or the right one (the top-right graph in Panel (a), i.e., $\tilde{q}(\tau)_+$). Panel (b) displays similar effects but for tail shocks, which increase the mass on the left or right distribution tails.

Panel (c) and (d) in Figure 10 show that the two asymmetric shocks have very different effects on the macroeconomic variables, for both the Scaling and Tail shock. Let us first focus on the effects of the asymmetric Scaling shock on the macroeconomic variables in Panel (c). Note that the decomposition of the symmetric case into $\tilde{q}(\tau)_-$ and $\tilde{q}(\tau)_+$ is exact only for initial perturbed distributions. However, we expect that the sum of the two median IRFs of the two asymmetric experiments approximate satisfactorily the symmetric one. The initial negative response of IP – bottom-left in Figure (c) – in the symmetric case (black solid line) is entirely due to the movements in the left tails of the distribution, shown by the blue dashed line. The same is true for the negative response of sentiment. The response of the CPI is as one would

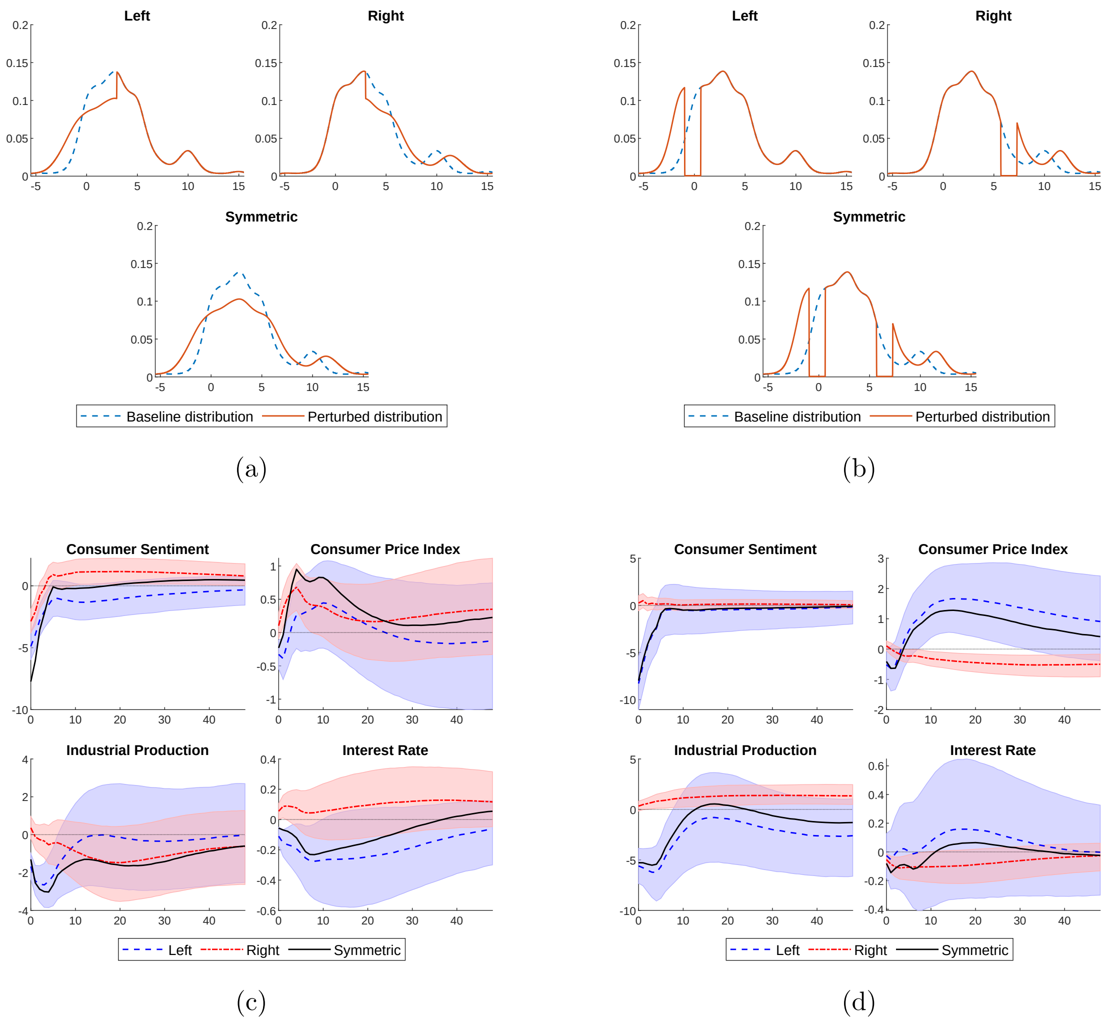


Figure 10: **Left-tail perturbations account for most of the macroeconomic effects.** The figure shows the macroeconomic effects of the asymmetric version of two dispersion shocks: the Scaling in Panel (a,c), and the Tail in Panel (b,d). Panel (a): contemporaneous effect of the left, right, and symmetric version of the Scaling shock. Panel (b) shows the same but for the Tail shock. Panel (c): IRFs of the macroeconomic variables after the three Scaling shock. For the asymmetric part we also show the 68% credible bands. Panel (d): shows the same but for the Tail shock.

expect: movements in the left tail induce a decrease in the CPI on impact, and the opposite for movements in the right tail, so that the CPI response for the symmetric case is not significantly different from zero. Finally, note that these differences are statistically significant, at least for the initial part of the IRFs.

The relative higher relevance of movements in the left tail is even more pronounced in the case of a Tail shock. In this case, shifting the right tail of the distribution to the right has barely any effect on sentiment, IP and CPI. The dynamics of the IRFs of the macroeconomic variables in the symmetric case are entirely driven by the stretching of the left tail of the distribution. Specifically, both the strong negative effects on sentiment and IP and the response of CPI almost coincides with the IRFs caused by the left tail movement. Hence, it is not surprising that the initial response of CPI is negative, inducing a recessionary effect on impact. The responses to a right tail shock are muted, and qualitatively go on the opposite direction: IP increases and CPI drops, while sentiment does not respond. Therefore, the answer to the initial question about which tail causes the symmetric tail shock to be recessionary is clear: it is all about the left tail shock. Moreover, this does not apply only to the response of CPI, as conjecture, but also to the negative response of IP.

By decomposing the symmetric shock in perturbations of the left and right part of the distribution, this analysis underscores the importance of asymmetries in understanding the transmission mechanisms of dispersion shocks, highlighting the dominant role played by movements in the left part of the distribution for both Scaling and Tail shocks. The Tail shock, in particular, illustrates that movements in the extreme tails of the distribution have opposite effects on IP and CPI, and that the left tail ultimately determines the outcome of the Tails shock as recessionary.

Who populates the left tail? The empirical evidence that emerged from the asymmetric dispersion shocks stimulates further investigation. Why do left-exogenous movements generate bigger macroeconomic fluctuations? To try to answer this question, we study the individual characteristics of the households forming the MSC. The MSC is a comprehensive survey, collecting inflation expectations, but also a variety of other information, including income and wealth, demographic statistics, personal finance, expectations and so on. Specifically, we want to assess if the probability of having low inflation expectations (i.e., being within the left part of the distribution) depends on some individual features.

Table 3 shows the results from two Pooled OLS regressions with time-fixed effects.[28] The dependent variable is a binary transformation of inflation expectations that takes the value of 1 if the individual i inflation expectation lies in a given interval, A. In row 1 the interval represents the full left tail, while the second row filters out excessively low values of inflation expectations. The results are the same. Remarkably, the probability of being in the tails is positively related to the probability of having a high income. Figure 11 shows that the distribution of the correlation between income and inflation expectations across time – i.e., given the cross-section for each time t – is significantly negative. Moreover, despite showing a downward bias in inflation being in the left tail, these households tend to be more optimistic, in terms of sentiment. Moreover,

[28] Individual fixed effects are not considered, given that the MSC has a rotating panel structure, and the same household remains in the survey for no longer than three periods. Thus, it is impossible to track the story of an individual across time.

$$\mathrm{P}\left[\mathrm{E}_{i,t}(\pi_{t+12}) \in A\right] = \mu_t + \mathbf{x}_{i,t}^{'}\boldsymbol{\beta} + \epsilon_{i,t}$$

Specification	**Hold Diploma**	**Hold Degree**	**Income (quintiles)**	**If Invest**	**Sentiment**	**R-squared**
$A = [-50, 1]$	-0.0099	0.0048	0.0100***	-0.0054*	0.0664***	0.0323
	(0.0065)	(0.0029)	(0.0012)	(0.0032)	(0.0012)	
$A = [0, 1]$	-0.0099	0.0046	0.0094***	-0.0024	0.0633***	0.0302
	(0.0063)	(0.0028)	(0.0011)	(0.0031)	(0.0011)	
$A = [0, 2]$	-0.0109	0.0105***	0.0124***	0.0205***	0.0789***	0.0407
	(0.0069)	(0.0031)	(0.0013)	(0.0034)	(0.0012)	

Table 3: **Pooled OLS regression on inflation expectations.** The dependent variable takes the value of 1 if individual i has an inflation expectation that falls within the set A, 0 otherwise. The regression includes both time and fixed effects. Number of observations = 114.847. For the sake of parsimony, we show the results of some relevant variables. The full specification is available in Table D.1 in Appendix D.

the coefficients suggest that people in the left tail of the distribution are relatively more educated – i.e., higher probability of holding a degree and lower of holding a diploma – even if they are only marginally significant (15% level). These results suggest that people in the left tail of the distribution are relatively more educated and have a relatively higher capacity to spend, given the higher income, providing some indirect support for the transmission mechanism of expectation shocks to the macroeconomic variables that we found in the BVAR analysis.

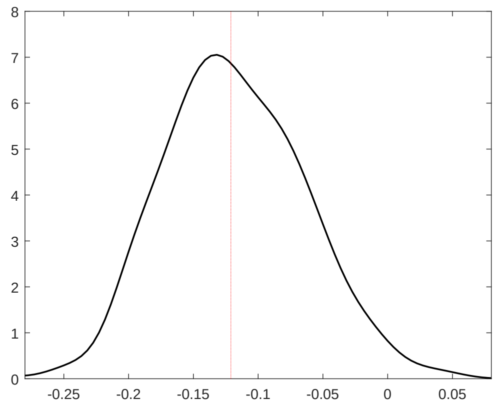
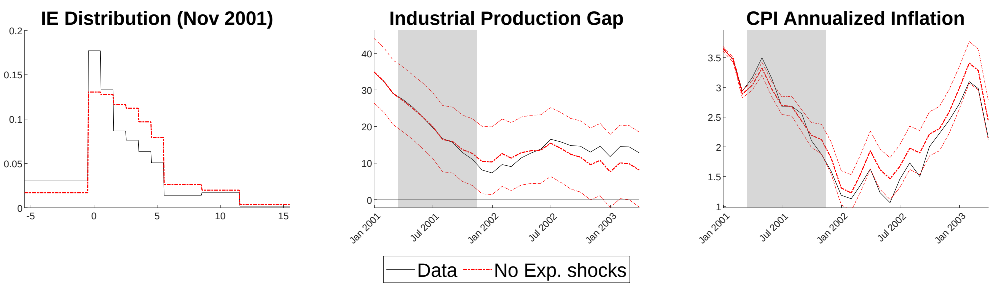


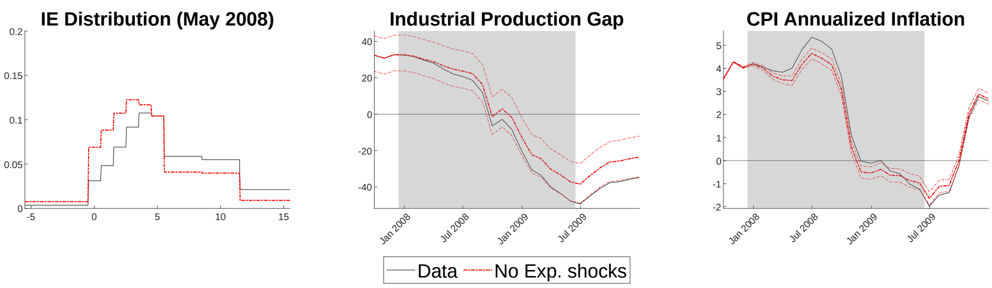
Figure 11: **The relation between inflation expectations and income is significantly negative.** The figure shows the distribution across time of the correlation between inflation expectations and income calculated using the cross section for each time t.

4.4 The role of shocks to inflation expectations during recession events

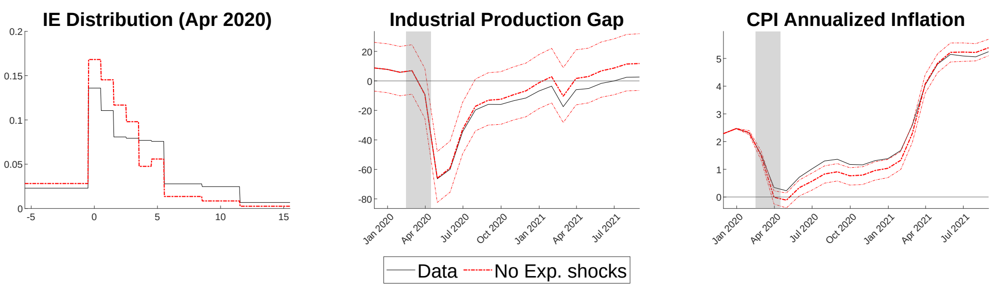
The goal of the experiments carried out so far was to investigate the effects of different perturbations to the inflation expectation distribution – Location, Scale, Kurtosis, Tail – on macroeconomic variables of interest. We learned that location and dispersion shocks are stagflationary, while tail shocks are recessionary, with especially strong effect coming from the left tail. However, the assumed perturbations are not necessarily observed in the sample considered, so one



(a) Dot-Com



(b) GFC



(c) Covid-19

Figure 12: **Shocks to the distribution of inflation expectations have relevant macroeconomic impacts during recessions.** The figure shows the effects of shocks to expectations during three recessions caused by the dot-com bubble (a), the GFC (b) and the Covid-19 outbreak (c). The first panel of each row compares observed and counterfactual inflation expectation distributions at the end of the horizon of each exercise. The central panels display observed and counterfactual industrial production, expressed in deviation from the estimated trend. The left panels compare annual inflation in percent. Data are plotted in continuous black lines and counterfactual developments in dashed red lines. For the counterfactual output and inflation we plot the 68% credible bands that account for the estimated parameter uncertainty.

might wander how relevant they are in the real world. Hence, we now analyze movements in the inflation expectation distributions during the last three recessions: the 2001 episode that followed the collapse of the dot-com bubble; the Great Financial Crisis (GFC) and the Covid-19 pandemic.

To understand the role of expectations, we construct counterfactual exercises for the three recession episodes, calibrating to zero the estimated shocks to inflation expectations for a given period of time. Figure 12 shows the effects of expectation shocks in the three cases, comparing the observed data (continuous black lines) and the counterfactual behavior of the inflation expectation distribution, the output gap and inflation (dashed red lines).

The first row of Figure 12 shows the results for the 2001 episode. We construct the experiment calibrating the expectation shocks to zero from March to November 2001: the period corresponding to the NBER recession. In the upper-left panel we compare the actual and counterfactual distributions of inflation expectations at the end of the horizon, that is November 2001.[29] The shocks to expectations operate a shift of the mean to the left (from 2.2% to 0.4%), and an important increase in the left tail (the proportion of respondent that expect negative numbers increase from 17% to 30%). In Section 4.3 we find that positive location shocks are stagflationary while tail shocks are recessionary. Then, a negative shift of the mean and a positive increase in the left tail have counteracting effects on output (upper-central panel) making the overall effect small.[30] On the other hand, both the increase in the left tail and the decrease in the mean have negative effects on inflation, with a peak of 0.4% in May 2002.

The counterfactual exercise for the GFC is constructed setting expectation shocks to zero for the first six months of the NBER recession, that is from December 2007 until May 2008. The central-left panel of Figure 12 shows the counterfactual and the observed distributions in May 2008: they have similar shapes, with the observed one clearly translated to the right. Shocks to expectations produce an increase of both the first and second moments: the mean is equal to 6.9% in the observed data, and to 4.4% in the one without expectation shocks; the standard deviation is equal to 5.7 in the data and to 5.2 in the counterfactual exercise; finally, the skewness is substantially the same for the two distributions. Then, we interpret the effects of the exogenous variation in expectations as a combination of a positive location and positive dispersion shocks. In Section 4.3 we find that both kind of disturbances have stagflationary effects. Without the impact of expectation shocks, output would have dropped less reducing the trough of the crisis of about 20%. The positive effect on inflation comes earlier: in July 2008 higher inflation expectations account for about 1% CPI inflation, partly contributing to explain the "missing disinflation" puzzle (e.g., Coibion and Gorodnichenko, 2015).

For the Covid-19 pandemic, we shut down expectation shocks for three months: February, March and April 2020.[31] Comparing the distributions in April 2020 we note that expectations

[29] Also for the other recession episodes we show the last month of the counterfactual exercise as representative of the effects of expectation shocks to the distribution. A richer set of distributions during the counterfactual periods are reported in Figures D.4-D.6 in Appendix D.

[30] The negative impact from the tail shock initially prevails, while the mean shock predominates at longer horizons. This result is consistent with the estimated DIRFs, as clear from panel (c) of Figure 7 and panel (c) of Figure 10.

[31] The Covid-19 pandemic period is outside our estimation sample. To construct the counterfactual, we use the posterior distribution of the parameters estimated in the original sample that ends in December 2019.

shocks move probability mass from the bulk to the right tail of the distribution (bottom-left panel of Figure 12). The effects are very comparable to what we label as "asymmetric Scaling shock", that for the right tail has a stagflationary impact: see panels (a) and (c) of Figure 10. To corroborate our interpretation, note that differently from the GFC, here the increase in the mean and the dispersion is also combined with an increase in the skewness, from 0.28 to 0.39. The bottom-central panel shows that expectation shocks have slowed down the post-pandemic recovery, with the industrial production gap closing about one year earlier in the counterfactual exercise. Finally, from the bottom-right panel we see a positive impact on inflation consistent with the nature of the expectation shock.

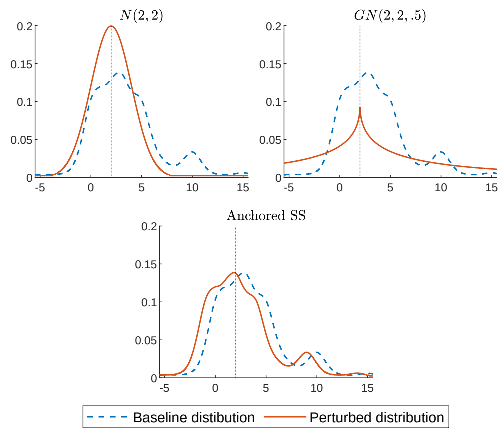
4.5 Central bank's communication

In this subsection, we investigate the potential implications of our results for monetary policy communication. Despite households are generally inattentive to monetary policy, our results suggest that there might be scope for a communication strategy, that should aim not only to convey the consensus (mean) or target, but to affect the whole distribution or, better, some targeted location in the distribution. Assume communication is effective so that it can influence households' expectations, then, one could ask whether it is more important to change the mean of the distribution towards the 2% target or to decrease the disagreement/uncertainty (dispersion shocks) or to target some specific part of the distribution (asymmetric tail shocks).

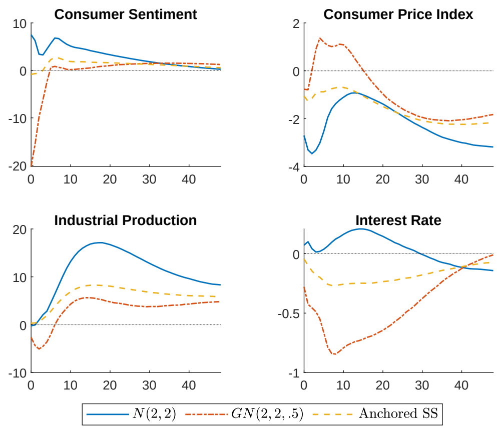
In order to investigate this, we study the effects of exogenous distribution shocks that discipline the expectations toward the 2% inflation target, as if these shocks were the result of effective communication. More specifically, we devise three scenarios, featuring different effects on the shape of the distribution on impact. The first scenario shifts the steady state distribution to the left so that the mean is equal to 2% – hence, we label it "Anchored steady state". This case describes a central bank communication policy which tries to steer the mean of the distribution, disregarding the shape. The second scenario corresponds to the, possibly ideal, case in which the communication shock makes the initial distribution of expectations equal to a Normal with mean and standard deviation equal to 2. In this case the cross-sectional dispersion is limited – the signal-to-noise ratio here is $\mu/\sigma = 1$ – and fairly disciplined. In the third scenario, we assumes that after the shock, the distribution is a GN with a mean and scale of 2 and a shape parameter of 0.5, which implies heavier tails than for the Normal distribution. This case corresponds to a communication policy able to steer the median to the target, but leading to higher dispersion.[32]

The first Panel of Figure 13 provides, as before, a graphical representation of the movements on impact of the distribution of expectations under the three scenarios, whereas Panels (b) and (c) show the IRfs of macroeconomic variables. Looking at Panel (b), a communication policy able to shift the median to the 2% target – yellow dashed lines – induces a positive and persistent response of IP – as well as sentiment – and a negative and persistent response of the CPI. This suggests that shifting inflation expectations to target is beneficial for the real

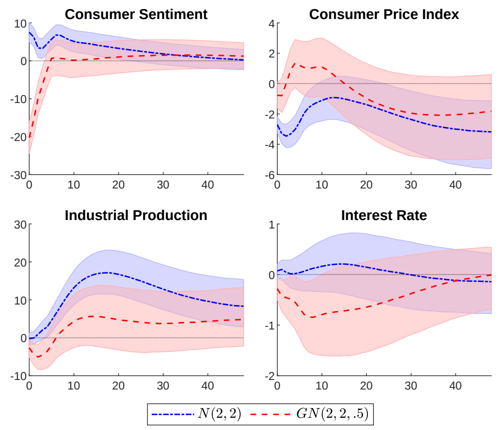
[32] Notably, we do not adjust the scale to make the last two shocks comparable as we did before. This is largely due to the first scenario only entailing a shift in the baseline distribution, thus precluding any meaningful comparison.



(a)



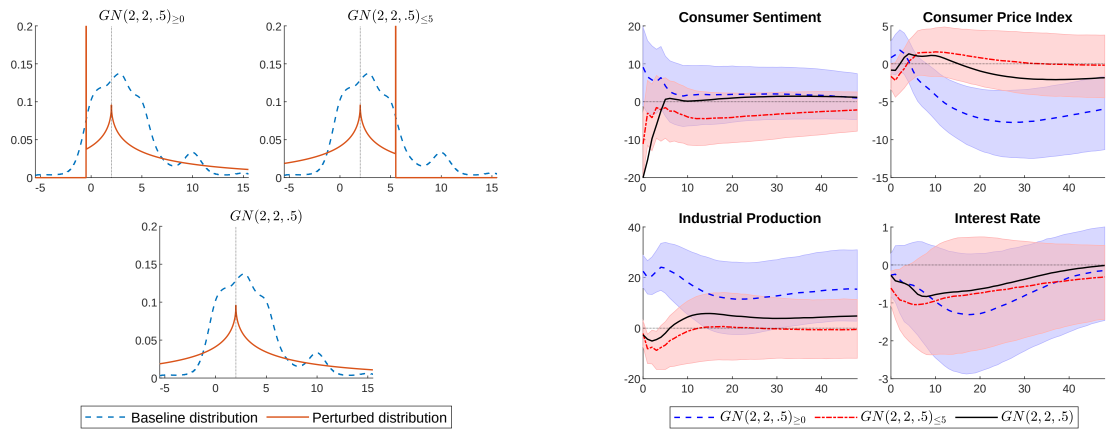
(b)



(c)

Figure 13: **Macroeconomic effects of three communication shocks.** Panel (a): Baseline and perturbed distributions of the three shocks considered. Panel (b): Median IRFs of the macroeconomic variables after the three shocks. Panel (c): comparison between the Normal and the Generalized Normal (fat tails) shocks together with the 68% credible bands.

economy, as this is a Location shock as in Figure 7 – recall from above that the median of the steady state distribution is 3.6%. The responses of the macroeconomic variables when the the initial distribution translates into a Normal $N(2,2)$ – solid blue lines – are very similar to those obtained when the shock produces a shift in the median, but much more pronounced. This demonstrates the additional, and substantial, benefit due to the decrease in the dispersion of the inflation expectation distribution, as we saw in the case of dispersion shocks. The third scenario stresses even more the importance of dispersion. This case represents a "muddled" communication policy, that is able to steer the median to target, but leads to a greater dispersion of the distribution. The responses of macroeconomic variables – dashed-dotted red lines – differ from the other two cases. Specifically, consumer sentiment drops on impact, the initial response of industrial production is negative, and the CPI, after an initial temporary drop, increases for some time. Panel (c) compares the second and third scenarios, showing that the responses of macroeconomic variables in these two cases are indeed statistically different. These differences stress the importance of the tails of the distribution: it is not sufficient to cause a shift of



(a) (b)

Figure 14: **Macroeconomic effects of truncated versions of the GN shock.** Panel (a): Baseline and perturbed distributions associated with the shocks. Panel (b): IRFs of the macroeconomic variables. Median for the non-truncated GN(2,2,0.5); median and 68% credible bands for the truncated distributions.

the mean towards the target, but communication should avoid leaving mass on the tails of the distribution.

Given the results in the previous subsections, logically, a further question follows: are the two tails alike, or should communication target a specific tail? The next experiment thus truncates the left and right tails in the $GN(2,2,0.5)$ distribution, as shown in Panel (a) of Figure 14. The probability masses in the left and right tails collapse to a value not lower than zero – $GN(2,2,0.5)_{\geq 0}$, top-left graph in Panel (a) – or not larger than 5 – $GN(2,2,0.5)_{\leq 5}$, top-right graph in Panel (a) – respectively. The comparison between the IRFs in Panel (b) shows that the negative reaction of sentiment and IP under the $GN(2,2,0.5)$ distribution is entirely driven by the (fat) left tail of the distribution. In the absence of the left tail, i.e., $GN(2,2,0.5)_{\geq 0}$, the responses of the macroeconomic variables are in line with the ones of the Normal $N(2,2)$ analyzed in the first scenario: sentiment and IP responses are positive and the CPI decreases persistently. Hence, a communication strategy that targets the left tail of the distribution, but leaves the right tail fat, causes a similar reaction as an ideal communication scenario, depicted here as a Normal $N(2,2)$. This result suggests that a central bank should focus its communication on convincing consumers that inflation will not be too low – negative in our experiment. Since these values can be considered evidently unrealistic, this strategy seems relatively simpler, but most of the improvements in the real activity seem to emerge through it.

5 Conclusion

We have developed a Monetary Policy BVAR, augmented with heterogeneous beliefs from the MSC, in order to shed light on the macroeconomic effects of shocks to the entire distribution of short-term inflation expectations. In line with Reis (2022), the main message of this paper lies in acknowledging the existence of essential information content in the higher order moments of the inflation expectation distribution that should be considered in applied work. The periphery of the distribution, often overlooked in parametric approaches, emerges as a location driving substantial macroeconomic fluctuations.

Our analysis reveals the following main results. First, and somewhat surprisingly, households' inflation expectations affect macroeconomic variables, but not as much vice versa. Second, shocks that increase the first two moments of the distributions are stagflationary, and models based on just mean and variance – without accounting for the whole cross-sectional heterogeneity – would underestimate the macroeconomic effects of these shocks. Third, the transmission mechanism of these shocks goes through consumer sentiment, and thus through the relationship between inflation expectations and the perceived future outcome by consumers, in line with the microeconomic survey evidence (e.g., Coibion et al., 2019, 2023; Weber et al., 2022). Fourth, shocks that stretch the tails of the inflation expectation distribution out are recessionary, since both industrial production and prices decrease. Fifth, by introducing asymmetry, we provide further evidence on the specific portions of the distribution that trigger the results. In particular, it emerges that left-tail perturbations account for the largest macroeconomic effects. Sixth, we show that households populating the left tail have relatively higher income and are more educated, suggesting that our findings are not due to poorly educated households with no spending ability. Seventh, we show that exogenous movements in inflation expectations played a role in shaping the dynamics of output and inflation during the last three recessions. Finally, we discuss implications of our results for central bank communication, revealing a possible complex trade-off for central banks. Communication should focus more on decreasing dispersion, and specifically moving mass from the left tail, where pessimistic inflation expectations lie, rather than on moving the mean towards the 2% target. In other words, communication targeting pessimistic inflation expectations might be the most effective.

Methodologically, our work contributes to the growing literature on addressing dimensionality challenges when incorporating heterogeneous beliefs into macroeconometric models. Our approach enables us to directly perturb the expectations of specific classes of agents for meaningful economic insights.

References

Allayioti, A., F. Monti, and M. Piffer (2023): "The transmission of monetary policy when agents fear extreme inflation outcomes," Tech. rep., mimeo.

Angeletos, G.-M., F. Collard, and H. Dellas (2018): "Quantifying Confidence," *Econometrica*, 86, 1689–1726.

Antolín-Díaz, J., I. Petrella, and J. F. Rubio-Ramírez (2021): "Structural scenario analysis with SVARs," *Journal of Monetary Economics*, 117, 798–815.

Ascari, G., S. Fasani, J. Grazzini, and L. Rossi (2023): "Endogenous uncertainty and the macroeconomic impact of shocks to inflation expectations," *Journal of Monetary Economics*.

Barrett, P. and J. J. Adams (2022): "Shocks to Inflation Expectations," Working Paper 2022/072, International Monetary Fund.

Barsky, R. B. and E. R. Sims (2011): "News shocks and business cycles," *Journal of Monetary Economics*, 58, 273–289.

——— (2012): "Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence," *American Economic Review*, 102, 1343–77.

Beaudry, P. and F. Portier (2006): "Stock Prices, News, and Economic Fluctuations," *American Economic Review*, 96, 1293–1307.

——— (2014): "News-Driven Business Cycles: Insights and Challenges," *Journal of Economic Literature*, 52, 993–1074.

Benhabib, J., P. Wang, and Y. Wen (2015): "Sentiments and Aggregate Demand Fluctuations," *Econometrica*, 83, 549–585.

Binder, C. and C. Makridis (2022): "Stuck in the Seventies: Gas Prices and Consumer Sentiment," *The Review of Economics and Statistics*, 104, 293–305.

Binder, C. and J. Ryngaert (2024): "Consumer and Firm Inflation Expectations," Forthcoming in the *Research Handbook of Inflation*, edited by G. Ascari and R. Trezzi, Edward Elgar Publishing, UK.

Binder, C. C. (2017): "Measuring uncertainty based on rounding: New method and application to inflation expectations," *Journal of Monetary Economics*, 90, 1–12.

——— (2020): "Long-run inflation expectations in the shrinking upper tail," *Economics Letters*, 186, 108867.

Blinder, A. S., M. Ehrmann, M. Fratzscher, J. De Haan, and D.-J. Jansen (2008): "Central bank communication and monetary policy: A survey of theory and evidence," *Journal of Economic Literature*, 46, 910–945.

CANDIA, B., O. COIBION, AND Y. GORODNICHENKO (2020): "Communication and the beliefs of economic agents," Working Paper 27800, National Bureau of Economic Research.

CARRIERO, A., T. E. CLARK, AND M. MARCELLINO (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137–154.

CHAN, J. C. (2021): "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 37, 1212–1226.

CHAN, J. C. AND I. JELIAZKOV (2009): "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 101–120.

CHANG, M., X. CHEN, AND F. SCHORFHEIDE (2024): "Heterogeneity and aggregate fluctuations," *Journal of Political Economy*, 132, 000–000.

CHEN, Y., Z. LIN, AND H.-G. MÜLLER (2021): "Wasserstein Regression," *Journal of the American Statistical Association*, 0, 1–14.

CLARK, T. E. AND T. DAVIG (2011): "Decomposing the declining volatility of long-term inflation expectations," *Journal of Economic Dynamics and Control*, 35, 981–999.

COIBION, O., D. GEORGARAKOS, Y. GORODNICHENKO, AND M. VAN ROOIJ (2023): "How Does Consumption Respond to News about Inflation? Field Evidence from a Randomized Control Trial," *American Economic Journal: Macroeconomics*, 15, 109–52.

COIBION, O. AND Y. GORODNICHENKO (2015): "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, 7, 197–232.

COIBION, O., Y. GORODNICHENKO, AND T. ROPELE (2019): "Inflation Expectations and Firm Decisions: New Causal Evidence," *The Quarterly Journal of Economics*, 135, 165–219.

DIEBOLD, F. X. AND C. LI (2006): "Forecasting the term structure of government bond yields," *Journal of Econometrics*, 130, 337–364.

D'ACUNTO, F., U. MALMENDIER, AND M. WEBER (2022): "What Do the Data Tell Us About Inflation Expectations?" Working Paper 29825, National Bureau of Economic Research.

HAYFIELD, T. AND J. S. RACINE (2008): "Nonparametric econometrics: The np package," *Journal of statistical software*, 27, 1–32.

HORVÁTH, L. AND P. KOKOSZKA (2012): "Inference for Functional Data with Applications," .

INOUE, A. AND B. ROSSI (2019): "The effects of conventional and unconventional monetary policy on exchange rates," *Journal of International Economics*, 118, 419–447.

JURADO, K., S. C. LUDVIGSON, AND S. NG (2015): "Measuring Uncertainty," *American Economic Review*, 105, 1177–1216.

KAMDAR, R. (2019): "The Inattentive Consumer: Sentiment and Expectations," 2019 Meeting Papers 647, Society for Economic Dynamics.

KOKOSZKA, P., H. MIAO, A. PETERSEN, AND H. L. SHANG (2019): "Forecasting of density functions with an application to cross-sectional and intraday returns," *International Journal of Forecasting*, 35, 1304–1317.

MANKIW, N. G., R. REIS, AND J. WOLFERS (2003): "Disagreement about Inflation Expectations," *NBER Macroeconomics Annual*, 18, 209–248.

MEEKS, R. AND F. MONTI (2023): "Heterogeneous beliefs and the Phillips curve," *Journal of Monetary Economics*.

NERI, S. (2023): "Long-term inflation expectations and monetary policy in the euro area before the pandemic," *European Economic Review*, 154, 104426.

PETERSEN, A., C. ZHANG, AND P. KOKOSZKA (2022): "Modeling Probability Density Functions as Data Objects," *Econometrics and Statistics*, 21, 159–178.

RAMSAY, J. O. AND B. W. SILVERMAN (2005): *Functional Data Analysis*, Springer.

REIS, R. (2022): "Losing the inflation anchor," *Brookings Papers on Economic Activity*, 2021, 307–379.

VILLANI, M. (2009): "Steady-state priors for vector autoregressions," *Journal of Applied Econometrics*, 24, 630–650.

WANG, M.-C. AND J. VAN RYZIN (1981): "A class of smooth estimators for discrete distributions," *Biometrika*, 68, 301–309.

WEBER, M., F. D'ACUNTO, Y. GORODNICHENKO, AND O. COIBION (2022): "The Subjective Inflation Expectations of Households and Firms: Measurement, Determinants, and Implications," *Journal of Economic Perspectives*, 36, 157–84.

ZHANG, C., P. KOKOSZKA, AND A. PETERSEN (2022): "Wasserstein autoregressive models for density time series," *Journal of Time Series Analysis*, 43, 30–52.

ZHU, C. AND H.-G. MÜLLER (2023): "Spherical autoregressive models, with application to distributional and compositional time series," *Journal of Econometrics*.

A Distributional Approximation and Parameter Reduction

In this appendix, we provide a more detailed discussion of the distributional approximation introduced in Section 3. First, we show how the form of $\mathbf{H}_D$, that determines the piecewise constant approximation, derives directly from the way we aggregate probabilities in equation (4). Then, we show how the method effectively reduces the number of parameters compared to the high-dimensional VAR that would be estimated without probability aggregation, thereby making the approach operational.

As in the main text, we focus on the discrete case, as in practice one does not use the continuous density $f_t^K(\pi)$, but rather a discrete approximation, which makes it vector valued. In our case, this is given by $\mathbf{f}_t^K$, that is a $101/\delta$ vector obtained by evaluating f_t^K evaluated at the thin grid $[-50, -50+\delta, \ldots, 50-\delta, 50]$, where we use $\delta = 1/100$. Once the infinite-dimensional object is mapped into the finite-dimensional vector, $\mathbf{p}_t$ and $\mathbf{f}_t^K$ can conceptually be treated in the same way. The left hand side of equation (4) can be rewritten as

$$\mathbf{w}_t^D = \tilde{\mathbf{H}}_D \mathbf{p}_t$$

where $\mathbf{H}_D$ is $[M \times 101]$ selection matrix given by:

$$\tilde{\mathbf{H}}_D = \begin{bmatrix} \boldsymbol{\iota}'_{q_1-q_0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\iota}'_{q_2-q_1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{0} & & \boldsymbol{\iota}'_{q_M-q_{M-1}} \end{bmatrix}$$

By expressing $\mathbf{p}_t$ as a function of $\mathbf{w}_t^D$ one needs to invert $\tilde{\mathbf{H}}_D$. This can be done via $\mathbf{H}_D = \tilde{\mathbf{H}}_D^{+}$, where $\mathbf{X}^{+}$ is the Moore-Pensore pseudo inverse of $\mathbf{X}$. The matrix $\tilde{\mathbf{H}}_D$ has a block structure, where each block of is of rank-1. For this reason, the pseudo inverse can be computed as the transpose of the pseudo inverse of each normalized block, resulting in the following form for $\mathbf{H}_D$:

$$\mathbf{H}_D = \begin{bmatrix} \frac{\boldsymbol{\iota}_{q_1-q_0}}{q_1-q_0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{\boldsymbol{\iota}_{q_2-q_1}}{q_2-q_1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{0} & & \frac{\boldsymbol{\iota}_{q_M-q_{M-1}}}{q_M-q_{M-1}} \end{bmatrix}$$

Suppose one is interested in fitting a VAR on the stacked vector $\mathbf{x}_t = [\mathbf{y}_t', \mathbf{p}_t]'$, where $\mathbf{y}_t$ is a N_y vector of macroeconomic variables. Without loss of generality, we model $\mathbf{x}_t$ as a VAR(1) without intercept. This specification defines the first N_y equations as:

$$\mathbf{y}_t = \boldsymbol{\Phi}_{y,y}\mathbf{y}_{t-1} + \boldsymbol{\Phi}_{y,p}\mathbf{p}_{t-1} + \mathbf{e}_{y,t}$$

where $\mathbf{e}_{y,t}$ is a N_y dimensional error of correlated innovations, $\boldsymbol{\Phi}_{y,y}$ and $\boldsymbol{\Phi}_{y,p}$ are, respectively, $[N_y \times N_y]$ and $[N_y \times 101]$ dimensional matrices to be estimated. The latter is very high dimensional and unfeasible to estimate in a full Bayesian setting, even in this simplistic setting with a single lag. With the Kernel case the problem is even magnified, given that the dimension of $\mathbf{f}_t^K$ is $1/\delta = 100$ higher than $\mathbf{p}_t$. Our proposed aggregation scheme replaces $\mathbf{p}_t$ with $\mathbf{w}_t^D$, leading to the following approximation for $\mathbf{\Phi}_{y,p}$:

$$\mathbf{\Phi}_{y,p} \approx \tilde{\mathbf{\Phi}}_{y,p}\tilde{\mathbf{H}}_D$$

where $\tilde{\mathbf{\Phi}}_{y,p}$ is $[N_y \times M]$ dimensional, whose estimation can be easily handled.

B SS-BVAR implementation details

This section gives the details on the hyperparameters setting and the posterior sampling steps for the model employed in this study. We allow for an asymmetric shrinkage among equations by setting $\lambda_2 = 0.2$, and penalize higher order lags more strongly with $\lambda_p = 1.5$. Regarding the prior scale matrix and degrees of freedom, we set $\mathbf{\Sigma}_0 = \text{diag}(\sigma_1^2, ..., \sigma_N^2)$ and $\nu = N + 2$, where σ_i^2 are the AR(4) scaling used to inform the prior variance of the lagged coefficients. Regarding the hyperprior of λ_1, we set $\alpha_1 = \alpha_2 = 0.001$.

For the posterior sampling steps, let $\mathbf{y} = [\mathbf{y}_1', ..., \mathbf{y}_T']'$ be the stacked version of the dataset. Given the prior framework outlined above and the likelihood function implied by the model in eq. (8), we can simulate from the joint posterior distribution using a Gibbs sampler that sequentially samples from the conditional posterior distributions: $\pi(\mathbf{\Sigma}|\mathbf{y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \lambda_1)$, $\pi(\boldsymbol{\phi}|\mathbf{y}, \mathbf{\Sigma}, \boldsymbol{\theta}, \lambda_1)$, $\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{\Sigma}, \boldsymbol{\phi}, \lambda_1)$, and $\pi(\lambda_1|\mathbf{y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{\Sigma})$.

- The sampling steps of $\mathbf{\Sigma}|\mathbf{y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \lambda_1$, and $\boldsymbol{\theta}|\mathbf{y}, \mathbf{\Sigma}, \boldsymbol{\phi}, \lambda_1$ are standard (Villani, 2009).

- Sampling $\boldsymbol{\phi}|\mathbf{y}, \mathbf{\Sigma}, \boldsymbol{\theta}, \lambda_1$ is also standard. Given the high dimensionality of the system, we employ the correct version of the triangular algorithm developed by Carriero et al. (2019) in conjunction with the precision sampler of Chan and Jeliazkov (2009).

- The full conditional posterior of λ_1 is Inverse Gamma:

$$\pi(\lambda_1|\mathbf{y}, \boldsymbol{\phi}, \mathbf{\Sigma}, \boldsymbol{\theta}) \propto \mathcal{IG}\Big(\frac{N^2P}{2} + \alpha_1, \frac{\boldsymbol{\phi}'\mathbf{C}^{-1}\boldsymbol{\phi}}{2} + \alpha_2\Big)$$

  where $\mathbf{C} = \text{diag}(c_1, ..., c_{N^2P})$, is a diagonal matrix that stores the constant part of the prior variance of each ϕ_i, $i = 1, ..., N^2P$.

- In an auxiliary specification, we assign different shrinkages to different portions of the system. Let's consider $\lambda^{(m\leftarrow e)}$, the one associated to the lagged parameters that belong to the expectations to macro block. To sample it, we must first define the index set $\mathcal{S}_{m\leftarrow e}$ that collect the indexes i such that θ_i is a coefficient associated with the macro to expectation block. The conditional posterior is given by

$$\pi(\lambda^{(m\leftarrow e)}|\mathbf{y}, \boldsymbol{\phi}, \mathbf{\Sigma}, \boldsymbol{\theta}) \propto \mathcal{IG}\Big(\frac{\#\mathcal{S}_{m\leftarrow e}}{2} + \alpha_1, \sum_{i\in\mathcal{S}_{m\leftarrow e}} \frac{\phi_i}{2c_i} + \alpha_2\Big)$$

C Scenario analysis

This section explains how the scenario analysis is conducted. The goal is to draw the structural shock that generates the movements of the N_q quantiles we have in mind, using just the shocks of the quantiles and the sentiment. We start by rewriting compactly the SS-BVAR in equation (8) in structural form for $t = T + 1$:

$$\mathbf{\Phi}_0 \bar{\mathbf{z}}_{T+1} = \mathbf{c} + \mathbf{u}_{T+1}, \quad \mathbf{u}_{T+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N) \tag{C.1}$$

where $\bar{\mathbf{z}}_t = \mathbf{z}_t - \boldsymbol{\mu}_t$ are the zero unconditional mean stable deviations, and $\mathbf{c} = \sum_{i=1}^{P} \mathbf{\Phi}_0 \mathbf{\Phi}_i \bar{\mathbf{z}}_{T+1-i}$. Moreover, let $\mathbf{r}_z$ be the $N_q \times 1$ vector of quantile movements we want to impose. They can be conveniently expressed in deviations from the constant part of equation (C.1), ie $\mathbf{r}_z = \mathbf{R}_z \mathbf{c} + \bar{\mathbf{q}}$. $\mathbf{R}_z = [\mathbf{I}_{N_q} \quad \mathbf{0}_{N_q, N_y+1}]$ is a $N_q \times N$ selection matrix, which picks the first N_q elements of $\mathbf{c}$, whereas $\bar{\mathbf{q}}$ represents the various movements imposed. $\bar{\mathbf{q}}$ varies across experiments, and is always depicted in the first Panel of each Figure. For the mean shock, $\bar{\mathbf{q}}$ is simply a vector of ones, implying a rightward movement of all the quantiles, that doesn't affect the shape of the distribution. The two restrictions can be written as:

$$\mathbf{R}_z \bar{\mathbf{z}}_{T+1} = \mathbf{r}_z, \quad \mathbf{R}_u \mathbf{u}_{T+1} = \mathbf{0}_{N-N_e}$$

$\mathbf{R}_u = [\mathbf{0}_{N_y, N_e} \quad \mathbf{I}_{N_y}]$ selects the last N_y structural shocks. Although the second restriction is on the structural shock, there is one to one map with $\bar{\mathbf{z}}_{T+1}$, ie $\mathbf{R}_u \mathbf{\Phi}_0 \bar{\mathbf{z}}_{T+1} = \mathbf{R}_u \mathbf{c}$. This allow us to rewrite both compactly as

$$\mathbf{R} \bar{\mathbf{z}}_{T+1} = \mathbf{r}$$

where $\mathbf{R} = [\mathbf{R}_z^{'}, \quad \mathbf{\Phi}_0^{'} \mathbf{R}_u^{'}]^{'}$, and $\mathbf{r} = [\mathbf{r}_z^{'}, \quad \mathbf{c}^{'} \mathbf{R}_u^{'}]$. The resulting structural shock can be obtained as

$$\boldsymbol{\mu}_u = (\mathbf{R} \mathbf{\Phi}_0^{-1})^{+} (\mathbf{r} - \mathbf{R} \mathbf{\Phi}_0^{-1} \mathbf{c})$$

where the $\mathbf{X}^{+}$ is the Moore Pensore inverse of $\mathbf{X}$.

Graphical visualization of scenario construction

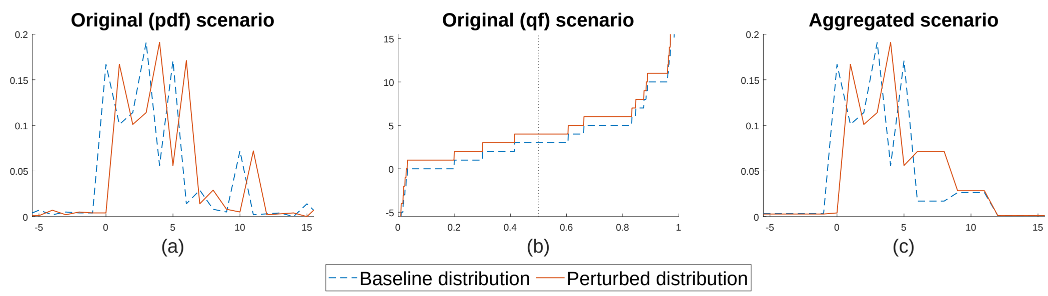


Figure C.1: **Example of Density Impulse construction: location shock.** In this figure we show the workflow to construct the experiment of the positive location shocks under the Discrete distributional approximation. Panel (a) shows the pdf function; Panel (b) shows the quantile function, Panel (c) shows the aggregated proportions (or probabilities). Each panel shows two curves: the original one (dashed-blue line) and the perturbed one (solid-orange line).

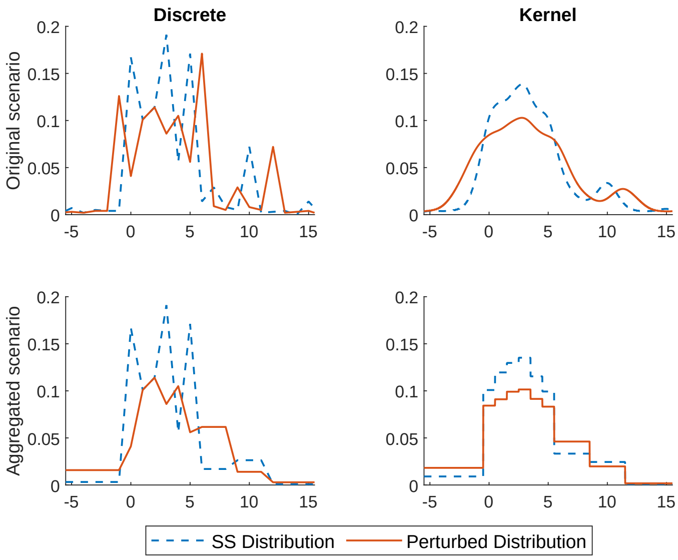


Figure C.2: **Zoom on the scenario construction: standard deviation shock.** In this figure we show the workflow to construct the experiment of the positive standard deviation shock under the two distributional assumptions: discrete (first column), and continuous (second column).

D Additional figures and results

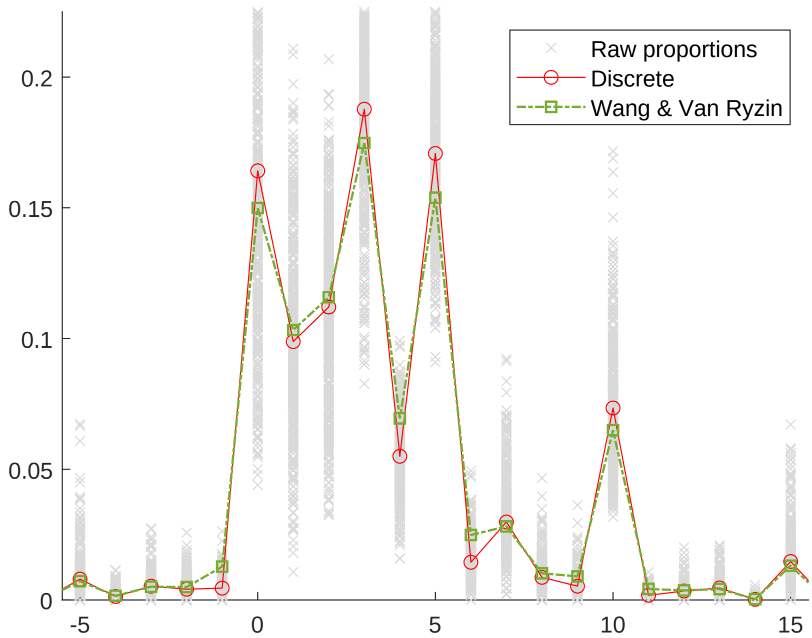


Figure D.1: The figure shows the average distributions over the temporal dimension of the "raw" empirical probabilities (red line), and the Kernel density (green dashed-dotted line) obtained using the Wang and Van Ryzin (1981) Kernel in conjunction with a cross-validation strategy for the bandwidth selection.

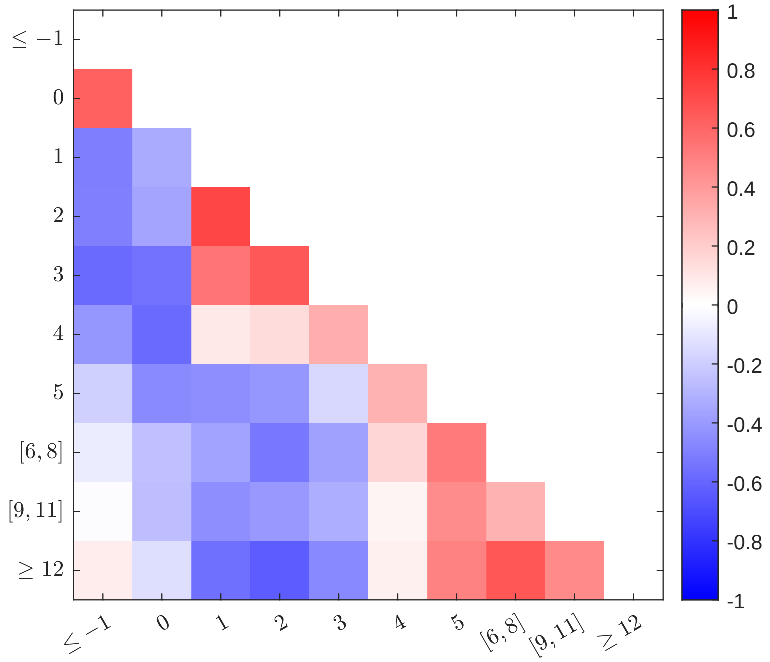


Figure D.2: The figure displays the lower triangular portion of the correlation matrix for the time series of aggregate probabilities associated with the discrete approximation $\mathbf{w}_t^D$.

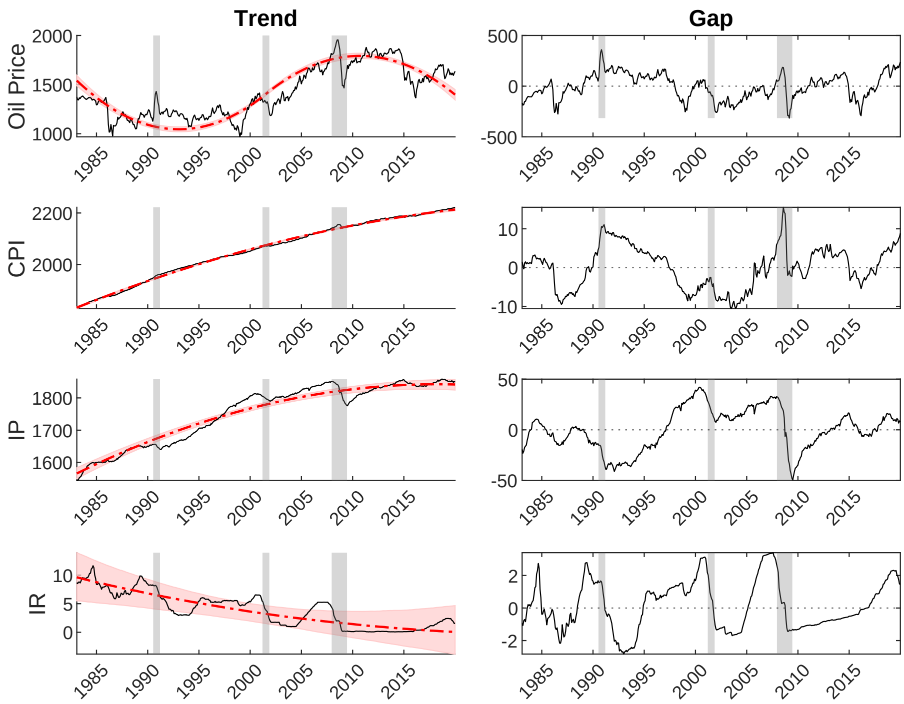


Figure D.3: Estimated trends and gaps for the three macroeconomic variables in the BVAR model. Red dashed-dotted line: median trend. Red shaded areas delimit the 68% credible bands.

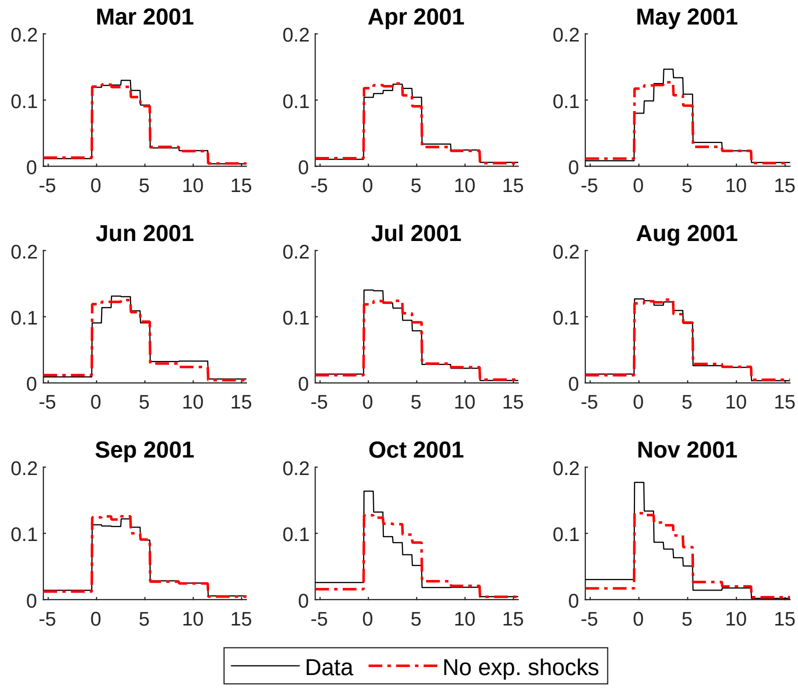


Figure D.4: Observed and counterfactual inflation expectation distribution in nine representative dates for the Dot-Com crisis.

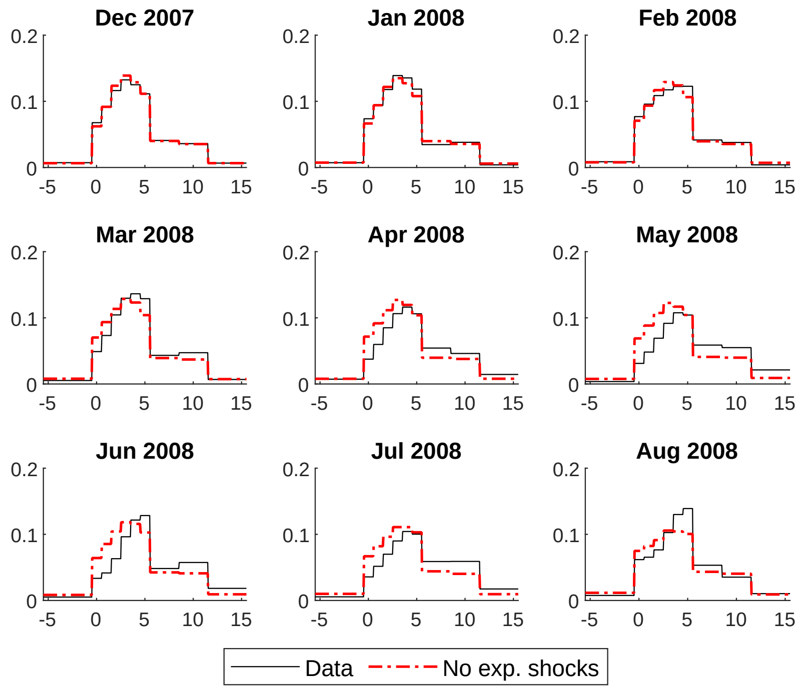


Figure D.5: Observed and counterfactual inflation expectation distribution in nine representative dates for the GFC.

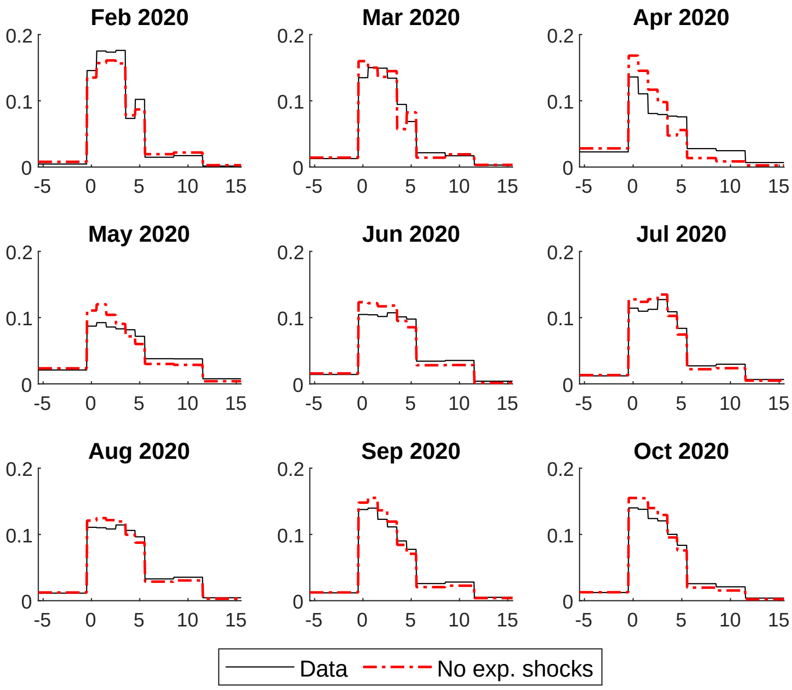
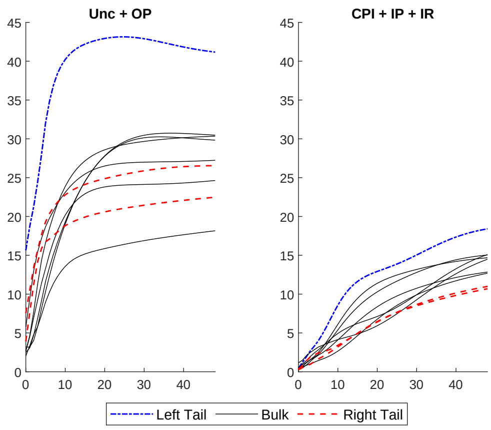


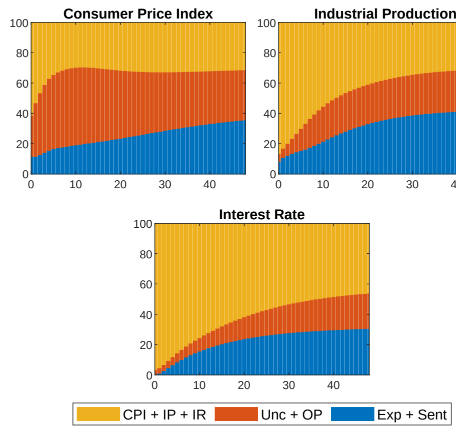
Figure D.6: Observed and counterfactual inflation expectation distribution in nine representative dates for the Covid-19 crisis.

D.1 Alternative identification schemes

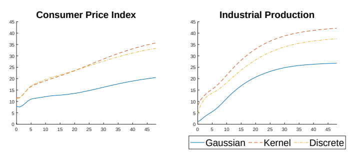
FEVDs



(a)



(b)



(c)

Figure D.7: **Different Identification: Interest rate exogenous to expectations, but endogenous to uncertainty indicators and the oil price.** Panel (a): Contribution of the two exogenous variables (first subplot) and the three macroeconomic variables (second subplot) to the FEVD of the nine expectations series at different horizons (x-axis). Different colors indicate the location in the domain of the proportions. Left tail (≤ 0, blue dashed-dotted line), bulk ($\in [1, 5]$, black lines), right tail (≥ 5, red dashed lines). Panel (b): Contributions of the three different identified blocks of indicators to the FEVDs of the macroeconomic variables at different horizons (x-axis). Panel (c): Comparison among the different approximations of the contribution of the expectation block to the FEVD of the macroeconomic variables.

FEVDs

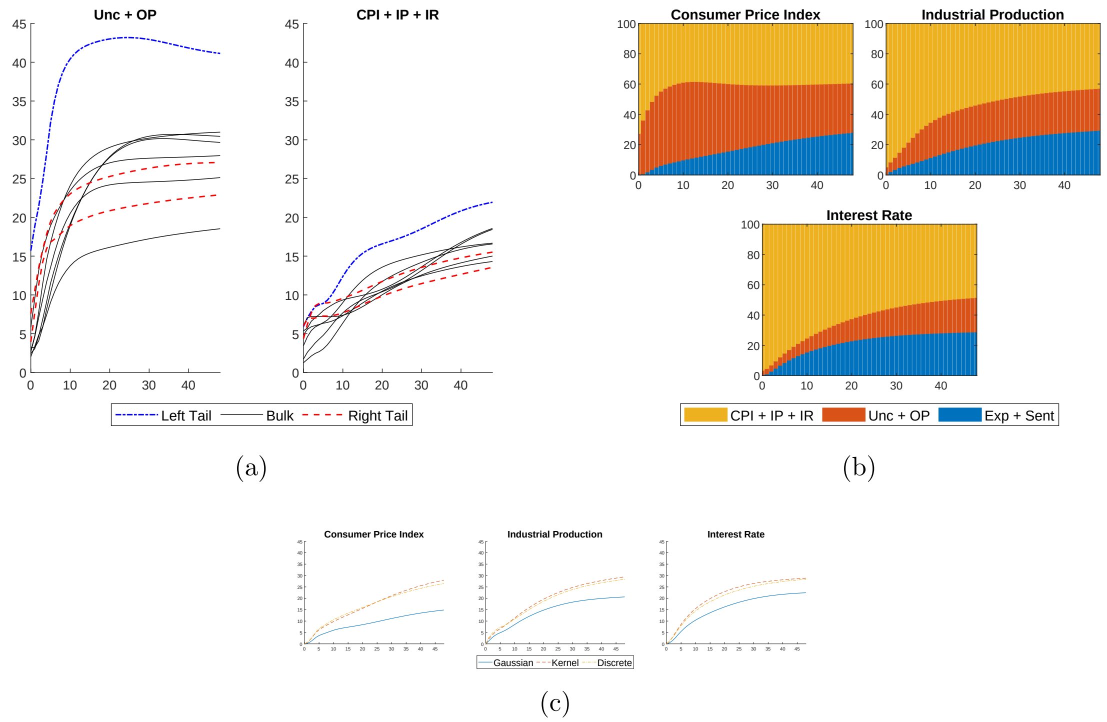


Figure D.8: **Different Identification: All the macroeconomic variables exogenous to expectations, but endogenous to uncertainty indicators and the oil price.** Panel (a): Contribution of the two exogenous variables (first subplot) and the three macroeconomic variables (second subplot) to the FEVD of the nine expectations series at different horizons (x-axis). Different colors indicate the location in the domain of the proportions. Left tail (≤ 0, blue dashed-dotted line), bulk ($\in [1, 5]$, black lines), right tail (≥ 5, red dashed lines). Panel (b): Contributions of the three different identified blocks of indicators to the FEVDs of the macroeconomic variables at different horizons (x-axis). Panel (c): Comparison among the different approximations of the contribution of the expectation block to the FEVD of the macroeconomic variables.

The following Table presents the full specification of the regression results in Table 3.

$$\mathrm{P}\big[\mathrm{E}_{i,t}(\pi_{t+12}) \in A\big] = \mu_t + \mathbf{x}_{i,t}^{'}\boldsymbol{\beta} + \epsilon_{i,t}$$

Variable	**Specification** $A = [-50, 1]$	$A = [0, 1]$
Birth cohort	-0.0268^{***}	-0.0218^{***}
	(0.0019)	(0.0019)
Squared effect	0.0007^{***}	0.0005^{***}
	(0.0001)	(0.0000)
If male	0.0111^{***}	0.0060^{**}
	(0.0027)	(0.0026)
If south	-0.0032	-0.0012
	(0.0039)	(0.0037)
If west	-0.0058	-0.0053
	(0.0042)	(0.0041)
If midwest	-0.0054	0.0010
	(0.0040)	(0.0039)
Hold diploma	-0.0099	-0.0099
	(0.0065)	(0.0063)
Hold degree	0.0048	0.0046
	(0.0029)	(0.0028)
Income (quintiles)	0.0100^{***}	0.0094^{***}
	(0.0012)	(0.0011)
If invest	-0.0054^{*}	-0.0024
	(0.0032)	(0.0031)
Sentiment	0.0664^{***}	0.0633^{***}
	(0.0012)	(0.0011)
Model summary		
Time FE	Yes	Yes
Robust SE	Yes	Yes
N	114847	114847
R-squared	0.0323	0.0302

Table D.1: Pooled OLS regression on inflation expectations, full specification.

E Michigan survey on Consumer attitudes

- *Mnemonic:* MSC.
- *Population*: Cross section of the general public.
- *Organization*: Survey research center, University of Michigan
- *N. of respondents*: 566 (mean), 480 (min), 1459 (max).
- *Type*: Short rotating Panel. For each month, an independent cross-section sample of households is drawn. The respondents chosen in this drawing are then reinterviewed six months later. The total sample for any one survey is normally made up of two- thirds of new respondents, and one-third being interviewed for the second time.
- *Timing*: variable, usually the fourth week of the month.
- *Forecast Horizon*: One year ahead (from January 1978), one and five year ahead (from April 1990).
- *Questions (for inflation expectations):*
  1. During the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now?
  2. By about what percent do you expect prices to go (up/down), on average, during the next 12 months?
- *Questions (for the Consumer Sentiment construction):*
  1. We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?
  2. Now looking ahead–do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?
  3. Now turning to business conditions in the country as a whole–do you think that during the next twelve months we'll have good times financially, or bad times, or what?
  4. Looking ahead, which would you say is more likely–that in the country as a whole we'll have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression, or what?
  5. About the big things people buy for their homes–such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items?