From Beliefs to Prices: Analyzing How Inflation Expectations Affect the Inflation Distribution*

Jose Barrales-Ruiz[†1], Azharul Islam[2], Mikidadu Mohammed[3], and Irina Panovska[4]

[1]Center of Economics for Sustainable Development (CEDES), Faculty of Economics, Business and Government, Universidad San Sebastian, Concepción, Chile. Email: jose.barrales@uss.cl
[2]School of Business Administration, Penn State Harrisburg, Middletown, PA, USA. Email: axi5208@psu.edu
[3]Completed paper while Visiting Scholar at The University of Texas at Dallas, Richardson, Texas, USA. Email: mikidadu@gmail.com
[4]School of Economic, Political, and Policy Sciences, The University of Texas at Dallas, Richardson, Texas, USA. Email: irina.panovska@utdallas.edu

August 11, 2025

Abstract

This paper investigates how the distribution of inflation expectations influences realized inflation across its entire distribution. We find that shocks to median inflation expectations increase the median inflation rate for up to three years, and subsequently generate persistent upside risks in inflation that endure for more than six years. By analyzing higher-order moments; specifically, the standard deviation and skewness of inflation expectations, we show that greater disagreement among agents has distinct impacts on the distribution of the inflation rate. While an increase in the standard deviation amplifies right-tail inflation risks and generates quantile-dependent effects at the 90th percentile of the inflation distribution, a negative skewness shock temporarily shifts the distribution without any quantile-dependent effects. Our results highlight the critical importance of the distributional properties of inflation expectations for understanding and managing inflation risks. The results also underscore the need for well-anchored expectations to promote macroeconomic price stability.

JEL classification: C31, E31

Keywords: Quantile Regression, Inflation expectations, Inflation, Expectations disagreement

*We would like to thank the seminar and conference participants at Universidad de Talca and the 2024 Southern Economic Association (SEA) Conference for their helpful comments and discussions. We would also like to thank Luiggi Donayre, Camilo Granados, and Victor Valcarcel for their helpful comments and suggestions. This paper was a chapter of Azharul Islam's PhD dissertation.

†This research was partially funded by the National Fund for Scientific and Technological Development (FONDECYT), National Research Agency of Chile, grant number 11250541.

1 Introduction

The global inflation surge of 2021–2023, the most synchronized and persistent uptick in price pressures since the 1970s, has reignited the debate about the mechanisms linking inflation expectations to realized inflation (Aastveit et al., 2024; Acharya et al., 2023; Adams & Barrett, 2024; Beaudry et al., 2024; Reis, 2020, 2023a, 2023b; Werning, 2022). Recent work by Coibion and Gorodnichenko (2025) highlights elevated and unanchored inflation expectations and the dangers that this may pose for monetary policy. While conventional wisdom holds that inflation expectations significantly influence inflation outcomes, the precise nature of this relationship, particularly how the entire distribution of expectations rather than just mean expectations affects inflation dynamics, remains incompletely understood. This paper contributes to this critical area of research by examining how both the central tendency and higher-order moments of inflation expectations distributions shape realized inflation across different quantiles. Our analysis employs quantile methods to capture potentially asymmetric effects and tail risks that standard mean-based approaches might overlook. This methodological approach is particularly relevant given the recent volatility in inflation observed globally, and the increasing importance of understanding the risks of the inflation tail for effective policy implementation. By documenting how heterogeneity in inflation expectations translates into asymmetric inflation outcomes, our findings offer valuable insights for monetary policymakers seeking to maintain price stability in an increasingly complex economic environment.

Recent empirical research has established that inflation expectations play a crucial role in driving inflation dynamics, with particularly compelling evidence emerging from the post-pandemic period. Using a structural vector autoregression model, Aastveit et al. (2024) examine the drivers of the recent inflation surge and highlight how inflation expectations significantly exacerbated inflation by amplifying the effects of global demand and supply shocks in the US, Canada, and New Zealand. Similarly, Beaudry et al. (2024) document the dominant role of inflation expectations in recent U.S. inflation dynamics through a Phillips curve framework. Reis (2023a) explores multiple hypotheses for the pandemic-era inflation surge, including the potential disregard of inflation expectations data by policymakers. Related research also identifies the discrepancy between market prices and long-term inflation expectations from surveys as a key driver of business cycle fluctuations (Reis, 2020).

Examining the pre-COVID literature reveals that while both the New Keynesian Phillips Curve (NKPC) and the hybrid NKPC frameworks theorize that inflation is a function of inflation expectations and economic slack, the empirical evidence quantifying the importance of expectations remains mixed. Hybrid NKPC models such as Gali et al. (2005) and Gali and Gertler (1999) found that the coefficient on inflation expectations significantly exceeds that on lagged inflation, a result echoed by Basistha and Nelson (2007) using forward-looking NKPC specifications. Contrasting perspectives emerge from Rudd (2022), who questions the fundamental relevance of inflation expectations in driving inflation dynamics, and Adams and Barrett (2024), who demonstrate that positive shocks to inflation expectations, defined as the gap between agents' expectations and the mathematical conditional expectation of inflation, can paradoxically lead to decreased inflation.

Recent evidence increasingly points to time variation and nonlinearity in the relationship between inflation expectations and realized inflation (Chang et al., 2022; Panovska & Ramamurthy, 2022). Lopez-Salido and Loria (2024) have documented that while inflation expectations symmetrically influence inflation across its distribution, other factors such as financial conditions, labor market slack, and oil price inflation contribute to asymmetry in the tails. Furthermore, an expanding body of research demonstrates that inflation outcomes are shaped not only by central tendency measures of household inflation expectations but also by the disagreement within these expectations—effectively, the higher-order moments of the household inflation expectations distribution. Research by Ahn and Farmer (2024) and Fofana et al. (2024) reveals that such disagreement is formed through prior beliefs, information processing, and demographic factors, and responds dynamically to macroeconomic shocks.

Motivated by this compelling evidence of nonlinearity and the growing literature examining inflation risks across both the left and right tails of the distribution, we adopt a quantile vector autoregression model (QVAR) to comprehensively analyze how inflation expectations and expectations disagreement transmit to realized inflation across different quantiles. This methodological approach allows us to capture potential asymmetries and nonlinearities that might be obscured in traditional mean-based estimation techniques, thereby providing a more nuanced understanding of the inflation expectations transmission mechanism.

Overall, our results provide evidence that the distribution of inflation expectations has important consequences for the distribution of the actual inflation rate. Specifically, a shock to the median inflation expectations increases the median inflation rate for up to three years following the shock. However, when examining the whole distributional response of the inflation rate after the shock, we observe a significant and persistent upside risk that remains for more than six years.

To assess the significance of these asymmetries, we conducted bootstrap analyses (where the evidence supported rejecting the null hypothesis of quantile independence), long-run analyses, and a series of robustness checks using both pre- and post-COVID data, as well as quarterly and monthly data. We further investigate how higher-order moments of inflation expectations affect the distribution of the inflation rate, utilizing the residualized standard deviation and skewness of inflation expectations as in Fofana et al. (2024). Our findings indicate that an increase in the standard deviation generates only a right-tail asymmetric risk in the actual inflation rate. In contrast, a negative shock to skewness—meaning fewer values in the left tail and more data clustered toward higher values (shifting the median leftward)—shifts the location of the inflation rate distribution for about one year after the shock, subsequently mildly increasing the right tail of the distribution.

In summary, our results provide compelling evidence that the distribution of inflation expectations has significant effects on the upside risk of inflation. This underscores the importance of well-anchored inflation expectations for maintaining the stability of future inflation rates.

The rest of this paper is structured as follows. Section 2 presents the methodology and data. Section 3 reports the findings. Section 4 provides further robustness checks. Section 5 provides some concluding remarks.

2 Methodology and Data

In this section, we present the econometric methods employed to analyze how shocks to inflation expectations—and its disagreement—affect the distribution of realized inflation rates. We also provide a detailed description of the data used in our analysis.

2.1 Methodology

Quantile vector autoregressive models (QVAR) are built upon quantile regressions, a concept which was introduced by Koenker and Bassett Jr (1978).[1] Although quantile regression has a long tradition in economics, its application to univariate time series is somewhat more recent with seminal contributions from Engle and Manganelli (2004) and Koenker and Xiao (2006). In the multivariate context, White et al. (2015) first introduced a reduced form estimation technique for QVAR models. Multivariate reduced form models however suffer from identification problems which seem to be more serious than in mean-based OLS estimation given that the law of iterated expectation does not hold in conditional quantiles.

The fundamental challenge in developing quantile vector autoregressive models lies in extending the concept of quantiles to higher dimensions, where no natural ordering exists (Wei, 2008). Several independent approaches have emerged to define and estimate multivariate quantile impulse responses. In the context of structural vector autoregressive models (SVARs), to our knowledge, two main methodological streams have developed. The first, represented by Montes-Rojas (2017, 2019, 2022), employs the multivariate directional quantile framework developed by Hallin et al. (2010), where quantiles are well-defined once a specific direction is chosen. The second approach, introduced by Chavleishvili and Manganelli (2024), implements a recursive identification strategy that decomposes any distribution into a product of marginal and univariate conditional distributions, building upon Wei (2008). This paper adopts the latter approach. In the recursive approach of Chavleishvili and Manganelli (2024), each quantile is estimated conditional on both the lagged variables and the contemporaneous variables from previously estimated equations. More formally, consider a vector of observed random variables $Y_t = (y_{1,t}, \dots, y_{K,t})'$ and a set of independent and identically distributed (iid) standard uniform random variables $U_t = \{u_{i,t}\}$ for $i = 1, \dots, K$. The random parameters SVAR(p) model can be written as:

$$Y_t = \omega(U_t) + A_0(U_t)Y_t + \sum_{j=1}^{p} A_j(U_t)Y_{t-j} \tag{1}$$

where $\omega(U_t)$ is a $K \times 1$ vector and $A_j(U_t)$ is a $K \times K$ matrix for $j = 1, \dots, K$ of parameters. The matrix $A_0(U_t)$ is a $K \times K$ lower triangular matrix representing the assumed recursive

[1] For more in-depth presentation on the quantile regression methodology see Koenker (2005, 2017)

identification. When Y_t is monotonically increasing in U_t, the τ_i–quantile function can be defined based on Equation (1) as:

$$Qy_{i,t}(\tau_i|Z_{i,t}) = Z'_{i,t}\beta_i(\tau_i) \tag{2}$$

where $Z_{i,t}$ is the conformable information set and $\beta_i(\tau_i)$ is a matrix containing all the parameters in model (1) (Koenker & Xiao, 2006). Estimation is carried out by minimizing the check function equation by equation following Koenker and Park (1996).

The structural impulse responses are obtained using the approach introduced by Chavleishvili and Manganelli (2024). Using the random coefficient representation of the quantile regression model as in (1), we can define the reduced form model as:

$$Y_t = v + \sum_{j=1}^{p} B_j(U_t)Y_{t-j} + \epsilon(U_t) \tag{3}$$

where the reduced form error is given by $\epsilon(U_t) = v(U_t) - v$, $v = E[v(U_t)]$, $v(U_t) = [I_K - A_0(U_t)]^{-1}\omega(U_t)$, and $B_j(U_t) = [I_K - A_0(U_t)]^{-1}A_j(U_t)$ for $j = 1, \dots, p$. The structural shocks can then be defined as:

$$C\varepsilon(U_t) = \epsilon(U_t), \qquad \varepsilon(U_t) \sim (0, I_K) \tag{4}$$

where matrix C must be estimated through simulation by assuming a short-run Cholesky identification, i.e.,

$$\Sigma_\epsilon = E[\epsilon(U_t)\epsilon(U_t)'] = C'C \tag{5}$$

The impulse response function can then be calculated by assuming that a shock of size $\delta = 1$ is changing the location of the whole distribution as:

$$\varepsilon^*(U_t) = \varepsilon(U_t) + \delta\iota \tag{6}$$

where ι is $K \times 1$ vector of zeros with a 1 in the position corresponding to the desired shocked variable. Therefore, the quantile impulse response function (QIRF) for $h = 0$ is given by:

$$\begin{aligned} \text{QIRF}_0(U_0, \delta) &= Y_0^* - Y_0 \\ &= C\delta\iota \end{aligned} \tag{7}$$

where Y_0^* represents the shocked variable. For the following periods, $h = 1, \ldots, H-1$, the QIRF can be obtained by iterating over Equation (3).

Based on these estimates, we compute the distributional QIRF (DQIRF) and the long-term impact coefficient to analyze how the distribution of inflation expectations, specifically its median, standard deviation, and skewness, affect the realized inflation rate. The DQIRF captures the dynamic response of the entire inflation rate distribution following a shock to inflation expectations. To operationalize this, we employ Algorithm 1 outlined in Montes-Rojas (2022) for estimating distributional responses.[2]

Algorithm 1

Consider the QIRF for $h = 0, \ldots, H-1$ by using B simulations, given the shock of size δ.

- For a finite grid of quantiles $0 < \tau_1 < \tau_2 < \cdots < \tau_N$. Construct and store the parameters matrices in Equation (1).
- Repeat the following steps for $b = 1, \ldots, B$:
  1. Randomly draw $\tau_0 = (\tau_1, \ldots, \tau_K)'$ from the sequence stored above. Compute the $\text{QIRF}_0(\tau_0, \delta)$. Define $u^{(0)} = \tau_0$.
  2. For $h = 1, \ldots, H-1$ randomly draw τ_h from the sequence stored above. Compute the $\text{QIRF}_h(\tau_h, \delta)$. Define $u^{(h)} = [\tau_{h-1}, \tau_h]$.
  3. Save $\text{QIRF}(b) = [\text{QIRF}_0, \ldots, \text{QIRF}_{H-1}]$.
- The collection $[\text{QIRF}(1), \ldots, \text{QIRF}(B)]$ are used to evaluate the distribution of the IRF at any horizon $h = 0, \ldots, H-1$.

The long-run impact coefficients for the inflation rate equation help us to determine whether the moments of inflation expectations influence tail risks in the actual inflation rate in the long run. To conduct inference on these coefficients, we employ the quantile bootstrap algorithm described in Chavleishvili and Manganelli (2024), referred to as Algorithm 2.

2.2 Data

The empirical analysis employs quarterly U.S. macroeconomic data spanning from 1960Q1 to 2024Q2, incorporating 2 lags in the estimation. The baseline model consists of five key

[2] We also carry surrogate methods to test our DQIRFs results against the null of quantile independence (described in detail in Section 3.1).

Algorithm 2

1. For a finite grid of quantiles $0 < \tau_1 < \tau_2 < \cdots < \tau_N < 1$. Construct and store the parameters matrices in Equation (1). Construct estimates of the conditional quantiles function $\hat{Q}_{y_{i,t}}(\tau_j|Z_{i,t})$ for each equation $i = 1, \ldots, K$ and quantile $j = 1, \ldots, N$.
2. For each $p < t \leq T$ and equation $i = 1, \ldots, K$, find τ_j from the grid that minimizes the distance $|y_{i,t} - \hat{Q}_{y_{i,t}}(\tau_j|Z_{i,t})|$. Collect and store these values as $\{\tau_t\}_{t=p+1}^{T}$.
3. For $b = 1, \ldots, B$: Draw independently $\left\{u_t^{(b)}\right\}_{t=p+1}^{T}$ from the sequence stored quantiles in $\{\tau_t\}_{t=p+1}^{T}$ and generate the bootstrap sample recursively following Equation (1).
4. The collection of bootstrap samples are used to perform inference.

variables: price index for energy goods and services (DNRGRG3M086SBEA), the output gap which is measured as the percentage deviation of real GDP (GDPC1) relative to real potential GDP from the Congressional Budget Office (GDPPOT), one-year ahead inflation expectations from the University of Michigan Survey of Consumers, the Personal Consumption Price Index (PCEPI) inflation rate, and the 3-month Treasury Bill (DTB3), downloaded from the Federal Reserve Economic Data Base (FRED).[3] Figure 1 plots the time series used in the baseline model. As a robustness check, in Section 4 we also present the results from QVAR analyses using quarterly data for the pre-COVID sample, and the results when we use monthly data for both the full and the pre-COVID samples.

The identification strategy relies on recursive ordering restrictions, with variables ordered as described above. First, following Kilian and Zhou (2022, 2023), we treat the price index for energy goods and services as the most contemporaneously exogenous variable in the system. Second, in line with Beaudry et al. (2024), we assume that the output gap affects both inflation expectations and realized inflation contemporaneously. Third, although the contemporaneous relationship between inflation expectations and the inflation rate remains debated, we adopt the view in Beaudry et al. (2024), which suggests that inflation expectations adjust more slowly to inflation surprises within a quarter. This also parallels the findings of Kocherlakota (2016) who finds that expectations respond sluggishly.[4] Accordingly, inflation expectations are ordered before the actual inflation rate. Finally, the interest rate is placed last in the ordering, reflecting the assumption that the Federal Reserve responds contemporaneously to

[3]Energy price inflation, output gap, and the inflation rate are measured as annualized percentage change.

[4]However, as a robustness check, we also reversed the ordering by placing the inflation rate before inflation expectations. The results indicate that inflation exhibits a weaker response to shocks in inflation expectations under this alternative ordering. Nevertheless, the key qualitative finding persists: the impulse response continues to display pronounced excess skewness relative to a standard SVAR framework, thereby underscoring the robustness of the main results. The results with this alternative ordering are available on request.

current economic conditions but does not influence the other variables within the same period.

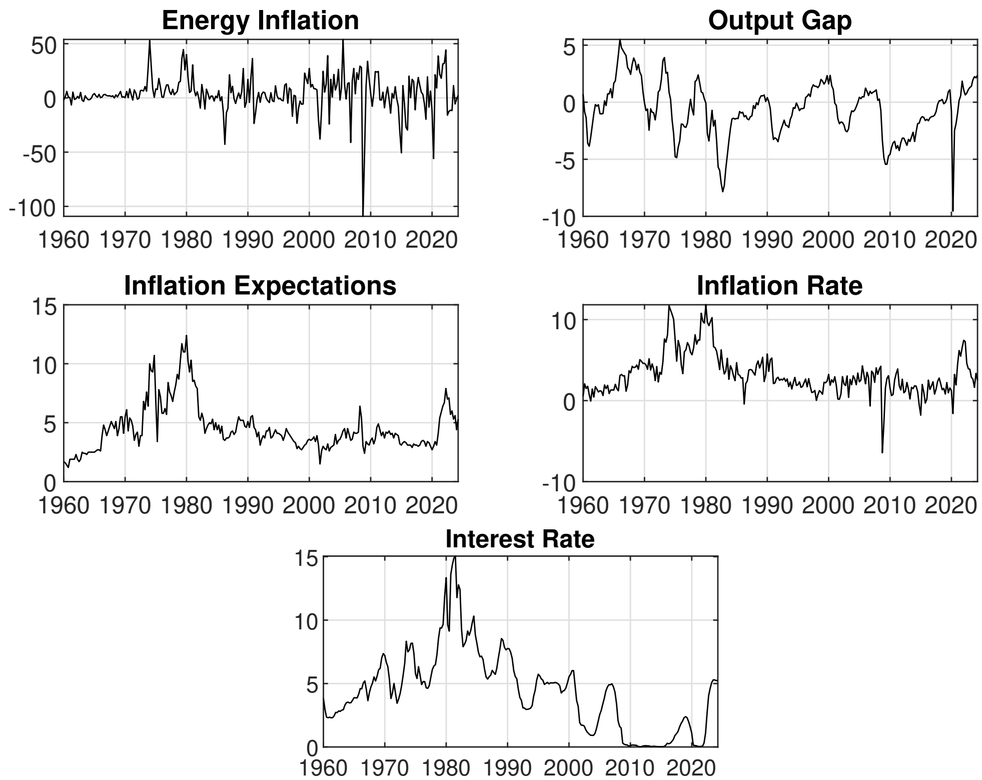


Figure 1: **Quarterly time series dataset** This figure shows the quarterly time series for the energy prices, output gap, inflation expectations, inflation, and interest rate from 1960Q1 to 2024Q2.

3 Results

In this section, we present the main results of the study. First, we provide evidence that an increase in median inflation expectations has an asymmetric effect on the distribution of the realized inflation rate. Second, we use the residualized median, standard deviation, and skewness of inflation expectations to: (1) verify that our findings are robust to shocks in the residualized median, and (2) demonstrate that different measures of disagreement in inflation expectations and the different moments of the expectation distribution have distinct implications for the realized inflation rate.[5]

[5]Following Fofana et al. (2024), we construct residualized measures of inflation expectations by using the University of Michigan inflation expectations survey and removing the influence of observable characteristics such as demographic, social, and economic factors from raw inflation expectations. Further details are provided in Subsection 3.2.

3.1 Distributional effects of a median inflation expectation shock

We first examine the impact of a one standard deviation shock to median inflation expectations on the inflation rate. The main findings, illustrated in Figure 2 and derived using Algorithm 1,[6] indicate that a shock to median expected inflation leads to a significant increase in mean inflation for up to three years before gradually returning to baseline levels. However, the distributional impulse response analysis highlights a notable and persistent upside risk in inflation's upper quantiles, which remains relevant for more than six years after the shock. These results are consistent with the finding in Lopez-Salido and Loria (2024) who find that both in the U.S. and the Euro area, inflation expectations have been the decisive inflation determinant over the last 20 years. These results underscore the inflationary risks associated with unanchored median inflation expectations and emphasize the importance of maintaining well-anchored expectations for effective monetary policy.

To formally evaluate whether the skewness of the estimated DQIRFs in Figure 2 is significant, we carry out a surrogate exercise. We first estimate an SVAR based on the data using the identification assumptions introduced in the previous section. Then, we simulate $M = 1,000$ linea SVARs using the estimated coefficients from a linear model, each of size $T = 400$, discarding the first 100 observations to approximately match our sample size. Then we estimate a QVAR model for each of the M realizations from the simulated linear SVARs (that by construction do not have any quantile dependence). Finally, we calculate the DQIRF from these models and calculate the skewness for each horizon to set the null hypothesis. The bootstrap comparison in Table 1 reveals that the observed skewness values exceed the surrogate distribution for horizons beyond 12 quarters, confirming the asymmetry of the responses is statistically significant. As expected, the DQIRF from our data set show strong positive skewness over longer horizons.

To further investigate the effect of median inflation expectations on the inflation rate, we use two approaches: (i) estimating the long-run coefficients for the inflation rate equation (see Equation (10) below), and (ii) calculating the cumulative (impulse) response of the inflation rate to a permanent 1 percentage point (pp) increase in inflation expectations at a 24-quarter

[6] Kernel density estimates were computed in MATLAB using the default normal (Gaussian) kernel and automatic bandwidth selection, after excluding the 0.1% lower and upper outliers.

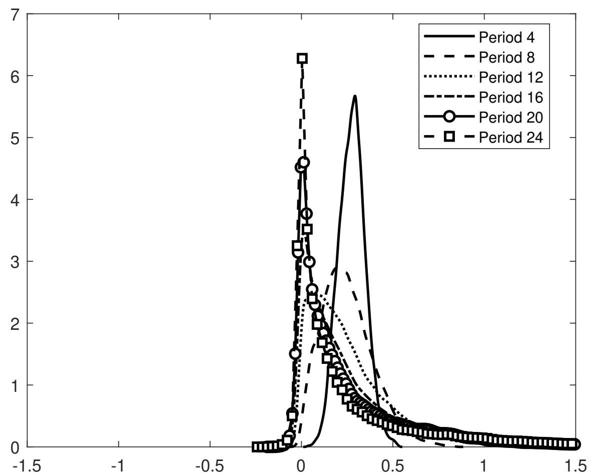


Figure 2: **Distributional QIRF** This figure shows the distributional QIRF to the inflation rate from a one standard deviation shock to the median inflation expectations. We carry out the calculations using Algorithm 1.

Horizon	Skewness from SVAR	Skewness from data
4	0.1179 (0.5143)	0.1374
8	0.2579 (0.5675)	0.7274
12	0.4240 (0.6744)	1.6470
16	0.5482 (0.8150)	2.8796
20	0.6130 (0.9985)	4.6782
24	0.6514 (1.2491)	7.3946

Table 1: **Surrogate skewness** This table shows the results from skewness from an SVAR simulation. The columns show the response horizon, mean and standard error (in parenthesis) from the SVAR surrogate, and skewness from data; respectively.

horizon. These two sets long-run parameters and responses help us understand whether the inflation rate has an asymmetric response to inflation expectations and, therefore, if inflation expectations can exert tail risk on the inflation rate. To calculate the long-run parameter for the inflation rate equation, we use the random parameter VAR(p) representation in Equation (1) and the algorithm for inference in Algorithm 2. With this approach, we can obtain the (possibly) quantile-dependent long-run reaction of the inflation rate distribution to a 1pp increase in the median annual inflation expectations. Under a linear VAR assumption, i.e., when the parameters are quantile-independent, the long-run parameter associated with inflation expectations should have an equal effect across the entire distribution.[7]

[7] Quantile-invariance in the long-run effect can be seen in Equation (1) when $\omega(U_t) = \omega$, $A_0(U_t) = A_0$, and $A_j(U_t) = A_j$ for $j = 1, \ldots, K$.

We calculated the long-run parameter from the inflation equation as follows. Equation (8) represents the SVAR model for the τ-quantile (excluding constants and error terms).

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ a_{21,\tau} & 1 & 0 & 0 & 0 \\ a_{31,\tau} & a_{32,\tau} & 1 & 0 & 0 \\ a_{41,\tau} & a_{42,\tau} & a_{43,\tau} & 1 & 0 \\ a_{51,\tau} & a_{52,\tau} & a_{53,\tau} & a_{54,\tau} & 1 \end{bmatrix}}_{A_0(\tau)} \underbrace{\begin{bmatrix} \Pi_t^{\text{ene}} \\ \widetilde{Y}_t \\ \Pi_t^{\text{exp}} \\ \Pi_t \\ R_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} b_{11,\tau}(L) & b_{12,\tau}(L) & b_{13,\tau}(L) & b_{14,\tau}(L) & b_{15,\tau}(L) \\ b_{21,\tau}(L) & b_{22,\tau}(L) & b_{23,\tau}(L) & b_{24,\tau}(L) & b_{25,\tau}(L) \\ b_{31,\tau}(L) & b_{32,\tau}(L) & b_{33,\tau}(L) & b_{34,\tau}(L) & b_{35,\tau}(L) \\ b_{41,\tau}(L) & b_{42,\tau}(L) & b_{43,\tau}(L) & b_{44,\tau}(L) & b_{45,\tau}(L) \\ b_{51,\tau}(L) & b_{52,\tau}(L) & b_{53,\tau}(L) & b_{54,\tau}(L) & b_{55,\tau}(L) \end{bmatrix}}_{B_\tau(L)} \underbrace{\begin{bmatrix} \Pi_t^{\text{ene}} \\ \widetilde{Y}_t \\ \Pi_t^{\text{exp}} \\ \Pi_t \\ R_t \end{bmatrix}}_{Y_t} \tag{8}$$

where Π_t^{ene}, $\widetilde{Y}_t$, Π_t^{exp}, Π_t, R_t represent the energy inflation, output gap, inflation expectations, realized inflation rate, and interest rate, respectively. Further, $A_0(\tau)$ represents the τ-quantile impact matrix of coefficients and $B_\tau(L)$ the τ-quantile matrix of coefficients with L as the lag operator. From the fourth row of Equation (8), it follows that

$$\begin{aligned} & a_{41,\tau}\Pi_t^{\text{ene}} + a_{42,\tau}\widetilde{Y}_t + a_{43,\tau}\Pi_t^{\text{exp}} + \Pi_t \\ & = b_{41,\tau}(L)\Pi_t^{\text{ene}} + b_{42,\tau}(L)\widetilde{Y}_t + b_{43,\tau}(L)\Pi_t^{\text{exp}} + b_{44,\tau}(L)\Pi_t + b_{45,\tau}(L)R_t \end{aligned} \tag{9}$$

Solving for the inflation rate Π_t and setting $L = 1$, we find that

$$\begin{aligned} [1 - b_{44,\tau}(1)]\Pi_t = -a_{41,\tau}\Pi_t^{\text{ene}} - a_{42,\tau}\widetilde{Y}_t - a_{43,\tau}\Pi_t^{\text{exp}} \\ + b_{41,\tau}(1)\Pi_t^{\text{ene}} + b_{42,\tau}(1)\widetilde{Y}_t + b_{43,\tau}(1)\Pi_t^{\text{exp}} + b_{45,\tau}(1)R_t \end{aligned} \tag{10}$$

The long-run effect from a 1pp increase in the inflation expectations Π_t^{exp} to the inflation rate Π_t can be calculated by computing the following coefficient $[-a_{43,\tau} + b_{43,\tau}(1)]/[1 - b_{44,\tau}(1)]$.

Table 2 presents the long-run responses for the quantiles $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$ across the columns. Each row displays the point estimates and their corresponding 68% confidence intervals. The last row reports the cumulative quantile impulse response.

The results indicate that as the quantile of interest, τ, increases, the long-term response also rises, and expectations have a larger effect and higher quantiles.[8] The last row of Table 2 also shows the cumulative IRF for different quantiles, which represents the response of inflation rate to a permanent 1pp increase in the annual inflation expectations. Overall, both sets of results in the table show that there are higher upside risks from inflation expectations to the distribution of the inflation rate.

	τ				
π	10%	25%	50%	75%	90%
16%	0.6070	0.5931	0.7045	0.7478	0.7545
50%	0.7254	0.7484	0.8555	0.9231	0.9236
84%	0.8360	0.8898	0.9924	1.0869	1.0873
Point Estimate	0.6854	0.7293	0.8752	0.9516	0.9543
Cumulative IRF	0.8242	1.4060	2.5398	4.4618	7.4357

Table 2: **Long-run inflation** Rows one to four shows the results for the long-run response from the inflation rate equation to a shock to the median annual inflation expectations in 1pp. Inference is provided using Algorithm 2. The last row shows the cumulative impulse response from a permanent 1pp shock to inflation expectations at quarter 24.

3.2 Shocks to higher-order moments of inflation expectations

The literature on inflation expectations has largely emphasized the importance of central tendencies, such as the median or mean values, in determining inflation rates. Our study extends this type of analysis by incorporating residualized higher-order moments of the distribution of inflation expectations, as proposed by Fofana et al. (2024). Their methodology focuses on removing observable characteristics such as demographic, social, and economic factors from inflation expectations to isolate residual disagreement. This residual disagreement is then captured through statistical measures such as the standard deviation and skewness, which reflect the dispersion and asymmetry in expectations across households. By focusing on these residualized higher-order moments, our results provide deeper insights into the dynamics of

[8] Additionally, we applied Algorithm 2 to simulated VAR models using the same data. As expected, the results showed that the long-run response is independent of the quantile.

inflation expectations disagreement that are not captured by conventional metrics.

Using data from the Michigan Survey of Consumers (MSC), we adopt the Fofana et al. (2024) residualization framework to replace conventional median inflation expectations with residualized higher-order moments (standard deviation and skewness) in our QVAR model. We compute the QIRFs to analyze how shocks propagate through the distribution of inflation expectations. Crucially, these residualized disagreement measures are time-varying, enabling us to explore how shocks to the higher moments of inflation expectations—rather than central tendencies—affect actual inflation dynamics.

Figure 3 plots the time series for the residualized median, standard deviation, and skewness derived from the survey data. To begin our analysis, Figure 4 examines the effects of a shock to the residual median inflation expectations. The results closely align with those presented in Figure 2, thereby serving as a robustness check for our earlier findings. As emphasized before, a shock to median inflation expectations leads to an upside risk for the inflation rate.

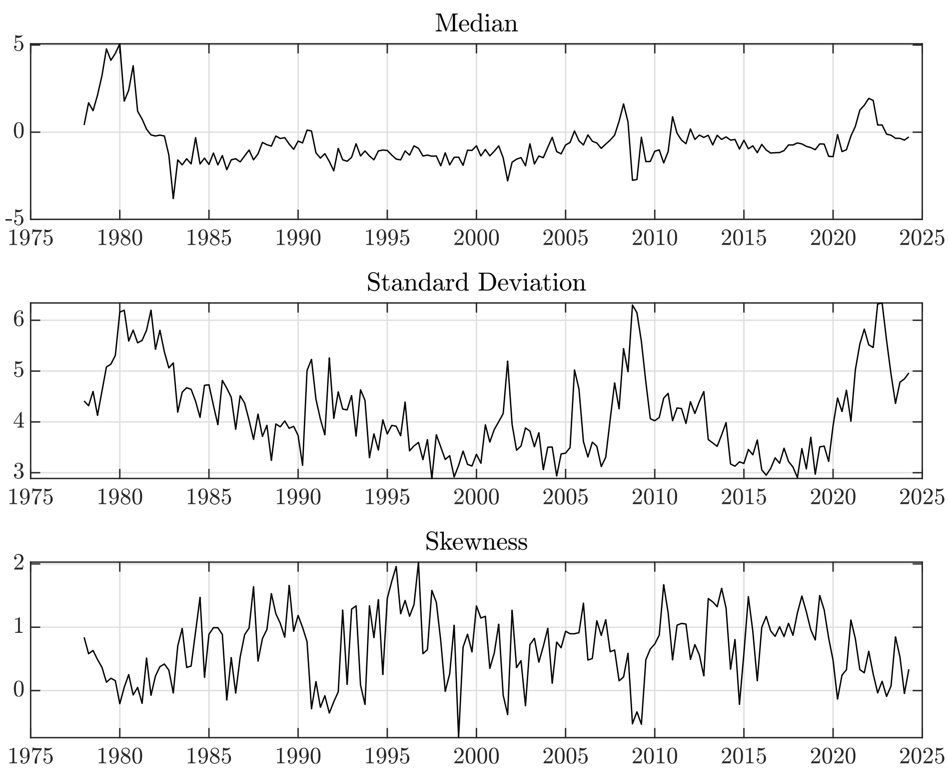


Figure 3: **Higher order moments time series** Quarterly time series of residualized higher order moments of the inflation expectations distribution, 1978Q1 to 2024Q2.

Second, a positive shock to the residual standard deviation, as shown in Figure 5a, increases only the risk of higher inflation rates on the QIRF distribution up to approximately one year after the shock occurs. The standard deviation (along with the interquartile range) is often used as a proxy for disagreement when it comes to inflation expectations. Interestingly, the response of inflation to a shock in the residual standard deviation resembles its response to a shock in the median inflation expectations after a year, although to a lesser degree.

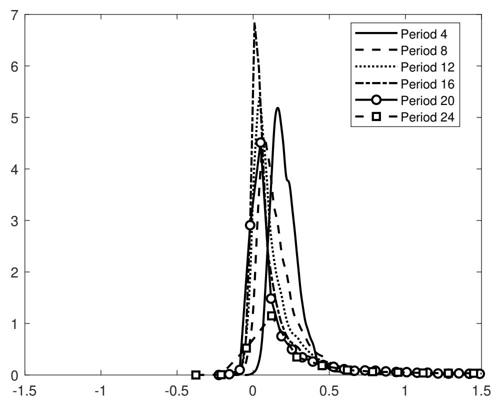
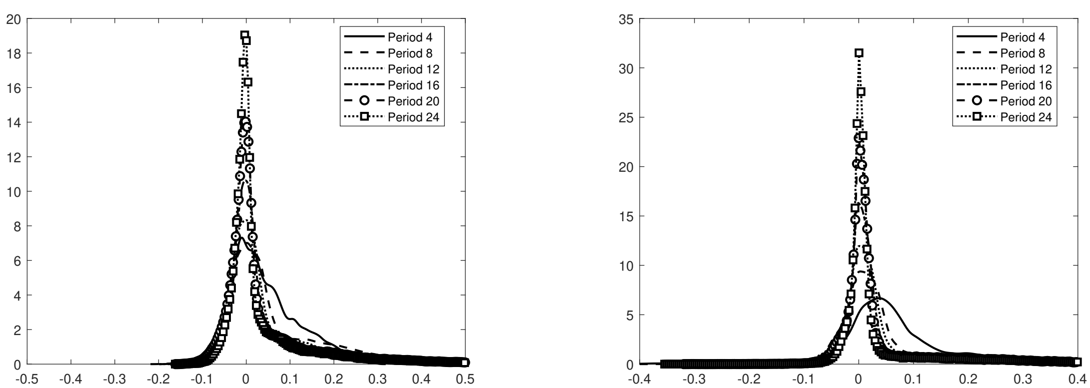


Figure 4: **Distributional QIRF for residualized median** This figure shows the distributional QIRF to the inflation rate from a one standard deviation shock to the residualized median inflation expectations. We carry out the calculations using Algorithm 1.



(a) Standard Deviation

(b) Skewness

Figure 5: **Distributional QIRF for higher order moments** This figure shows the distributional QIRF over inflation rate from a one standard deviation positive shock to residual standard deviation and a one standard deviation negative shock to the residual skewness of the inflation expectations distribution. We carry out the calculations using Algorithm 1.

Next, we explore a negative shock to skewness meaning that fewer values are at the left tail of the distribution leaving most data clustered toward higher values. Interestingly, in this case, the median would move to the left. Figure 5b shows our results, which indicate that a decrease in skewness affects the inflation rate positively after a year, increasing it by almost 0.5 percentage points. This goes in the opposite direction of the same effect of a decrease of the median in Figure 4. After a year, over the medium run, the effect on inflation converges to a distribution with a higher left tail, but with a non-negligible risk of an increase in inflation in the right tail.

	τ				
π	10%	25%	50%	75%	90%
16%	0.4528	0.5072	0.6695	0.7139	0.7288
50%	0.7602	0.7964	0.9193	0.9165	1.0154
84%	0.9998	1.0315	1.1138	1.1010	1.2717
Point Estimate	0.8205	0.9484	0.9202	1.1045	1.4472
Cumulative IRF	0.5533	0.8496	1.3968	2.6043	5.3865

Table 3: **Long-run inflation from residualized median inflation expectations** Rows one to four shows the results for the long-run response from the inflation rate equation to a shock to the residualized median annual inflation expectations in 1 percentage point (pp). Inference is provided using Algorithm 2. The last row shows the cumulative impulse response from a permanent 1pp shock to the residualized inflation expectations at quarter 24.

	τ				
π	10%	25%	50%	75%	90%
16%	-1.0517	-0.8381	-0.6736	-0.3240	0.6123
50%	-0.0311	0.1638	0.2969	0.4822	1.3236
84%	0.6707	0.8605	0.9828	1.1512	2.0659
Point Estimate	0.3937	-0.0566	0.0004	2.0006	1.6422
Cumulative IRF	-0.5551	-0.2046	0.2295	1.2439	3.1818

Table 4: **Long-run inflation from residualized standard deviation of inflation expectations** Rows one to four shows the results for the long-run response from the inflation rate equation to a shock to the residualized standard deviation in 1 unit. Inference is provided using Algorithm 2. The last row shows the cumulative impulse response from a permanent 1 unit shock to the residualized standard deviation in inflation expectations at quarter 24.

Our findings contribute to the literature on expectation disagreement and inflation dynamics (Mankiw et al., 2003) and demonstrate that different measures of disagreement play a critical role in understanding their effects on the distribution of actual future inflation rates. The QIRF in Figure 5 reveals distinct mechanisms: while a shock to the standard deviation, which can be used as a proxy for expectation disagreement, primarily induces rightward skewness in the distribution (similar to the median shock). A negative skewness shock exerts a transitory upward location shift on inflation up to one year. Beyond this horizon, the distributional impacts of skewness shocks become less significant. Interestingly, the long-run effects of these shocks diverge: while the impacts of the shocks to the standard deviation remain significant at extremely upper quantiles ($\tau = 0.9$), the effects of shocks to the skewness dissipate in the long run, i.e., they are quantile independent. Tables 3, 4, and 5 summarize these findings. For instance, the results for the residual median in Table 3 largely align with our earlier results. However, increases in the residual standard deviation, as shown in Table 4, exhibit significant effects only in the long run for the 90% quantile. In contrast, evidence from Table 5 suggests that residual skewness has no significant quantile dependent effect over the longer run.

	τ				
π	10%	25%	50%	75%	90%
16%	-0.8172	-0.9557	-1.1695	-0.9907	-2.4422
50%	0.0297	-0.1189	-0.0508	-0.0405	-1.0869
84%	1.1271	1.0007	1.5085	1.2308	0.2269
Point Estimate	-0.5271	0.3378	-0.0953	-0.8479	-3.4442
Cumulative IRF	-2.7938	-0.8025	-0.2997	0.0583	0.4237

Table 5: **Long-run inflation from residualized skewness of inflation expectations** Rows one to four shows the results for the long-run response from the inflation rate equation to a shock to the residualized skewness in 1 unit. Inference is provided using Algorithm 2. The last row shows the cumulative impulse response from a permanent 1 unit shock to the skewness of inflation expectations at quarter 24.

4 Robustness

We conduct additional robustness checks to assess the impact of a shock to median inflation expectations on the distribution of the actual inflation rate. First, to evaluate whether our results were affected by the recent run up in inflation, we reestimate our model by using pre-COVID quarterly data. Second, we employ comparable monthly data covering both the pre- and post-COVID periods. The overall results were qualitatively and quantitatively similar to our baseline results, and the corresponding figures and tables are presented below.

4.1 Quarterly results: Before COVID

Figure 6 presents the distributional impulse responses of the realized inflation to an one standard deviation shock in median inflation expectations, using Algorithm 1 and quarterly data from 1960Q1 to 2019Q1. The findings are consistent with the full-sample results, showing that a shock to median inflation expectations raises the realized inflation for up to four years and reveals upside risks in the upper quantiles of the inflation distribution. The long-run coefficients in Table 6 show similar findings, indicating that the long-run response of inflation to a shock in inflation expectations increases with higher quantiles. Furthermore, the magnitude and the shape of the responses presented in Figures 2 and 6 are very similar, as are the long-run responses presented in Table 2 and 6, indicating that our results are not driven by the inflation run up following 2020.

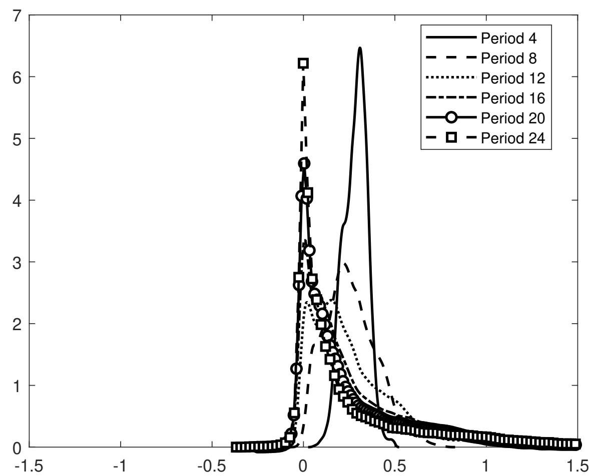


Figure 6: **Distributional QIRF** This figure shows the distributional QIRF from a median inflation expectations shock to the inflation rate using only pre-COVID data. We carry out the calculations using Algorithm 1.

	τ				
π	10%	25%	50%	75%	90%
16%	0.6700	0.6881	0.7645	0.7782	0.7568
50%	0.7734	0.8261	0.8988	0.9326	0.9388
84%	0.8843	0.9612	1.0270	1.0837	1.1295
Point Estimate	0.6819	0.8076	0.9108	0.9396	0.9860
Cumulative IRF	0.8257	1.4300	2.5820	4.6620	7.6528

Table 6: **Long-run inflation** Rows one to four shows the results for the long-run response from the inflation rate equation to a shock to the median annual inflation expectations in 1pp. Inference is provided using Algorithm 2. The last row shows the cumulative impulse response from a permanent 1pp shock to inflation expectations at quarter 24.

4.2 Monthly data: Full sample

We also conduct monthly robustness checks to analyze the effect of inflation expectations on inflation using data spanning from 1978M1 to 2024M6 for the full sample, with the pre-COVID subsample ending at 2019M12. The estimation incorporates six lags. Similar to the QVAR analysis with quarterly data, the monthly analysis includes five key variables: price index for energy goods and services (DNRGRG3M086SBEA), industrial production (INDPRO) as a proxy for economic activity, one-year-ahead inflation expectations (MICH) from the University of Michigan Survey of Consumers, the Personal Consumption Expenditures Price Index (PCEPI) inflation rate, and the 3-month Treasury Bill rate (DTB3).[9] Figure 7 presents the corresponding time series. The identification strategy follows the same recursive ordering restrictions used in the quarterly analysis, with variables ordered as listed above. Distributional quantile impulse responses (DQIRF) are in Figures 8 and 9. Also, the long-run results are reported in Tables 7 and 8. The results again largely agree with our baseline findings; with the exception of Table 8 where the maximum long-run response for the inflation equation, i.e., rows one to four, is attained at median quantile ($\tau = 0.5$). However, it is important to note that in this case there was substantial uncertainty associated with the responses at higher quantiles, and the 68% confidence interval at $\tau = 90\%$ was wide in this case.

[9] As with the quarterly data, energy price inflation, industrial production growth, and inflation rate are measured as annualized percentage changes. Because there is no standardized benchmark for the long-run potential level of industrial production, and different statistical assumptions about the trend behavior can produce different estimates for the cyclical behavior of a variable (see, inter alia, Hamilton, 2018 or Morley and Panovska, 2020. To avoid taking a stand about the nature of the trend of industrial production, we use the growth rate of industrial production to control for the state of the business cycle. As shown in the two subsequent subsections, the results obtained using this specification are quite similar to the quarterly results that use the CBO output gap in the QVAR.

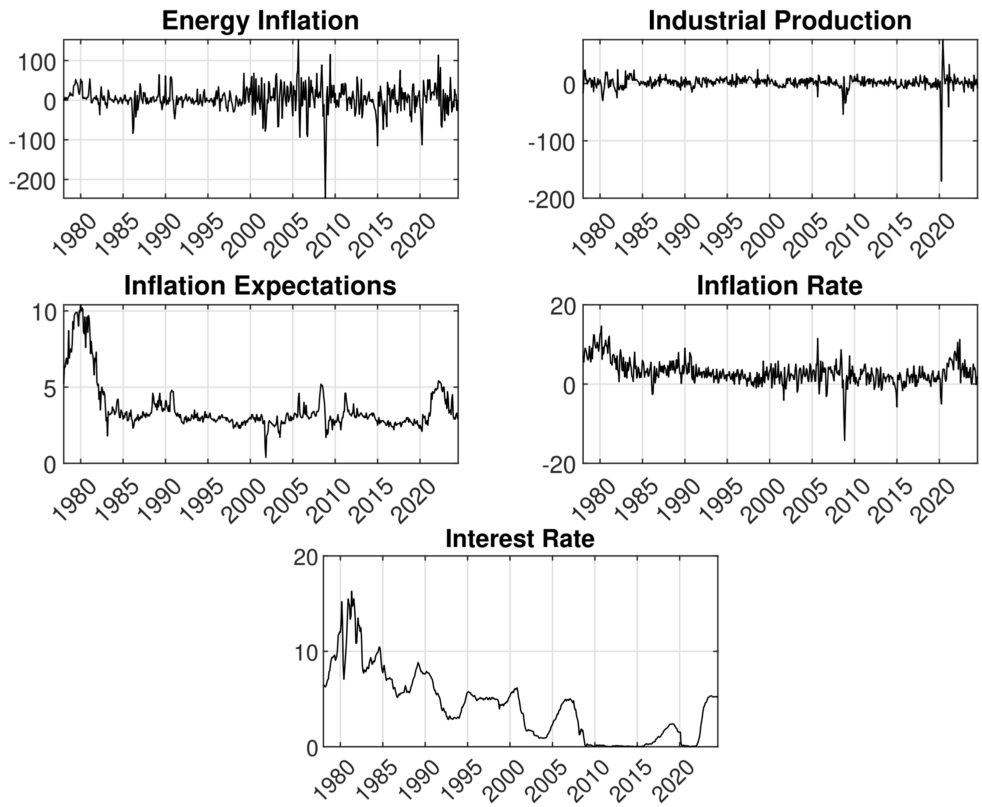


Figure 7: **Monthly time series dataset annualized** This figure shows the monthly time series for the energy prices, industrial production, inflation expectations, inflation, and interest rate from 1978M1 to 2024M6.

4.2.1 Monthly results: Full sample

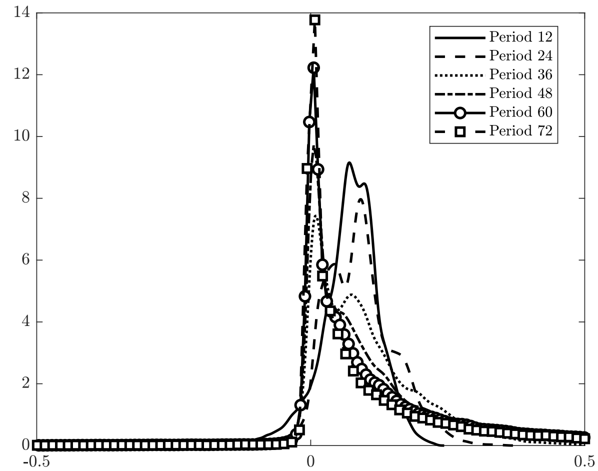


Figure 8: **Distributional QIRF** This figure shows the distributional QIRF from a median inflation expectations shock to the inflation rate. We carry out the calculations using Algorithm 1.

	τ				
π	10%	25%	50%	75%	90%
16%	0.4832	0.4788	0.5239	0.4850	0.1639
50%	0.7806	0.7611	0.8023	0.8566	0.8250
84%	1.1562	1.1230	1.1341	1.3521	1.7998
Point Estimate	0.9120	0.7152	0.8513	0.8719	0.9106
Cumulative IRF	0.5129	0.9477	2.0696	3.7200	6.3252

Table 7: **Long-run inflation** Rows one to four shows the results for the long-run response from the inflation rate equation to a shock to the expectation inflation in 1pp. Inference is provided using Algorithm 2. The last row shows the cumulative impulse response from a permanent 1pp shock to the inflation expectations at 72 months.

4.2.2 Monthly results: Pre-Covid

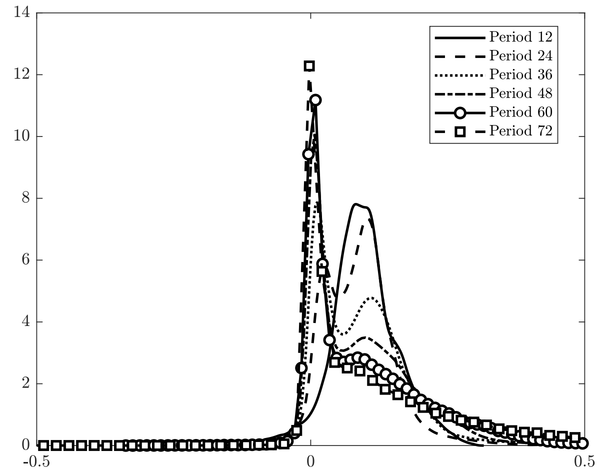


Figure 9: **Distributional QIRF** This figure shows the distributional QIRF from a median inflation expectations shock to the inflation rate. We carry out the calculations using Algorithm 1.

	τ				
π	10%	25%	50%	75%	90%
16%	0.5741	0.5550	0.6255	0.5555	0.4849
50%	0.7485	0.7363	0.7901	0.7479	0.6874
84%	0.9488	0.9351	0.9771	0.9734	0.9237
Point Estimate	0.7499	0.6958	0.8671	0.7983	0.6513
Cumulative IRF	0.5302	0.8413	2.2311	3.7703	5.4034

Table 8: **Long-run inflation** Rows one to four shows the results for the long-run response from the inflation rate equation to a shock to the expectation inflation in 1pp. Inference is provided using Algorithm 2. The last row shows the cumulative impulse response from a permanent 1pp shock to the inflation expectations at 72 months.

5 Conclusion

This study provides new evidence on how inflation expectations shape the entire distribution of realized inflation rates. Our findings show that while shocks to the median of inflation expectations lead to a significant but temporary increase in average inflation—lasting up to three years—these shocks also generate a persistent upside risk in the upper tail of the inflation distribution, with effects that endure for more than six years. Importantly, we find that higher-order moments of the expectations distribution—specifically, dispersion (standard deviation) and skewness—play distinct roles in shaping inflation outcomes. A positive shock to the standard deviation of expectations amplifies right-tail inflation risks, significantly increasing inflation at higher quantiles, whereas a negative shock to skewness causes a temporary upward shift in the median inflation rate without notable quantile-dependent effects. These results highlight that different moments of the expectations distribution transmit to inflation in asymmetric and nonlinear ways, which are not captured by conventional mean-based models.

Collectively, our findings underscore the importance for policymakers of monitoring not just the central tendency but also the heterogeneity and asymmetry in inflation expectations, as de-anchoring in higher-order moments can propagate persistent inflation risks even when median expectations appear stable. Paralleling the findings of Coibion and Gorodnichenko (2025), we also find that expectations play an important role, especially when inflation is elevated. This distributional perspective offers valuable insights for the design of monetary policy, particularly in periods of heightened macroeconomic uncertainty, and suggests that future research should further explore how these effects evolve across different economic regimes.

References

Aastveit, K. A., Bjørnland, H. C., Cross, J. L., & Kalstad, H. O. (2024). *Unveiling inflation: Oil shocks, supply chain pressures, and expectations* (tech. rep.). BI Norwegian Business School Working Paper.

Acharya, V. V., Crosignani, M., Eisert, T., & Eufinger, C. (2023). *How do supply shocks to inflation generalize? evidence from the pandemic era in europe* (tech. rep.). National Bureau of Economic Research.

Adams, J. J., & Barrett, P. (2024). Shocks to inflation expectations. *Review of Economic Dynamics*, *54*, 101234.

Ahn, H. J., & Farmer, L. E. (2024). Disagreement about the term structure of inflation expectations.

Basistha, A., & Nelson, C. R. (2007). New measures of the output gap based on the forward-looking new keynesian phillips curve. *Journal of Monetary Economics*, *54*(2), 498–511.

Beaudry, P., Hou, C., & Portier, F. (2024). *The dominant role of expectations and broad-based supply shocks in driving inflation* (tech. rep.). National Bureau of Economic Research.

Chang, Y., Gómez-Rodríguez, F., & Hong, M. G. H. (2022). *The effects of economic shocks on heterogeneous inflation expectations*. International Monetary Fund.

Chavleishvili, S., & Manganelli, S. (2024). Forecasting and stress testing with quantile vector autoregression. *Journal of Applied Econometrics*, *39*(1), 66–85.

Coibion, O., & Gorodnichenko, Y. (2025). *Inflation, Expectations and Monetary Policy: What Have We Learned and to What End?* (Tech. rep.). National Bureau of Economic Research Working Paper 33858.

Engle, R. F., & Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.

Fofana, S., Patzelt, P., & Reis, R. (2024). Household disagreement about expected inflation.

Gali, J., & Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, *44*(2), 195–222.

Gali, J., Gertler, M., & Lopez-Salido, J. D. (2005). Robustness of the estimates of the hybrid new keynesian phillips curve. *Journal of Monetary Economics*, *52*(6), 1107–1118.

Hallin, M., Paindaveine, D., & Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L1 optimization to halfspace depth. *The Annals of Statistics*, *38*(2), 635–669.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics*, *100*(5), 831–843.

Kilian, L., & Zhou, X. (2022). Oil prices, gasoline prices, and inflation expectations. *Journal of Applied Econometrics*, *37*(5), 867–881.

Kilian, L., & Zhou, X. (2023). A broader perspective on the inflationary effects of energy price shocks. *Energy Economics*, *125*, 106893.

Kocherlakota, N. R. (2016). *Sluggish inflation expectations: A markov chain analysis* (NBER Working Paper No. 22009). National Bureau of Economic Research.

Koenker, R. (2005). *Quantile regression* (Vol. 38). Cambridge: Cambridge university press.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics*, *9*(1), 155–176.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.

Koenker, R., & Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, *71*(1-2), 265–283.

Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, *101*(475), 980–990.

Lopez-Salido, D., & Loria, F. (2024). Inflation at risk. *Journal of Monetary Economics*, 103570.

Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about inflation expectations. *NBER Macroeconomics Annual*, *18*, 209–248.

Montes-Rojas, G. (2017). Reduced form vector directional quantiles. *Journal of Multivariate Analysis*, *158*, 20–30.

Montes-Rojas, G. (2019). Multivariate quantile impulse response functions. *Journal of Time Series Analysis*, *40*(5), 739–752.

Montes-Rojas, G. (2022). Estimating impulse-response functions for macroeconomic models using directional quantiles. *Journal of Time Series Econometrics*, *14*(2), 199–225.

Morley, J., & Panovska, I. B. (2020). Is business cycle asymmetry intrinsic in industrialized economies? *Macroeconomic Dynamics*, *24*(6), 1403–1436.

Panovska, I., & Ramamurthy, S. (2022). Decomposing the output gap with inflation learning. *Journal of Economic Dynamics and Control*, *136*, 104327.

Reis, R. (2020). *The people versus the markets: A parsimonious model of inflation expectations*. CFM, Centre for Macroeconomics.

Reis, R. (2023a). The burst of high inflation in 2021-22: How and why did we get here? In M. D. Bordo, J. H. Cochrane, & J. B. Taylor (Eds.), *How monetary policy got behind the curve—and how to get back* (pp. 203–252). Hoover Institution Press.

Reis, R. (2023b). What can keep euro area inflation high? *Economic Policy*, *38*(115), 495–517.

Rudd, J. B. (2022). Why do we think that inflation expectations matter for inflation?(and should we?) *Review of Keynesian Economics*, *10*(1), 25–45.

Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *Journal of the American Statistical Association*, *103*(481), 397–409.

Werning, I. (2022). *Expectations and the rate of inflation* (tech. rep.). National Bureau of Economic Research.

White, H., Kim, T.-H., & Manganelli, S. (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, *187*(1), 169–188.