The Dominant Role of Expectations and Broad-Based Supply Shocks in Driving Inflation

Paul Beaudry, Chenyu Hou and Franck Portier*

March 2024

version 2.1

Abstract

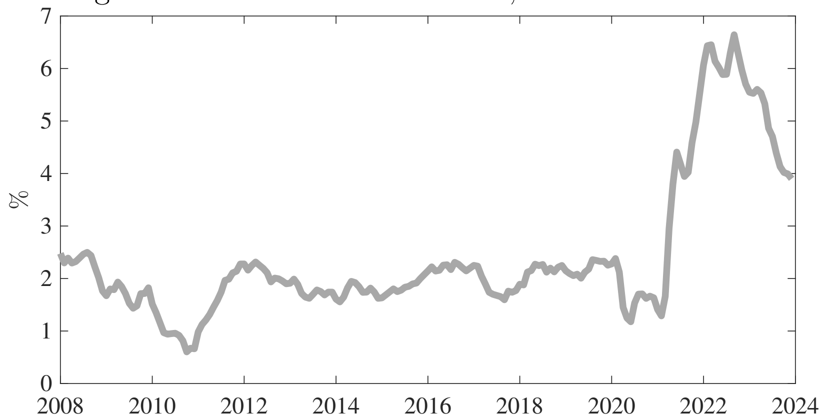
The object of this paper is to assess the role of supply shocks, labour market tightness and expectation formation in explaining bouts of inflation. We begin by showing that a quasi-flat Phillips curve, which was popular prior to the pandemic, still fits the post-2020 US data well and that changes in short term inflation expectations induced by supply shocks likely played a major role in the recent inflation episode. We then document features of the joint dynamics of inflation and inflation expectations. Given the difficulty of reproducing these dynamics under rational expectations, we propose and evaluate a model with imperfect information and bounded rationality. In our model, agents see sectoral inflations as being driven by a component common to all the sectors of the economy and by sector-specific shocks. When supply shocks affect many sectors (what we refer to as a broad-based supply shock), agents infer that the common component of inflation has increased, which drive persistent inflation dynamics through their effect of expectations. We show that departure from full rationality is minor, but that it is enough for broad-based supply shocks to be amplified and propagated over time in a manner needed to explain the data.

*Paul Beaudry: Vancouver School of Economics, University of British Columbia and NBER, paul.beaudry@ubc.ca. Chenyu Hou: School of Management and Economics, Chinese University of Hong Kong, Shenzhen, houchenyu@cuhk.edu.cn. Franck Portier: Department of economics, University College London and CEPR, f.portier@ucl.ac.uk.

Introduction

While inflation in many countries has been coming down quickly from the heights of 2022, core inflation has been coming down at a slower pace and this can be challenging for central banks who aim for a timely return to a pre-pandemic level of near 2%. For the US, this pattern for core inflation can be seen in Figure 1.

Figure 1: US Core CPI Inflation, 2008m1-2023m12



Notes: The observed series is Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average. We plot monthly observations of the percent change from one year ago to obtain a slightly smoother picture. In the rest of the paper, we will use quarter-to-quarter changes.

One of the main tools used by central bankers to understand inflation is the expectations-augmented Phillips curve which links inflation to expectations of future inflation and overall market tightness, either measured by an output gap, an unemployment gap or vacancy-to-unemployment ratio.[1] When viewed through the lens of the Phillips curve, the persistent period of high inflation that is not explained by supply shocks can be attributed to either an overly-hot labour market or by elevated expectations of inflation. A third possibility is that the Phillips curve is a rather unstable object and should be considered of limited value for thinking about current inflation. The later more unorthodox view is sometimes motivated by perceived repeated failures of the Phillips curve. For example, in the decade between the "Great Financial Crisis" and the pandemic –instead of under-predicting inflation as has recently been common– the Phillips curve often leads to a "missing deflation puzzle".

This paper begins by examining what a pre-2020 view of the Phillips curve tells us about the recent determinants of inflation and whether such a Phillips curves actually remains a good tool for understanding inflation. We complement our Phillips curve analysis with a VAR analysis of the joint dynamics of inflation, inflation expectations and labour market

[1]The later has been advocated as a better proxy for labour market tightness by, among others, Ball, Leigh, and Mishra [2022], Michaillat and Saez [2022] and Benigno and Eggertsson [2023].

tightness, where inflation expectations used in both cases are taken from the Michigan Survey of Consumers. The VAR analysis is especially relevant as it will become the target for our later structural analysis of how inflation and inflation expectations interact to create a rich joint dynamics following different supply shocks. Both our Phillips curve exploration and our VAR analysis come to similar conclusions. First, the recent episode does not suggest any important break from past behaviour. Second, both analyses suggest that inflation is mainly driven by expectations and quasi-i.i.d. supply shocks, with labour market tightness playing a very secondary role. In other words, the data maintain support for a close-to-flat Phillips curve view with short term inflation expectations being an important driver of inflation. In particular, the VAR analysis highlights how inflation expectations formed at t almost perfectly match realized inflation at $t + 1$, while inflation and inflation expectations are almost unrelated to labour market outcomes. We finish this section by showing why a Philips curve model with rational expectations does poorly in accounting for the joint dynamics of inflation and inflation expectations. .

We then turn to examining some of the empirical determinants of household inflation expectations. In particular, we show that these inflation expectations react to sectoral inflation measures in a way that deviates markedly from their importance in the overall CPI basket. Instead, inflation expectations seem to be influenced by that common component of sectoral data, where the main common component places very different weights than the CPI across items/sectors. So agents appear to update their expectations very differently if inflation is driven by a broad-based increase in prices (as measured by the common component) than if the same level of inflation is driven by only one or two items. This observation will be a key element in our model of inflation expectations.

In the fourth section of the paper we propose a model of the joint determination of inflation and inflation expectations that builds on the previously mentioned empirical patterns with the aim of explaining the VAR patterns. For the determinants of inflation, our model simply imports the New Keynesian Philips Curve specification we discuss and evaluate in Section 1. For the determinant of inflation expectations, we make two departures relative to a full information rational expectations benchmark. First, we endow agents with two inflation realizations; a realization of headline inflation and a realization of an inflation signal that can be interpreted as reflecting broad-based shocks. Second, we model the agents as quite sophisticated but not fully informed of the data generating process. They extract a common component from their sectoral inflation signals using a Kalman filter and use this common component to predict future inflation.[2] In this filtering process, agents treat the re-

[2]Lansing [2009] has introduced boundedly-rational inflation expectations in a New-Keynesian model, with a focus on low frequency properties of inflation.

alization of the common component in inflation as uncorrelated with the noise in their signal even though the two may not be independent in equilibrium due to the role of expectations themselves driving inflation. In doing so, agents depart from a fully rational benchmark. We show that such a model of inflation expectations can capture very well the joint dynamics of inflation and inflation expectations, as described by the VAR presented in Section 2, while simultaneously being consistent with easily observable moments of the inflation process, so that agent's beliefs are almost self-confirming.

While the narrative behind our model is most easily expressed in reference to the role of broad-based inflation shocks in driving inflation expectations, we do not actually use the disaggregated data in the estimation of our model. Instead, the "broad-based" inflation signal is treated as a latent variable and we complete our analysis by showing that our estimate of this latent variable correlates closely with the common component directly estimated from the disaggregated data.

The narrative that arises from our model of inflation and inflation expectations can be expressed as follows. Inflation can be seen as driven by two different types of shocks: first by narrow supply shocks that create very little dynamics even when substantial since such shocks are not readily transmitted to expectations as they do not create a perceived generalized increase in prices; second by a broad-based supply shock that transmits to inflation expectations and thereby creates something close to a self-fulling inflation episode. Such a model is shown to explain why inflation can remain stable during long periods even in the presence of many narrow but unsynchronized supply shocks. But, in contrast, inflation tends to broaden and persist when the economy is hit simultaneously by several supply shocks in different sectors.

1 Should We Maintain or Throw out the Flat Phillips Curve View?

Prior to the COVID-19 pandemic, many studies of the Phillips curve suggested that its slope was quite flat. Although the identification of the slope of the Phillips curve can be difficult, the work by Hazell, Herreño, Nakamura, and Steinsson [2022] advanced on this point by exploiting cross-state variation. Their finding was that the slope of the Phillips curve was significantly positive, but nevertheless quite small. Furthermore, they found that this flatness property did not arise post 90's, but appears to be a feature of the data going back into the 1960s. They concluded that inflation expectations likely played a dominant role in the inflation episode on the 1970s and early 1980s. In related work (Beaudry, Hou, and Portier [2023]), we found similar results with the inflation expectations drawn from the Michigan Survey of Consumers helping to explain much of the variation in inflation since late 1960.

In this section, we aim to examine whether the recent inflation episode should lead one to revise/reject/update the flat Phillips curve view. To this end, we start with the following very parsimonious view of the (quarterly) Phillips curve:

$$\pi_t = \beta\pi^e_{t+1} + \gamma_g \, \text{gap}_t + e_t, \tag{1}$$

π_t is quarter-to-quarter Headline CPI inflation, π^e_{t+1} is the mean of the one-year-ahead expectation of CPI inflation drawn from the Michigan Survey of Consumers.[3] We choose this measure of expectations as it goes back all the way to the 1960s. As prices are set by firms, using a measure of business expectations could be preferred. However, such measures are not readily available over such a long time span.[4] The gap represents labour market tightness and is measured by the minus unemployment gap.[5]

We initially do not estimate this Phillips curve, but choose commonly accepted parameters values. In particular, γ_g is set to 0.0138 according to what estimated in Hazell, Herreño, Nakamura, and Steinsson [2022] [6] and β is set to 0.99, which is a common value in the literature. We refer to this as our baseline Phillips curve. Note that given the low value of γ_g, that Phillips curve is pretty flat.

From this baseline Phillips curve, we construct residuals $\widehat{e}_t = \pi_t - 0.99\pi^e_{t+1} - 0.0138 \, \text{gap}_t$ over the period 2008q1-2023q1 and plot them against our measure of the gap. This is the plot in Panel (a) of Figure 2.[7] The dark dots are Phillips curve residuals for the post-2020 period, while the grey dots cover the period from 2008q1 to 2019q4. We overlayed

[3]Our preference would be to use quarter-to-quarter expected inflation. However, the Michigan Survey of Consumers reports expectations for CPI inflation for the next year. To extract a quarter estimate from this data we rescale the one-year-ahead expected inflation assuming survey respondents believe that quarter-to-quarter inflation follows an AR(1) process with persistence $\tilde{\rho}$, that needs not to be equal to the actual persistence of inflation. The estimated $\tilde{\rho} = 0.89$. This adjustment only affects the point estimate of β. For more details, we refer to Beaudry, Hou, and Portier [2023].

[4]The Cleveland Fed Survey of Firms' Inflation Expectations (SoFIE) is available since 2018Q2. Over the sample 2018Q2-2023Q1, the correlation between the Michigan Survey of Consumers' inflation expectations and the Cleveland Fed Survey of Firms' ones is 0.9330

[5]The unemployment gap is computed as the unemployment rate minus the noncyclical rate of unemployment from FRED (series name *NROU*).

[6]In Hazell, Herreño, Nakamura, and Steinsson [2022], the authors provide an implied aggregate slope of the Phillips curve, with R-CPI as the measure of inflation and negative unemployment gap as the measure of market slackness. From footnote 24 in Hazell, Herreño, Nakamura, and Steinsson [2022], this aggregate slope of the Phillips curve is $= 0.58 \times 0.0062 + 0.42 \times 0.0243 = 0.0138$.

[7]In our analysis for post-2008 data, we exclude the 2020q2 observation given that measuring labour market tightness at a time of massive lock downs is very controversial.

on these figures two estimated relations between the Phillips curve residuals and labour market tightness over 2008q1 to 2019q4; one in linear form and one in cubic form. The coefficient for this regression are presented in Columns 1 and 2 of Table 1. We also calculate the standard deviation σ_ϵ of the Phillips curve residuals for this sample as well as that from the prior sample running from 1968q1 to 2007q4. The first thing to note is that the standard deviation of the Phillips curve residuals are only slightly higher over the period 2008q1-2023q1 than over the period 1968q1-2007q4, although we are considering Headline inflation and not controlling for supply shocks. More interestingly, we detect no significant link between these residuals and the unemployment gap.[8] If the residuals were strongly and positively associated with the labour market tightness measure, including possibly in a non-linear fashion, this would put in question the validity of a flat Phillips curve view in the more recent data. However, these forecast errors do not suggest that a steeper Phillips curve would better explain the more recent data episode. The same exercise is repeated using Core CPI inflation as a measure of inflation. As it can be seen in Columns 3 and 4 of Table 1 and on Panel (b) of Figure 2, there is again no sign of a steepening of the Phillips curve. Furthermore, the standard deviation of the Phillips curve residuals is now smaller post 2008 as compared to previously. If anything, this flat Phillips curve fits better post 2008 data than over the period 1968-2007.

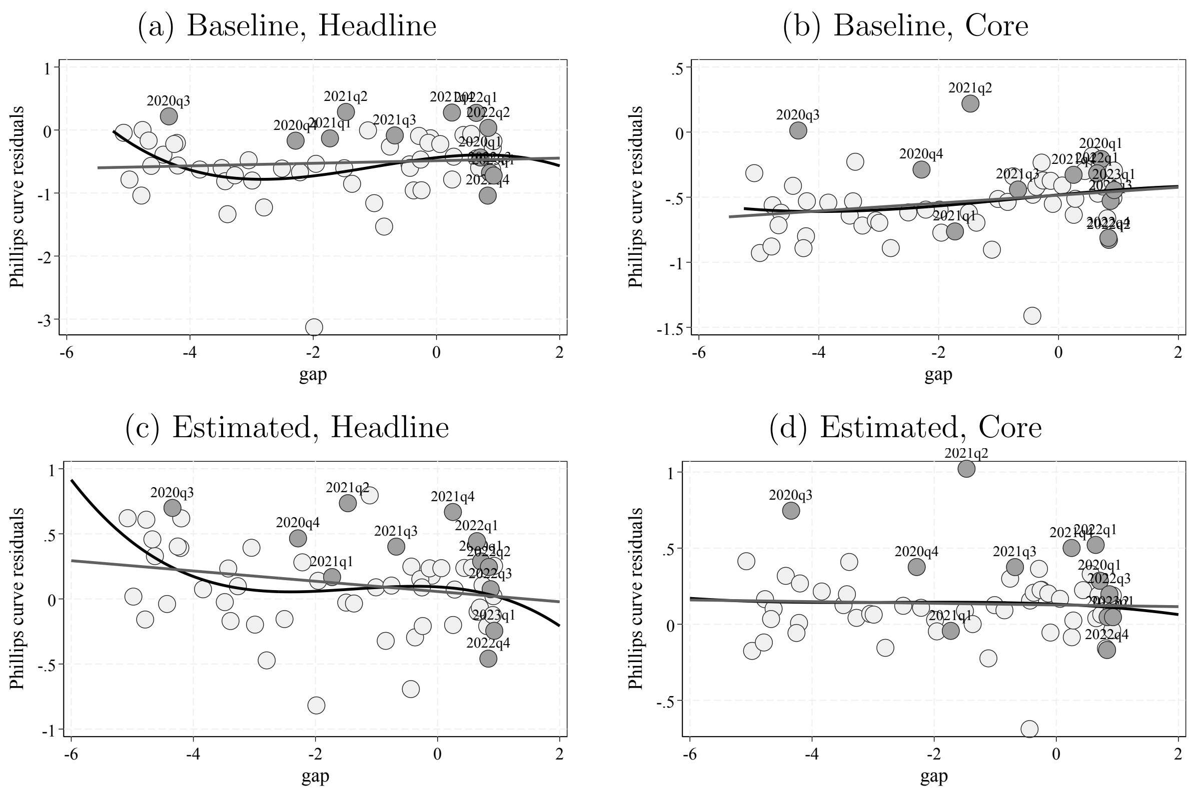
In addition to this baseline Phillips curve, we also directly estimate a Phillips curve over the sample 1969q1 to 2007q4, and look at the implied out of sample forecast errors for the period 2008q1-2023q1. Details of this estimation is provided in the appendix and follow Beaudry, Hou, and Portier [2023]. Since the determinants of inflation in Phillips Curve are endogenous, we follow Barnichon and Mesters [2020] and estimate our Phillips curve by Instrumental Variables using estimated monetary shocks as instruments.[9] To reduce the risk of biases, we estimated the Phillips curve using Core inflation, and as argued in Beaudry, Hou, and Portier [2023], we include the real interest rate as an additional explanatory variable to capture a potential cost channel of monetary policy.[10] The estimated Phillips curve is

[8]Endogeneity issues can complicate inference from this figure. Under the null hypothesis that our Phillips curve is well specified, forecast errors should reflect supply or markup shocks. In contrast, if the slope is mis-specified, then the forecast errors would also include a term directly related to labour market tightness. In this later case, such mis-specification should show up as a systematic relationship between forecast errors and labour market tightness. However, a strong endogenous relation between supply shocks and labour market tightness could hide this effect. Although we see this as a possibility, we believe it is likely of second order importance over this sample. We have explored the potential relevance of this issue by using high frequency identified monetary shocks to instrument labour market tightness when regressing forecast errors on labour market tightness. We have not found evidence of endogeneity bias.

[9]To be specific, we use twelve lags of the Romer and Romer's [2004] shocks (extended by Wieland and Yang [2020]) and their square as instruments.

[10]Also note that the presence of direct cost channel matters in the context of asking which expectations are more relevant when estimating Phillips curve. See a comprehensive discussion in Reis [2023].

Figure 2: Out-of-Sample Residuals from Phillips Curves



Notes: Panels (a) and (b) of this figure plots the out-of-sample residuals of the baseline Phillips curve (1) against a measure of labor market tightness (minus) the U.S. Congressional Budget Office unemployment gap), for two measures of inflation (Headline and Core CPI). The gray lines show the estimated linear or cubic relation between residuals and labor market tightness (see Table 1 for the estimated coefficients). Light dots correspond to pre-2020 observations and dark ones to post-2020. We exclude 2020q2 from this graph. Panels (c) and (d) of this figure repeat the analysis with the augmented estimated Phillips curve (2).

Table 1: Projection of the Philips Curve Residuals $\widehat{e}$ on the Gap, 2008-2023

	Headline Baseline		Core Baseline		Headline Estimated		Core Estimated	
	linear	nonlinear	linear	nonlinear	linear	nonlinear	linear	nonlinear
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
gap	0.02	0.11	0.03	0.04	-0.04	-0.03	-0.01	-0.02
	(0.035)	(0.113)	(0.016)	(0.056)	(0.022)	(0.072)	(0.016)	(0.055)
gap^2		-0.05		-0.00		-0.04		-0.01
		(0.087)		(0.043)		(0.056)		(0.042)
gap^3		-0.02		-0.00		-0.01		-0.00
		(0.015)		(0.007)		(0.009)		(0.007)
N	60	60	60	60	60	60	60	60
$\sigma_{\widehat{e},60-07}$	.324		.315		.276		.291	
$\sigma_{\widehat{e},08-23}$	.528		.256		.338		.245	

Notes: in columns (1) to (4), the Phillips curve residuals are obtained from the baseline Phillips curve (1). In columns (5) to (8), they are obtained from the augmented Phillips curve (2) estimated over the sample 1969Q1-2007Q4. $\sigma_{\widehat{e}}$ is the standard deviation of the Phillips curve residuals. The standard errors of estimated coefficients are between parentheses.

therefore

$$\pi_t = \beta\pi^e_{t+1} + \gamma_g \text{ gap}_t + \gamma_r(i_t - \pi^e_{t+1}) + e_t. \tag{2}$$

Table 2 displays the results from estimation. The coefficients we obtain for γ_g is quite close to that our of baseline Phillips curve, with the exception that we obtain a slightly steeper Phillips curve than that reported in Hazell, Herreño, Nakamura, and Steinsson [2022]. Our estimate our γ_g is .04 and significant at standard levels, which is more than twice that of .0138 from Hazell, Herreño, Nakamura, and Steinsson [2022], but not significantly different from it. Our estimate for β is also .99.

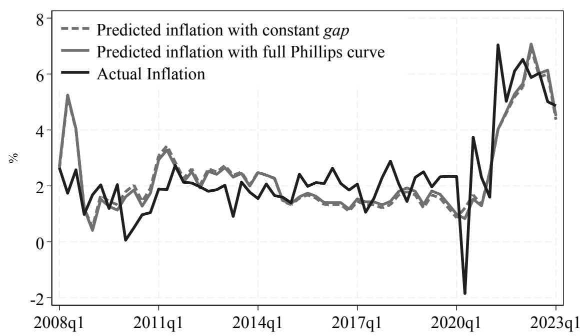
In Panel (d) of Figure 2, we report the out of sample Phillips curve residuals from our estimated model for the period 2008q1-2023q1 plotted against our measure of labour market tightness, with again a linear and cubic fit superimposed. We see a very similar pattern to that observed in Panels (a) and (b), with very little link between the forecast errors and labour market tightness, even though these are out of sample forecast. The estimated coefficients for these linear and cubic fits are displayed in Columns (7) and (8) of Table 1 So again, these observations do not suggest that a stepper Phillips curve is helpful/needed to explain the 2008q1-2023q1 period. Note that the fit of this estimated Phillips curve, as measured by the standard deviation of the residuals $\sigma_{\widehat{e}}$ is better out-of-sample than in-sample. In appendix A, we confirm similar finding using the vacancy to unemployment rate as an alternative measure of labour market tightness, as well as exploring the effect of the use market base measures of expectations. For completeness, panel (c) of Figure 2 and Columns (5) and (6) of Table 1 repeat the analysis when using the Headline CPI to estimate the Phillips curve. Results are similar.

The overall out-of-sample fit of the baseline Phillips curve can also be seen in Panel (a) and (b) (for Headline) and (c) and (d) (for Core) of Figure 3. As can be seen, the forecasted inflation tracks actual inflation reasonably well, with exceptions such as 2021q2, where actual inflation was well above predicted, as is consistent with a strong supply shock in that quarter associated with transport bottlenecks. In Panels (a) and (c), we also plot a counterfactual path of inflation over the period assuming that the unemployment gap had been equal to its sample mean over the forecast horizon. Here we see that this counterfactual inflation series is very similar to the predicted inflation path using the full model. This reflects the very weak direct role that labour market tightness is playing in our estimated Phillips curve given that the slope is quite small. The second counterfactual we examine imposes that inflation expectations are constant over the forecast period and set at their average. This counterfactual path is reported together with our forecasted rates of inflation in Panel (b). and (d). Here we see clearly that it is expected inflation that mainly drives the fit of the predicted inflation, as when we omit the role of inflation expectations, the fit greatly deteriorates.

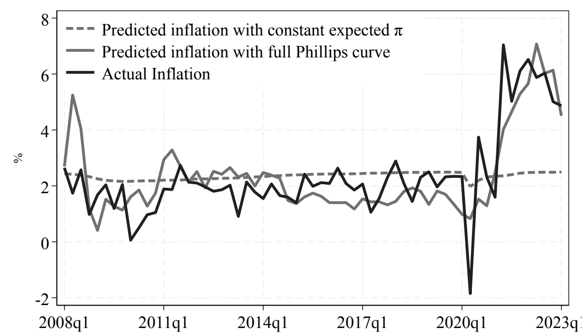
In summary, Figures 2 and 3 illustrate how a very simple three-variable Phillips curve, either estimated prior to the recent periods of missing deflation and high inflation, or based on the work of Hazell, Herreño, Nakamura, and Steinsson [2022], fits the post-2007 data and post-2020 quite well. The lack of a relationship between forecast errors and labour market tightness indicates that the Phillips curve is likely still quite flat. The forecast error are also close to i.i.d., which is consistent with them being interpreted as supply shocks. The ongoing flatness of the Phillips curve implies that persistently high levels of core inflation in 2022 and early 2023 are unlikely explained by labour market tightness. Instead, it appears primarily driven by elevated levels of inflation expectations. Such observations can be both reassuring and uncomfortable. On the one hand, they support the view that the Phillips curve framework likely remains a relevant framework for thinking about inflation developments. However, they also suggest that the main driver of inflation is expected inflation itself. The latter observations can be uncomfortable as it questions/downplays the traditional role of pure demand management as key to controlling inflation and instead pushes forward the central importance of expectations management– and the more amorphous role of inflation psychology– in controlling inflation.

Figure 3: Counterfactual Simulations from the Baseline Phillips Curve

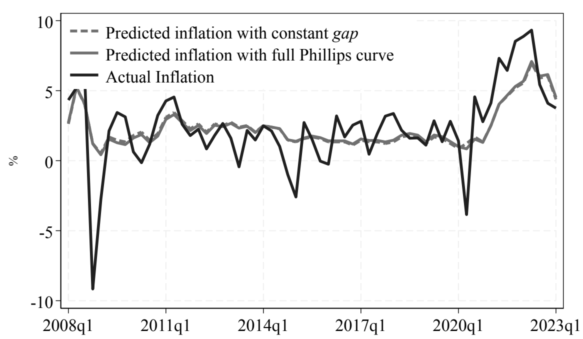
(a) With gap at its mean, Core



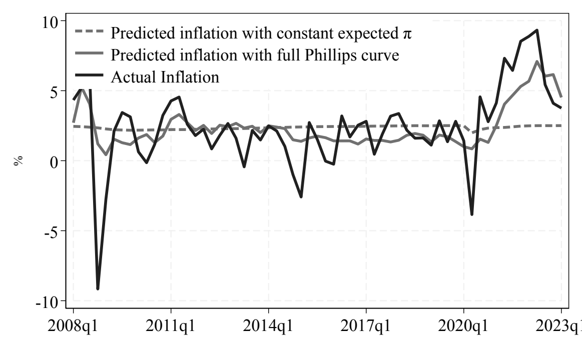
(b) With inflation expectations at their mean, Core



(c) With gap at its mean, Headline



(d) With inflation expectations at their mean, Headline



Notes: These counterfactual simulations are done using the baseline Phillips curve (1), using either Headline or Core CPI.

Table 2: Estimated Phillips Curves, 1969-2007

	Headline	Core
β	1.15*	0.99*
	(0.031)	(0.048)
γ_g	0.07*	0.04*
	(0.020)	(0.015)
γ_r	0.13*	0.25*
	(0.031)	(0.042)
Observations	144	144
J Test	15.201	10.684
(jp)	(0.887)	(0.986)
Weak ID Test	7.825	12.245

Notes: All results are using IV-GMM procedure, Newey-West HAC standard errors with six lags are reported in parentheses. The constant term is omitted from the table. Real oil price and its lag are also omitted for the regression with Headline CPI. All regressors are instrumented using six lags of Romer and Romer's [2004] shocks (as extended by Wieland and Yang [2020]) and their squares as instruments. A $\star$ indicates significance at 5%. Sample is 1969Q1-2007Q4.

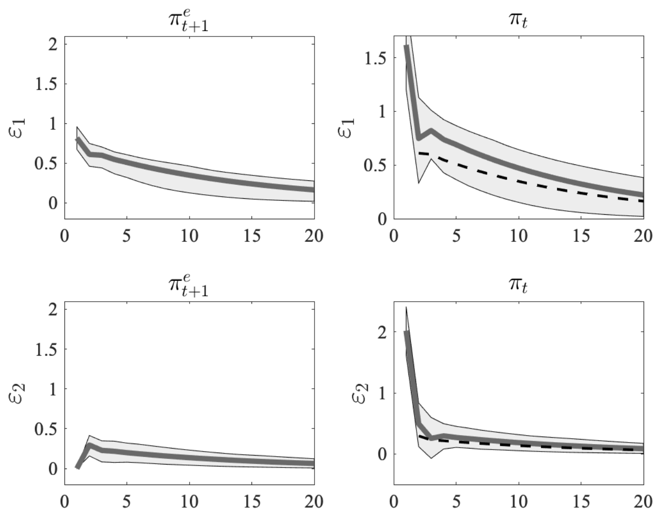
2 The Joint Dynamics of Inflation, Inflation Expectations and Labour Market Tightness

In this section we first provide VAR based evidence to give further support to the view that inflation dynamics appears to be largely explained by inflation expectations, with labour market tightness playing a very secondary role. We then show that our baseline Phillips curve model with rational expectations and full or incomplete information is unlikely to account for the observed joint dynamics.

2.1 VAR Evidence

To this end, Figure 4 plots impulse response generated from a bi-variate VAR using inflation and the inflation expectations estimated over the period 1969q1-2023q2. The data for inflation remains headline inflation and that for expectations is again drawn from the Michigan Survey of Expectations. The figure corresponds to impulse response associated with a Choleski orthogonalization of the VAR residuals, where the shock ε_2 does not have impact effect on inflation expectations while ε_1 is unrestricted but to be orthogonal to ε_2. We do not claim that these impulses responses capture the effects of structural shocks. Instead, we simply view them as a means of summarizing properties of the data. In particular, these two shocks will eventually be interpreted as combinations of the structural shock in our modelling of section 4. In Figure 4, together with the response of inflation to the two shocks $\{\pi_t\}_{t=1}^{20}$, we also represent (dashed line) the expected inflation response $\{\pi^e_{t+1}\}_{t=1}^{19}$ shifted by one period. Two properties of this joint dynamics can be observed. First, inflation expectations

Figure 4: Impulse Responses in the 2-VAR (π_t, π^e_{t+1})



Notes: this Figure plots the impulse responses to a one standard deviation shocks ε_1 and ε_2. These shocks are obtained from a Choleski orthogonalization. The estimated VAR features two lags of Headline CPI inflation and of the Michigan Survey of Consumers inflation expectations. The dashed line is the expected inflation response $\{\pi^e_{t+1}\}_{t=1}^{19}$ shifted by one period. Sample is 1969Q1-2023Q1. Shaded area represents the 95% confidence band.

formed at time t match very closely realized inflation at $t+1$. This is true for both shocks, and therefore for any linear combination of these shocks. Accordingly, it suggests that either agents are very good at predicting inflation or instead that expected inflation may be driving inflation. Second, these impulse responses suggest that there is some combination of shocks (that forms ε_2) that don't transmit to expected inflation and this leads to very temporary rise in inflation. While on the other had, there is some combination of shocks (that forms ε_1) that have a large effect of inflation, transmit to expected inflation, and this is associated with persistent inflation. Indeed, in the response to ε_1, the first and second serial correlation of inflation response are 0.45 and 0.44, while for ε_2 they are 0.19 and 0.06.

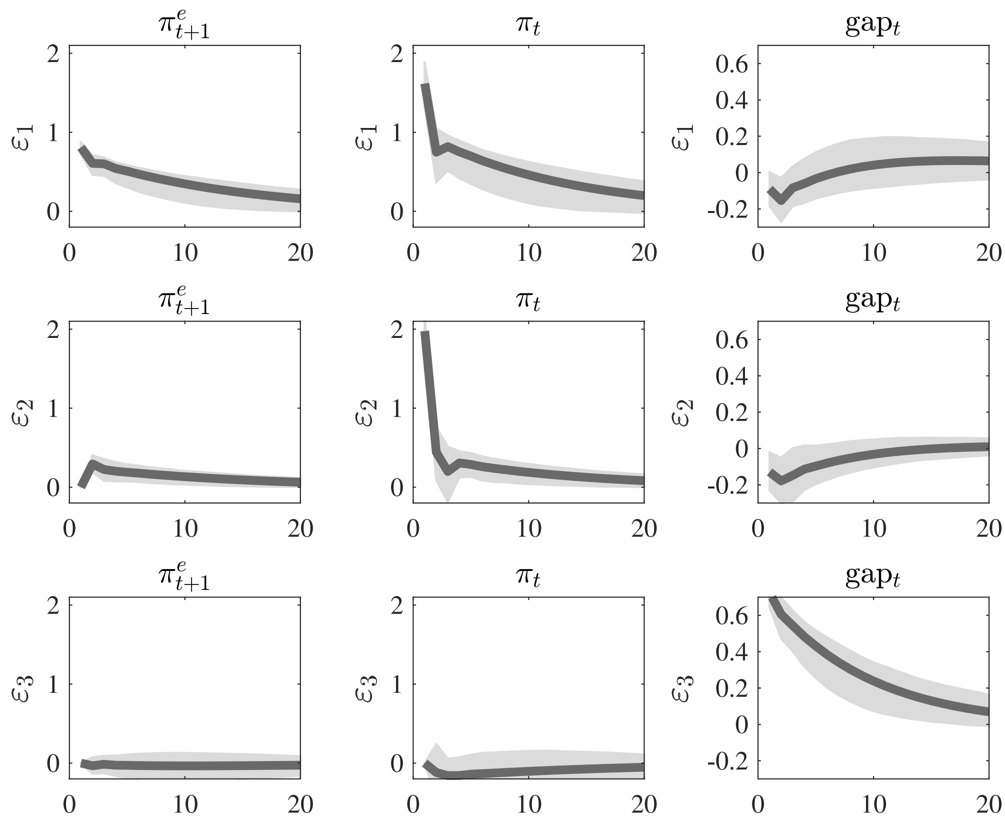
In Appendix C, we show that these impulse responses are essentially unchanged if we change the number of lags or if we drop post-2008 data from the estimation. It also holds if the VAR is estimated on 2008-2023.[11] This suggests that the recent inflation dynamics

[11]In that case, we keep only one lag in the VAR as the sample is small.

continue to obey the dynamics that was observed earlier, suggesting no substantial break in the process.

In Figure 5, we present the impulse responses from a trivariate VAR that builds on our bi-variate VAR by including (minus) the unemployment gap as the third variable, again using a Choleski indentification; ε_3 affects only the gap on impact, ε_2 affects only the gap and inflation on impact and ε_1 is unrestricted but to be orthogonal to the two other shocks.

Figure 5: Impulse Responses in the 3-VAR ($\pi_t, \pi^e_{t+1}, \text{gap}_t$)



Notes: this Figure plots the impulse responses to a one standard deviation shocks ε_1, ε_2 and ε_3. These shocks are obtained from a Choleski orthogonalization. The estimated VAR features two lags of Headline CPI inflation, the Michigan Survey of Consumers inflation expectations and of (minus) the employment gap. Sample is 1969Q1-2023Q1. Shaded area represents the 95% confidence band.

The main feature we want to highlight from this figure is the quasi complete separation between inflation and inflation expectations on the one hand and labour market tightness on the other hand. This can seen examining how the sub-block of impulse response to ε_1 and ε_2 for inflation and inflation expectations generated by this three variable VAR is almost identical to that generated by our two-variable VAR. Furthermore, we can see that the shock to unemployment ε_3 has essentially no impact on inflation or inflation expectations, while the two other shocks that drive almost all the variance of inflation and expected inflation have very effect on unemployment. In summary, these two (non-structural) VARs appear consistent with the view that the Phillips curve is likely very flat and that the persistent component of inflation may well be driven by expected inflation.

2.2 The Challenge in Explaining the Joint Dynamics of Inflation and Inflation Expectations under Rational Expectations

As we have argued above, inflation expectations appear to potentially be an important driver of inflation. However, this may only be a proximate cause. It may well be the case that both inflation and inflation expectations are driven by a third variable, and expectations are simply good at capturing this force. Accordingly, we ask in this subsection whether the joint dynamics of inflation and inflation expectations reflected in our bi-variate VAR could be produced by our baseline Phillips curve model under rational expectations, where the dynamics of the labour market tightness and supply shocks would be the more fundamental factor. To that end, first assume that inflation is generated from the baseline Phillips curve

$$\pi_t = \beta \pi^e_{t+1} + \gamma_g \, \mathrm{gap}_t + e_t, \tag{3}$$

with $e_t \sim N(0, \sigma_e^2)$, $\beta = .99$, $\gamma_g = 4 \times 0.0138$[12] and where we use quarterly (annualized) Headline inflation, the Michigan Survey of Expectations and (minus) the unemployment gap. As discussed in the previous section, the fit of the baseline Phillips curve hasn't worsened in the recent period, as illustrated on Figure 6.

We also assume that the gap follows an exogenous AR(1) process:

$$\mathrm{gap}_t = \rho \, \mathrm{gap}_{t-1} + v_t, \tag{4}$$

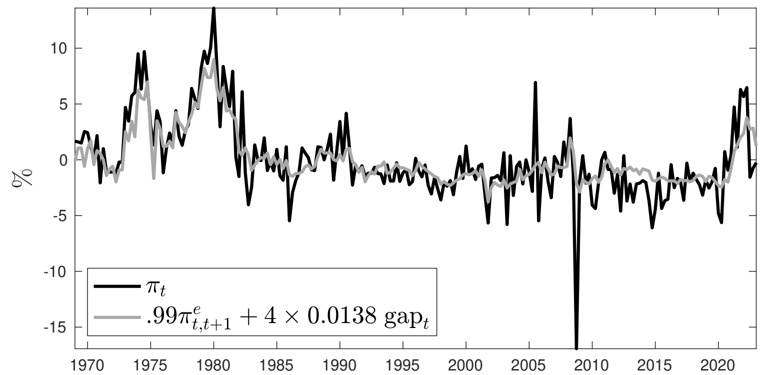
with $v_t \sim N(0, \sigma_v^2)$. We estimate an AR(1) process for gap_t.[13] The estimated parameters are $\rho = 0.89$ (s.e. 0.03) and $\sigma_v^2 = 9.05$ (s.e. 0.33). We obtain an estimate of $\sigma_e^2 = 5.00$ from the residual of (3).

Now let's solve that Phillips curve and gap process model under Rational Expectations and with or without full information, and show that data generated by such models cannot replicate our VAR findings.

[12] We multiply the Hazell, Herreño, Nakamura, and Steinsson's [2022] estimate by four because we use here anualized quarterly inflation and expected inflation.

[13] When we estimate an AR process with more lags, only the estimate on the first lag is significant.

Figure 6: The Fit of the Baseline Phillips Curve, 1969-2023



Notes: this Figure plots Headline CPI inflation together with inflation predicted by the baseline Phillips curve (1)

Full Information Rational Expectations (FIRE): When economic agents form rational expectations, they use the correct model (3)-(4) and form expectations $\pi^e_{t+1} = E^{FIRE}_t[\pi_{t+1}]$. Under full information, gap$_{t-1}$, v_t and e_t are observed. The solution to the model is given by:

$$\pi_t = \frac{\gamma_g}{1-\beta\rho} \text{ gap}_t + e_t, \tag{5}$$

$$E^{FIRE}_t[\pi_{t+1}] = \frac{\gamma_g \rho}{1-\beta\rho} \text{ gap}_t. \tag{6}$$

Full Information Rational Expectations (FIRE) with persistent supply shock: Looking at the properties of the Phillips curve residual $\widehat{e}$, it is significantly autocorrelated, although the autocorrelation coefficient is small. When we estimate an AR(1) process, we find low autocorrelation, with an estimated $\rho_e = 0.165$ (standard deviation = 0.05). Could this admittedly low persistence reverse the results we obtain with the i.i.d. supply shocks FIRE model? To answer that question, we also consider an extension of the FIRE model where we allow e_t to be persistent. The solution to such a FIRE model with persistent e_t is given by:

$$\pi_t = \frac{\gamma_g}{1-\beta\rho} \text{ gap}_t + \frac{1}{1-\beta\rho_e} e_t, \tag{7}$$

$$E^{FIREp}_t[\pi_{t+1}] = \frac{\gamma_g \rho}{1-\beta\rho} \text{ gap}_t + \frac{\rho_e}{1-\beta\rho_e} e_t. \tag{8}$$

Incomplete Information Rational Expectations (IIRE): Here we look at a counterpart of our Section 4 main model but with rational expectations. We assume that agents correctly understand the model (3)-(4), but that they do not observe the gap and the aggregate supply shock e_t. Instead, we assume that the economy is composed of N sectors with sector-specific supply shocks. In sector j, sectoral inflation is generated by:

$$\pi_{j,t} = \beta \pi^e_{t+1} + \gamma_g \text{ gap}_t + e_{j,t}, \tag{9}$$

where $e_{j,t} \sim N(0, \sigma_j^2)$. We assume that agents observe a signal that adds-up demand inflationary pressures γ_g gap$_t$ and a subset $\mathcal{S}$ of the sector-specific supply shocks e_{jt}:

$$s_t = \gamma_g \text{ gap}_t + \sum_{j \in \mathcal{S}} e_{j,t} \tag{10}$$

and we will use the notation

$$\epsilon_t = \sum_{j \in \mathcal{S}} e_{j,t}. \tag{11}$$

with $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. Economic agents have rational expectations and know that inflation is generated by

$$\pi_t = \beta \pi^e_{t+1} + \gamma_g \text{ gap}_t + \epsilon_t + w_t, \tag{12}$$

where $w_t = e_t - \epsilon_t$ is the residual supply shock. The agents use Bayes Rule and the knowledge of the model to form beliefs of e_t and gap$_t$, and to compute $\pi^e_{t+1} = E_t^{IIRE}[\pi_{t+1}]$. Denoting $\mathcal{K}$ the Kalman gain in this computation, the solution to the model is:[14]

$$E_t^{IIRE}[\pi_{t+1}] = (1 - \mathcal{K}\gamma_g)\rho E_{t-1}^{IIRE}[\pi_t] + \frac{\gamma_g \rho \mathcal{K}}{1 - \beta\rho}(\gamma_g \text{ gap}_t + \epsilon_t), \tag{13}$$

$$\begin{aligned} \pi_t = & \beta(1 - \mathcal{K}\gamma_g)\rho E_{t-1}^{IIRE}[\pi_t] + \frac{\gamma_g^2 \mathcal{K}\rho\beta + (1-\beta\rho)\gamma_g}{1 - \beta\rho} \text{ gap}_t \\ & + \frac{\mathcal{K}\rho\beta\gamma_g + (1-\beta\rho)\gamma}{1 - \beta\rho}\epsilon_t + w_t. \end{aligned} \tag{14}$$

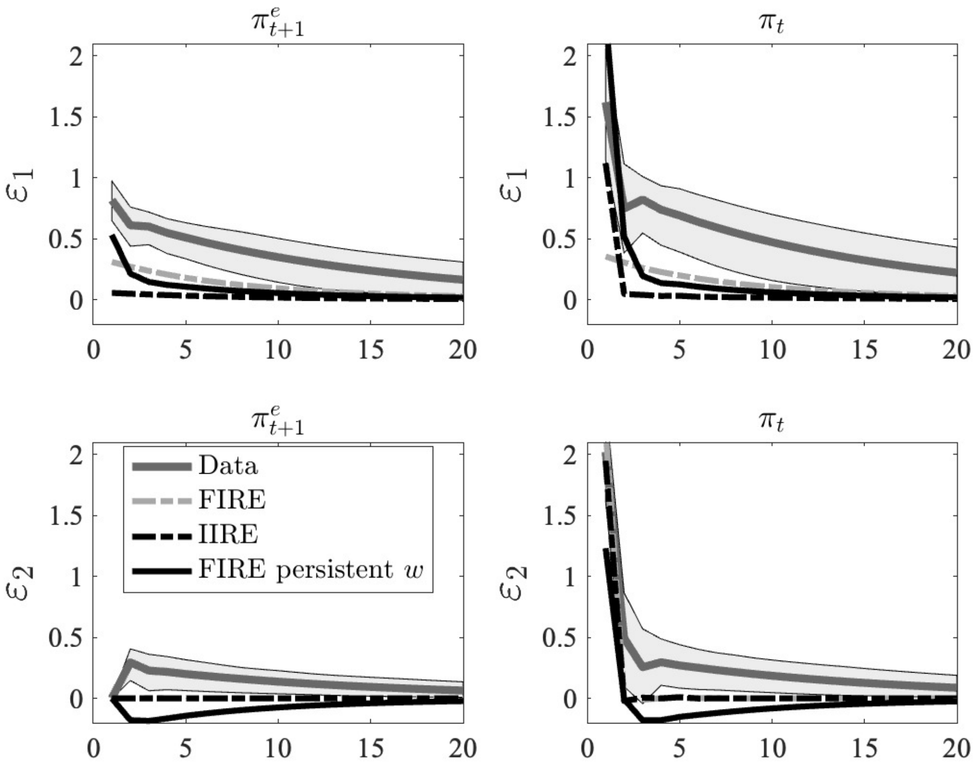
We estimate σ_ϵ by Simulated Method of Moments. The model is simulated 200 times, the (π_t, π^e_{t+1}) VAR is estimated in each simulation and σ_ϵ is set to minimise the distance between the average Impulse Response Functions obtained from the model with their empirical couterpart, when the same Choleski decomposition is used for model and data. The resulting estimate is $\sigma_\epsilon = 1.08$, and this implies $\sigma_w = 1.96$.

Results: For each of the three models considered, using the estimated parameters, we simulate 200 random samples with the same length as the data. With these generated data,

[14] The detailed derivations are left to Appendix F.

we estimate the bi-variate VAR and see how well it can match the impulse responses derived from our empirical bi-variate VAR. The result can be seen in Figure 7. The impulse responses from the generated data do not match that derived from actual data neither in shape nor level, even if in the case of IIRE model we estimate σ_ϵ to deliberately match these IRFs.

Figure 7: The Joint Process of π and π^e_{t+1} in the Data and in Models with Rational Expectations



Notes: on this Figure, the solid grey line plots the impulse responses to a one standard deviation shocks ε_1 and ε_2 estimated with data from Michigan Survey of Consumers inlfation expectations and Headline CPI. Sample is 1969Q1-2023Q1. Shaded area represents the 95% confidence band. The other lines plot the average impulse responses (over 200 simulations of length 216) obtained from the same VAR estimated on simulated data, when the Data Generating Process is either the FIRE model (5)-(6) assuming w_t is i.i.d, the IIRE model (13)-(14) or the FIRE model assuming w_t is persistent.

We can also illustrate the performances of these models in matching the joint dynamics of inflation and inflation expectations by looking at the set of moments presented in Table 3. We see that none of the rational expectations models can create inflation as volatile as the data. The simulated data have much lower persistence and variances in both realized and expected inflation, as compared to the actual data.

In part, this should not be too surprising. Given the rather small slope of the Phillips

Table 3: Moments of π and π^e_{t+1} in the Data and in Models with Rational Expectations

	$var(\pi_t)$	$cov(\pi_t, \pi_{t-1})$	$cov(\pi_t, \pi_{t-2})$	$var(\pi^e_{t+1})$	$corr(\pi^e_{t+1}, \pi^e_t)$
Data	12.52	7.70	6.99	3.92	0.90
FIRE	5.49	0.45	0.39	0.42	0.87
IIRE	5.10	0.05	0.04	0.00	0.87
FIRE, persistent w_t	7.59	1.61	0.57	0.61	0.64

Notes: For the three models, the moments are computed as the average across 200 random samples of same length than the data.

curve implied by the work of Hazell, Herreño, Nakamura, and Steinsson [2022] and the low persistence of the supply shocks, it is hard to see how shocks to labor market tightness or supply shocks could generate important volatility in inflation expectations. In particular, the difficulty for the rational expectations models to match the VAR observations reflects the difficulty to generate both large and persistent responses of inflation expectations. Rational agents understand that the underlying persistent factor moving inflation plays only a limited role. As a result, they shouldn't expect future inflation to have persistent movement as well. Moreover, limited information does not help with the situation. In fact, the IIRE case has the most muted response among all the three models and generates the least volatile as well as the least persistent inflation. This is not only because the rational agent understands the persistent labor market condition plays little role. They also realize that their signals are not informative because of the large noise-to-signal ratio due to low γ_g. As a result, their expectations do not respond much to the movements in labor market conditions.

Our take away from this this section is that the VAR dynamics for inflation and inflation expectations are unlikely to be explained by a Rational Expectations New Keynesian Phillips curve model. In the next section, we explore what may drive inflation expectations.

3 The Information Encoded in Inflation Expectations

In this section we provide some preliminary evidence for our conjecture that inflation expectations may be driven by a perceived common component across sector-specific inflations. Our narrative is quite simple. Inflation is defined as a generalized increase in prices. So when agents (both firms and households) see prices increase in many sectors (although not necessarily in *all* sectors), they take this to be sign that inflation is taking hold. This affects their expectations and this incite firms to increase prices.

To provide support to this conjecture, we take the set of sectoral inflation data for the 24 expenditure categories used by the BLS to construct the CPI (see Table D.1 in Appendix D for a description) and extract the common component of these sectoral inflations. To do so, we estimate a Dynamic Factor Model[15] in which sectoral inflations are driven by one common persistent factor $\mathcal{C}_t$ which follows an AR(1) process. The model takes the following form:[16]

$$\pi_{j,t} = \alpha_j \, \mathcal{C}_t + e_{j,t}, \tag{15}$$

$$\mathcal{C}_t = \rho_{\mathcal{C}} \, \mathcal{C}_{t-1} + v_t, \tag{16}$$

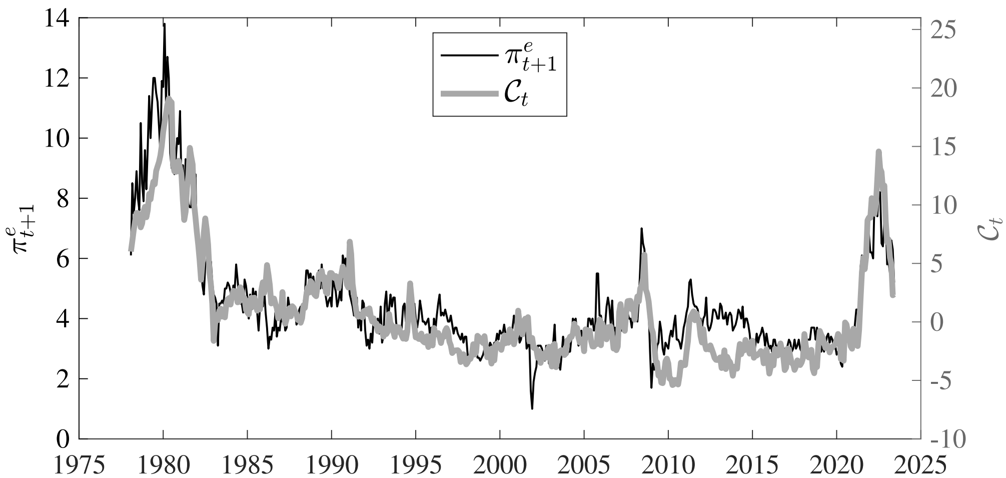
where j stands for the sector/category and $\pi_{j,t}$ is the month-to-month inflation of each sector/category j. How the common component $\mathcal{C}_t$ affects different $\pi_{j,t}$ may be different, and is captured by α_j. Data are monthly over the sample 1978m1-2023-m5 for seventeen sectoral inflations and shorter for the remaining seven (see Table D.1). We estimate that model with Kalman Filter Smoothing using a Maximum Likelihood Estimator that accounts for the missing observations. Estimated parameters are $\rho_{\mathcal{C}}$ and $\{\alpha_j, \sigma_j^2\}_{j=1}^{24}$. Note that σ_v^2 is not separately identifiable from α_j's, so that we normalize it to 1. The parameters estimates we obtained are presented in Table D.2 in Appendix D. The important element to note is that all the α_j's are positive, so that $\mathcal{C}_t$ is indeed capturing a common component in inflation. In Figure 8, we plot expected inflation and the common component $\mathcal{C}_t$. As can be seen, the common component of sectoral inflations tracks the expected inflation quite well. It is interesting to compare the (normalized) weights of each category in the common components to the expenditure share of that category, as used by the BLS to compute CPI. The weights in the common component are computed as the relative Kalman gain associated to each sectoral inflation in updating the common component. Those two sets of weights are represented in Figure 9. It is evident that are very different. The correlation between these two sets of weight is 0.15, with a 95% confidence set [-0.01, 0.32]. This shows that the common component encodes some information about broad-based shock, as the weights significantly differ from the CPI ones. We then regress our measure of expected inflation (from the Michigan Survey of Consumers) on the common component of the disaggregated inflation data and annualized month-to-month CPI (headline) inflation. The resulting regression is (with standard errors between parenthesis):

$$\pi^e_{t+1} = \underset{(0.09)}{0.66} + \underset{(0.02)}{0.78} \; \mathcal{C}_t + \underset{(0.01)}{0.09} \; \pi_t^{HL} + u_t, \tag{17}$$

[15] See Stock and Watson [2011] for a thorough review.

[16] There are no constants in the model because we use standardized sectoral inflations $\pi_{j,t}$.

Figure 8: Michigan Survey of Consumers Inflation Expectations π^e_{t+1} and the Common Component $\mathcal{C}_t$.

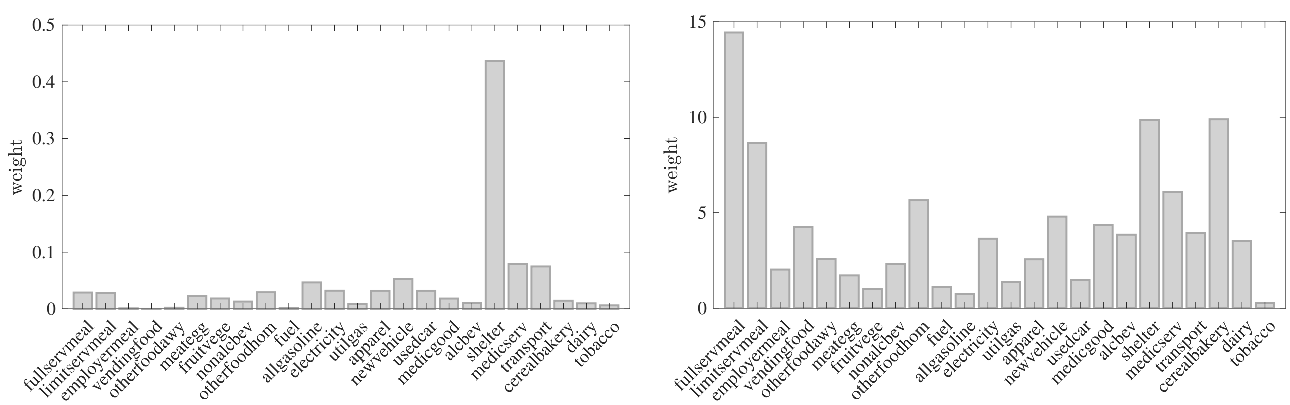


Notes: this Figure plots inflation expectations (as measured by the Michigan Survey of Consumers) together with the common component $\mathcal{C}_t$. $\mathcal{C}_t$ is obtained from the estimation of (15) and (16) on the 24 expenditure categories used by the BLS to construct the CPI.

Figure 9: Comparing the Weights in the Common Component and in the CPI Basket

(a) BLS weights

(b) Common Component weights



Notes: Panel (a) plots the weight of each category in the CPI basket, while panel (b) plots the (normalized) weights in the common component $\mathcal{C}_t$.

where the data are at a monthly frequency with 545 observations and an adjusted R-Squared of 0.82. The common component enters significantly in the regression on top of Headline CPI, which indicates that economic agents (here the households) may be using it to forecast inflation. The adjusted R-Squared of regressing inflation expectations from the Michigan Survey of Consumers on only common component is 0.80 and the adjusted R-Squared of regressing inflation expectations on only Headline CPI is merely 0.45.

4 A Bounded Rational Model of Inflation and Inflation Expectations Dynamics with Broad-Based Supply Shocks

In this section our aim is to take up the following challenge. We want to combine our baseline Phillips curve with a particular bounded rationality theory of inflation expectations to see whether it can match the bi-variate VAR for inflation and expected inflation. Our theory of expectations will build on the idea that agents may form expectations by extracting a common component from disaggregated data. However, agent's perception of the data generating process will be slightly simpler than the actual data generating process. In particular, we will assume that they don't conceptualize the data generating in its structural form. Instead they view persistent inflation as driven by a common component that they need to infer from the data.

4.1 The Model

The model has two main departures from the Full Information Rational Expectations case. First, agents observe only a subset of the sectoral inflations when forming their inflation expectations. This is the source of imperfect information. Second, they believe that sectoral inflations depend on a common persistent component and a transitory sector-specific component, but fail to fully understand the relations between these two components. The later is the source of bounded rationality.

Sectors: The model directly adopts the baseline aggregate Phillips Curve discussed in the previous sections as part of the data generating process:

$$\pi_t = \beta \pi^e_{t+1} + \gamma_g \, \text{gap}_t + e_t. \tag{18}$$

The headline aggregate inflation π_t is a weighted average of sector-specific inflations. Consider that each sectoral inflation obeys a sectoral Phillips Curve similar to the aggregate one:[17]

$$\pi_{j,t} = \beta\pi^e_{t+1} + \gamma_g \text{ gap}_t + e_{j,t}, \tag{19}$$

where j stands for a specific sector and $e_{j,t}$ is the i.i.d sector-specific shock with distribution $N(0, \sigma_j^2)$. We will describe the expectations formation process after we have introduced the agents perceived laws of motion. As in Section 2.2, the gap follows an AR(1) process:

$$\text{gap}_t = \rho \text{ gap}_{t-1} + v_t. \tag{20}$$

Assuming there are N categories in the CPI, headline inflation is the weighted average of the sectoral series:

$$\pi_t = \sum_{j=1}^{N} \omega_j \pi_{j,t}, \tag{21}$$

where the ω_js are the sectoral weights, and $\sum_j \omega_j = 1$. We can therefore rewrite the Phillips curve as

$$\pi_t = \beta\pi^e_{t+1} + \gamma_g \text{ gap}_t + \sum_{j=1}^{N} \omega_j e_{j,t}. \tag{22}$$

Imperfect Information: We assume that economic agents observe only a subset or the sectoral inflations.[18] This subset of sector indexes j is denoted $\mathcal{S} \subseteq \{1, 2, ..., N\}$, and its cardinality is $m \leq N$. Agents will use these sectoral inflations as signals $s_{j,t} = \pi_{j,t}$ to form expectations. In our framework, it is mathematically equivalent to replace the set of signals $\{s_{j,t}\}_{j\in\mathcal{S}}$ by one aggregate signal s_t. Indeed, for independent signals $j = 1, ..., m$ with different precisions $\frac{1}{\sigma_j^2}$, observing m different signals is equivalent to observing a signal with precision: $\sum_{j=1}^{m} \frac{1}{\sigma_j^2}$.[19]

[17] Note that we could allow for different responses of disaggregate inflations to expectations (β_j) and labor market conditions ($\gamma_{g,j}$). This would give some extra degrees of freedom to the model in our quantitative analysis later. At this stage, we abstract for this extra layer of heterogeneity between sectors and assume $\beta_j = \beta$ and $\gamma_{g,j} = \gamma_g$ for all j. We could also introduce both aggregate expected inflation expectations and sector j specific expected inflation. Again, we keep minimal the deviations from the standard model, and only introduce aggregate expected inflation.

[18] We do not take a stand on what are the causes for the agents not being fully attentive to all the sectoral inflations. Many frameworks such as rational inattention (Sims [2003]) or sparsity (Gabaix [2014]) can create such a partial information environment.

[19] See Appendix E for the derivation.

Perceived Law of Motion and Bounded Rationality: We have shown in Section 2 that neither the FIRE nor the IIRE models could easily match the joint dynamics if inflation and inflation expectations, as obtained from a simple VAR. Accordingly, in this section we build on our analysis in Section 3 by assuming that agents form their inflation expectations by extracting a common component from their observations of sectoral inflations. In particular, the agents are assumed to observe $\{\pi_{j,t}\}_{j\in\mathcal{S}}$, and to form expectations using this information.[20] The element of bounded rationality is that they perceive the law of motion for sectoral inflations as being composed of a common persistent component $\widetilde{z}_t$ and sector-specific transitory components $\widetilde{e}_{j,t}$, instead of the sectoral Phillips curves (22). Their estimate of the persistent component will be used to form expectations about future aggregate inflation.[21]

Formally, agents have the following perceived law of motion of the economy:

$$\pi_{j,t} = \widetilde{z}_t + \widetilde{e}_{j,t} = s_{j,t}, \tag{23}$$

$$\pi_t = \widetilde{z}_t + \sum_{j\in\mathcal{S}} \omega_j \widetilde{e}_{j,t} + \sum_{j\notin\mathcal{S}} \omega_j \widetilde{e}_{j,t}, \tag{24}$$

$$\widetilde{z}_t = \widetilde{\rho}_z \widetilde{z}_{t-1} + \widetilde{\varepsilon}_{z,t}, \tag{25}$$

with $\widetilde{e}_{j,t} \sim N\left(0, \widetilde{\sigma}_j^2\right)$ $\forall j$ and $\widetilde{\varepsilon}_{z,t} \sim N(0, \widetilde{\sigma}_z^2)$. The parameters $\widetilde{\rho}_z$, $\widetilde{\sigma}_j^2$ and $\widetilde{\sigma}_z^2$ denote the perceived persistence and variances for the law of motion. Here also, the perceived independent $\widetilde{e}_{j,t}$ with different precisions $\frac{1}{\widetilde{\sigma}_j^2}$ can be collapsed into a single $\widetilde{\varepsilon}_t$ with precision: $\frac{1}{\widetilde{\sigma}_\epsilon^2} = \sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2}$.

The agents endowed with this Perceived Law of Motion will be quite sophisticated. They will use the observed sectoral inflation series to infer the state $\widetilde{z}_t$ – the perceived common component affecting inflation – and take these beliefs about $\widetilde{z}_t$ to form expectations of aggregate inflation. However, they are bounded rational in the sense that they treat the state-space system (23)-(25) as a standard signal-extraction problem in which the noise generated by not observing all the sectors is treated as uncorrelated with the state $\widetilde{z}_t$. Because the sectoral inflations are common to all the agents, when the aggregate expectation responds to the signals, these signals will affect the state variable, which will feedback to the signals via the Phillips curve (22). As agents fail to recognise that the hidden state may be correlated with the noise in their signals, this has the potential to create a positive feedback loop from noise to inflation.[22] This mis-interpretation of the data is the key to creating an amplification

[20] For example, D'Acunto, Malmendier, Ospina, and Weber [2020] documents prevalent evidence that consumers use prices they are exposed to in their daily shopping experience to form inflation expectations.

[21] Such a signal-extraction formulation is widely used in the literature about inflation expectations. Examples are among others Coibion and Gorodnichenko [2015a] and Bordalo, Gennaioli, Ma, and Shleifer [2020].

[22] In the Perceived Law of Motion, the information relevant for forming expectation is the common com-

mechanism in our expectations formation model.[23]

Solving the Model: Using equation (23) together with the aggregate inflation (24), we obtain the observational equations from the agent's perspective:

$$\begin{cases} s_t &= \widetilde{z}_t + \widetilde{\epsilon}_t \\ \pi_t &= \widetilde{z}_t + \widetilde{\eta}_t \end{cases} \tag{26}$$

whereas the actual signal and aggregate inflation are generated by:

$$\begin{cases} s_t &= \beta\pi^e_{t+1} + \gamma_g \text{ gap}_t + \epsilon_t \\ \pi_t &= \beta\pi^e_{t+1} + \gamma_g \text{ gap}_t + \eta_t \end{cases} \tag{27}$$

In these equations, s_t is an average signal from the m independent sectoral signals, with precision $1/\widetilde{\sigma}^2_\epsilon$. Using (22) and (24), we denote $\eta_t = \sum_{j=1}^N \omega_j e_{j,t}$ and $\widetilde{\eta}_t = \sum_{j=1}^N \omega_j \widetilde{e}_{j,t}$, with $\eta_t \sim N(0, \sigma^2_\eta)$ and $\widetilde{\eta}_t \sim N(0, \widetilde{\sigma}^2_\eta)$. ϵ_t stands for the broad-based shock. The information encoded in s_t is also reflected in π_t, so that ϵ_t and η_t are correlated with correlation ϱ. For the same reason, $\widetilde{\epsilon}_t$ and $\widetilde{\eta}_t$ are correlated with correlation $\widetilde{\varrho}$. It is possible that $\varrho \neq \widetilde{\varrho}$ and both are non-zero.[24]

The agents face the state-space model described by (25) and (26). They solve a signal-extraction problem to form belief about the perceived common component, $\widetilde{z}_t$, and use it to form expectations about inflation. Because expectations formed in the current period will also affect current inflation and signals the agents see, we need to be specific about the timing of expectations formation and when expectations affect inflation. In the baseline version of the model, we will allow information extracted from the disaggregated data to be simultaneously determined with inflation, while we will assume that aggregate inflation (CPI) is released to households with a lag. This implies that any residual information in aggregate inflation does not feedback contemporaneously to itself.[25] Accordingly, we solve

ponent. As a result, the agent will treat the transitory sectoral supply shock $\widetilde{e}_{j,t}$ as a noise. However, $\widetilde{e}_{j,t}$ is different from a pure "noise shock" as in Lorenzoni [2009], which is typically interpreted as a demand shock. The crucial difference is that $\widetilde{e}_{j,t}$ will have direct impact on actual inflation even if it doesn't pass into expectation.

[23] Another feature of the above observational equation is that the signals are weighted by their precisions that do not depend on their actual weights ω_j in the aggregate price index. In other words, a category that has a very low weight in the aggregate price index can be more informative as a category that has significant weight but with a lower signal-to-noise ratio. This feature is in line with our finding in section 3 that the weights of disaggregate price indices in the CPI basket are not highly correlated with the estimated weights from Kalman Filter Smoothing.

[24] See Appendix E for a complete derivations of (27) and (26).

[25] In Appendix I, we reestimate the model under the assumption that inflation expectations are reported *after* observing current inflation. Differences in the results are minor.

the model adopting the following steps, which is easiest to present as having each period t being composed of two sub-periods:[26]

Step 1: In the first sub-period of t, the gap_t realizes. Economic agents form expectations using the signal s_t. We denote the expected inflation formed in this sub-period as $\pi_{t+1|t,0}$. This is the expectation π^e_{t+1} that enters in the Phillips curves (18) and (19). Consistent with the treatment in section 1 and section 2, we take these expectations as those observed in the Michigan Survey of Consumers.

Step 2: In the second sub-period of t, agents also observe aggregate inflation π_t. They use this extra information to update their belief about future inflation, denoted as $\pi_{t+1|t,1}$. This will be the prior on inflation that will be carried over to the next period $t+1$.

Step 3: Both expectations $\pi_{t+1|t,0}$ and $\pi_{t+1|t,1}$ are formed from beliefs on the common component $\widetilde{z}_t$. We denote these beliefs as $\widetilde{z}_{t|t,0}$ and $\widetilde{z}_{t|t,1}$. From equations (25) and (26), we see that the links between expected inflation and the common component are:

$$\pi_{t+1|t,k} = \widetilde{z}_{t+1|t,k} = \widetilde{\rho}_z \widetilde{z}_{t|t,k} \quad \forall k = 1, 2. \tag{28}$$

Step 4: Denote the innovations in agents' perceived information (signal and aggregate inflation)-generating process (26) as:

$$\begin{pmatrix} \widetilde{\epsilon}_t \\ \widetilde{\eta}_t \end{pmatrix} \sim N(\mathbf{0}, \widetilde{R}), \quad \widetilde{R} \equiv \begin{pmatrix} \widetilde{\sigma}^2_\epsilon & \widetilde{\varrho}\widetilde{\sigma}_\epsilon\widetilde{\sigma}_\eta \\ \widetilde{\varrho}\widetilde{\sigma}_\epsilon\widetilde{\sigma}_\eta & \widetilde{\sigma}^2_\eta \end{pmatrix}. \tag{29}$$

The agents form weights on $\widetilde{\epsilon}_t$ and $\widetilde{\eta}_t$, according to the Kalman filter, using their beliefs that s_t and π_t are correlated. Denoting ι_n a $n \times 1$ vector of ones, the stationary Kalman filter is given by:

$$\widehat{K} = \widetilde{\sigma}^2 \iota_2' (\iota_2 \widetilde{\sigma}^2 \iota_2' + \widetilde{R})^{-1}, \tag{30}$$

$$\widetilde{\sigma}^2 = \widetilde{\rho}_z^2 (\widetilde{\sigma}^2 - \widehat{K} \iota_2 \widetilde{\sigma}^2) + \widetilde{\sigma}_z^2, \tag{31}$$

where $\widetilde{\sigma}^2$ is the stationary posterior variance of belief on $\widetilde{z}_t$. Now note the $\widehat{K}$ is an 1×2 vector, with first element being the weight on the signal coming from the observed sectoral inflations, and the second being the weight on aggregate inflation π_t: $\widehat{K} \equiv (K \quad k)$.

Step 5: In the first sub-period of t, agents observe s_t and form a nowcast of $\widetilde{z}_t$:

$$\widetilde{z}_{t|t,0} = (1 - K)\widetilde{z}_{t|t-1,1} + K s_t. \tag{32}$$

[26] An alternative formulation would be thinking of the belief on z_t is formed conditional on information set $\mathcal{I}_t = \{\pi_{t-1}, s_t\}$ in each period t. The two will give equivalent representation of the model.

Step 6: In the second sub-period of t, agents observe π_t and update belief on $\widetilde{z}_t$:

$$\widetilde{z}_{t|t,1} = ((1-K-k)\widetilde{z}_{t|t-1,1} + Ks_t + k\pi_t). \tag{33}$$

Following these six steps, the solution of the model will be given by:

$$\pi^e_{t+1} = \pi_{t+1|t,0} = \frac{\widetilde{\rho}_z(1-K)}{1-K\beta\widetilde{\rho}_z}\widetilde{z}_{t|t-1,1} + \frac{\widetilde{\rho}_z K\gamma_g}{1-K\beta\widetilde{\rho}_z}\text{gap}_t + \frac{\widetilde{\rho}_z K}{1-K\beta\widetilde{\rho}_z}\epsilon_t, \tag{34}$$

$$\pi_t = \beta\pi_{t+1|t,0} + \gamma_g\text{gap}_t + \eta_t \tag{35}$$

$$\begin{aligned}\widetilde{z}_{t+1|t,1} = {} & \widetilde{\rho}_z\left(1-(K+k)\frac{1-\beta\widetilde{\rho}_z}{1-K\beta\widetilde{\rho}_z}\right)\widetilde{z}_{t|t-1,1} + \widetilde{\rho}_z\frac{(K+k)\gamma_g}{1-K\beta\widetilde{\rho}_z}\text{gap}_t \\ & + \widetilde{\rho}_z\left(\frac{K+\widetilde{\rho}_z Kk\beta}{1-K\beta\widetilde{\rho}_z}\right)\epsilon_t + k\widetilde{\rho}_z\eta_t.\end{aligned} \tag{36}$$

$$\text{gap}_t = \rho\text{gap}_{t-1} + v_t \tag{37}$$

Note that signal s_t and aggregate inflation π_t are realized according to (27), so expectations $\pi^e_{t+1} = \pi_{t+1|t,0}$ and signal s_t will be simultaneously determined.

The above expectations formation process has the potential to create persistent bouts of inflation in a quasi self-confirming way. When an adverse (positive) broad-base supply shock ϵ_t hits the economy, economic agents observe an inflation hike in many disaggregate inflation series. This causes them to update their view of the common component $\widetilde{z}_t$. If agents believe that their signal is quite informative, this will increase their expected inflation. Because the disaggregated inflations are public signals, the average expectation will increase which drives up the disaggregate prices they observe. This feedback channel can thereby amplify the response of expected inflation to broad-base shocks on impact.[27] Moreover, in the second sub-period, agents observe π_t which is also pushed up by expectations. This further supports their belief of positive realization in the common component. As a result, economic agents enter next period with a prior of high inflation. This then creates very persistent responses of both actual and expected inflation to broad-base price shocks. Furthermore, after the initial period of the shock, actual inflation will react almost one-to-one with expectations, which corresponds to the observation we made in the empirical VAR.

[27]Note that this amplification comes from the fact that the agents fail to understand expectations $\pi_{t+1|t,0}$ is part of the perceived common component $\widetilde{z}_t$. More precisely, they fail to understand the hidden state is correlated with the noise of their signal. This arises because the signals the agents observe are endogenous.

4.2 Illustrating the Respective Roles of Imperfect Information and Bounded Rationality using a Static Version of the Model

It may be useful to first consider a static model that conveys most of the intuitions for the mechanism at play in the dynamic one. In that static model, equilibrium inflation π depends on expected inflation (within the period) π^e and a shock that is i.i.d normal $e \sim N(0, \sigma_e^2)$:

$$\pi = \beta\pi^e + e, \tag{38}$$

with $\beta \in [0, 1[$. Economic agents do not observe inflation, but a noisy signal S:

$$S = \pi + \epsilon. \tag{39}$$

with $\epsilon \sim N(0, \sigma_\epsilon^2)$.

Consider first the Rational Expectation solution in which $\pi^e = E[\pi|_S]$. Agents filter the signal S to infer a conditional expectation of e, understanding that their conditional expectation of π influences inflation via equation (38). The equilibrium inflation is then given by:

$$\pi^{RE} = \frac{1-(1-\kappa)\beta}{1-\beta}e + \frac{\kappa\beta}{1-\beta}\epsilon, \quad \text{with} \quad \kappa = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_\epsilon^2}. \tag{40}$$

Both the noise ϵ and the shock e contribute the variability of π.

It is convenient to look at the case where the variance of fundamental shock e goes to zero ($\sigma_e^2 \to 0$). In that case, $\kappa = 0$ so that equilibrium inflation $\pi^{RE} = 0$, even if the signal S is noisy ($\sigma_\epsilon^2 \neq 0$). It is the same solution with full information ($\sigma_\epsilon^2 = 0$) or with imperfect information ($\sigma_\epsilon^2 \neq 0$).

Consider now a model with Bounded Rationality in which, instead of knowing (38), agents believe that inflation is i.i.d:

$$\pi = \widetilde{e} \tag{41}$$

where $\widetilde{e} \sim N(0, \widetilde{\sigma}_e^2)$. Observing the signal S, agents reads it as $S = \widetilde{e} + \varepsilon$, and believe that $\widetilde{e}$ and ε are orthogonal. Doing that signal extraction, one can obtain expected inflation and replace it in (38) to obtain equilibrium inflation:

$$\pi^{BR} = \frac{1}{1-\widetilde{\kappa}\beta}e + \frac{\widetilde{\kappa}\beta}{1-\widetilde{\kappa}\beta}\epsilon \tag{42}$$

with $\widetilde{\kappa} = \frac{\widetilde{\sigma}_e^2}{\widetilde{\sigma}_e^2 + \sigma_\epsilon^2}$. It is useful to look again at the limit case $\sigma_e^2 \to 0$. We obtain in that case

$$\pi^{BR} = \frac{\widetilde{\kappa}\beta}{1 - \widetilde{\kappa}\beta}\epsilon. \tag{43}$$

Here the noise accounts for *all* the variability of inflation, whereas it was totally eliminated with rational expectations. The reason is that because of their beliefs (41), the agents fail to see that inflation is orthogonal to the noise. Because $\widetilde{e}$ is a latent variable, there is no way they can detect that this correlation is not zero. Therefore, noise shocks ϵ impact expectations, therefore inflation, therefore expectations, etc..., which explains the "multiplier" term $\frac{1}{1-\widetilde{\kappa}\beta}$ in the equilibrium inflation equation.

One should nevertheless note that generically the variance of the signal S will not be equal to the agents belief $\widetilde{\sigma}_e^2 + \sigma_\epsilon^2$. It is reasonable to believe that agents would eventually notice that discrepancy and will change their belief to reach a "self-confirming" situation in which $\widetilde{\sigma}_e^2 + \sigma_\epsilon^2$ equals to the variance of the signal S. Would that kill the current "noise multiplier" we have previously highlighted? There are potentially multiple self-confirming equilibrium in this setup.[28] If $\beta > \frac{1}{2}$, the self-confirming equilibrium most likely to arise under learning is $\widetilde{\sigma}_e^2 = \frac{2\beta-1}{(1-\beta)^2}\sigma_\epsilon^2$. In that case, noise shocks are magnified in a quasi-self-fulfilling way, all the volatility of inflation comes from the noise ϵ, and the model equilibrium is qualitatively different from the case of Rational Expectations. Futhermore, the agents' belief that inflation is volatile cannot be contradicted by the observation given (41), that is *under their model of the economy*. As we will see, a similar mechanism is at play in our dynamic model.

4.3 Quantitative Analysis

Our aim not is to estimate the model's parameters by matching the impulse responses of the (π_{t+1}^e, π_t) VAR while simultaneously imposing that the parameters governing the agents' perceived law of motion be supported by the data. In particular, we want the parameters of the perceived law of motion to be consistent with that implied by the observed variance as well as the first and second auto-covariances of aggregate inflation. In that sense, the model is disciplined in two ways: first it aims at reproducing the VAR impulse responses and second it ensures that there are no large discrepancies between what agents believe and what they observe.

[28] An obvious self-confirming equilibrium is one in which $\widetilde{\sigma}_e^2 = 0$. If the agents believe that inflation is a constant, they will attribute all the variations in the signal to the noise ($\widetilde{\kappa} = 0$), and we will have indeed $\sigma_{\pi\text{BR}}^2 = 0$. We are back to Rational Expectations and inflation is not reacting to the noise. But under simple learning schemes, it can be shown that the agents will not converge to this equilibrium.

We fix the values of β, ρ, γ_g, σ_η, and σ_v at the values discussed in Section 1 when we use the baseline Phillips curve. We estimate the other parameters $\widetilde{\rho}_z$, $\widetilde{\sigma}_\epsilon$, $\widetilde{\varrho}$, ϱ, $\widetilde{\sigma}_z$, $\widetilde{\sigma}_\eta$ and σ_ϵ by minimum distance as in Gourieroux, Monfort, and Renault [1993]. Denote the vector of parameters to be estimated as Θ, the estimates are given by:

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \begin{pmatrix} \widehat{\mathcal{M}} - \frac{1}{N}\sum_n^N \widehat{\mathcal{M}}_T^n(\Theta) \\ \\ \mathcal{M}_{\text{beliefs}}(\Theta) - \mathcal{M}_{\text{model}}(\Theta) \end{pmatrix}' W \begin{pmatrix} \widehat{\mathcal{M}} - \frac{1}{N}\sum_n^N \widehat{\mathcal{M}}_T^n(\Theta) \\ \\ \mathcal{M}_{\text{beliefs}}(\Theta) - \mathcal{M}_{\text{model}}(\Theta) \end{pmatrix}. \tag{44}$$

In this minimization problem, we minimize the weighted norm of a vector composed of two sets of elements. The first set corresponds to fitting the data. $\widehat{\mathcal{M}}$ is the vector of moments to match –i.e. the 20 periods impulse response functions estimated from the bivariate VAR in section 2.1. $\widehat{\mathcal{M}}_T^n(\Theta)$ is computed as the average estimated IRF from N simulated random sample of the model with parameter Θ. T is the length for the simulated random sample, which is set to be the same as our actual sample 1969-2023. The second set corresponds to the requirement for consistency of beliefs. $\mathcal{M}_{\text{beliefs}}(\Theta)$ gathers the variance, first and second order auto-covariances of inflation as computed from the perceived law of motion, while $\mathcal{M}_{\text{model}}(\Theta)$ gathers the same moments as generated by the model. In our baseline estimation, we use the identity matrix as the weighting matrix W. The parameters obtained from the New Keynesian Phillips Curve and the minimum distance estimation are reported in Table 4:

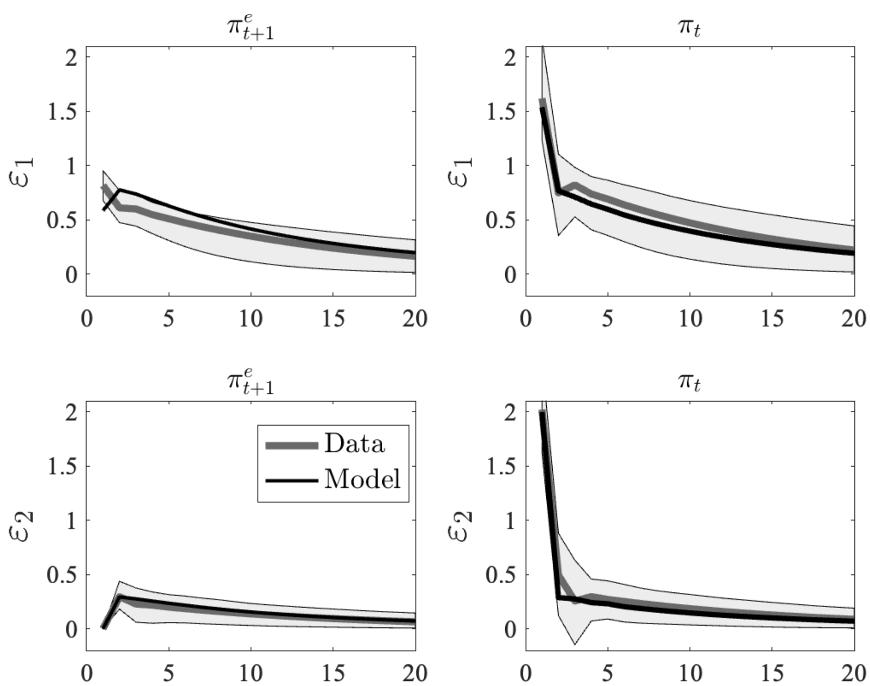
Table 4: Parameters

Phillips Curve Parameters			
β	0.99	ρ	0.89
σ_v	3.02	γ_g	0.0138
σ_η	2.24		
Estimated Parameters (Minimum Distance)			
$\widetilde{\sigma}_z$	0.73	σ_ϵ	5.88
$\widetilde{\rho}_z$	0.96	$\widetilde{\sigma}_\eta$	2.22
ϱ	0.43	$\widetilde{\sigma}_\epsilon$	2.24
$\widetilde{\varrho}$	0.96		

Notes: The Phillips curve estimates use the baseline estimation with Hazell, Herreño, Nakamura, and Steinsson's [2022] estimate of γ_g. σ_η is implied by the variance of the residual e from the Phillips curve.

Figure 10 shows how the estimated model matches well the empirical impulse responses we obtained in section 2.[29] In the signal-extraction problem, economic agents form expectations using the perceived law of motion and their subjective parameters. The large $\widetilde{\sigma}_z$ and $\widetilde{\rho}_z$ make the agents believe the disaggregated inflation series are very informative about the common component, $\widetilde{z}_t$. However, in reality, most of the variations in the disaggregate signals they observe are actually coming from the broad-base supply shock ϵ_t, as σ_ϵ point estimate is relatively large. As a result, they respond to the broad-base supply shock more aggressively. For this reason, this model with imperfect information and bounded rationality can trigger larger movements of expectations than those with rational expectations, and full or imperfect information. Meanwhile, their beliefs of such high $\widetilde{\sigma}_z$ and $\widetilde{\rho}_z$ are supported by the realized aggregate inflation that they observe, as we have imposed that consistency of beliefs restriction in our estimation.

Figure 10: Impulse Responses from Data and Estimated Model



Notes: The thick gray line is the IRF from the 2-VAR with actual data; the black line is the average IRF from simulated data across 200 random samples. The estimated VAR is a VAR(2) with Cholesky decomposition, ordering π^e_{t+1} first. The shaded area represents 95% confidence interval.

[29] In Appendix G, we also include the IRFs using actual and simulated data from our baseline model with an alternative orders of variables for the Choleski decomposition. Because the Choleski VAR is here used as a way to summarize the joint dynamics of inflation and inflation expectations, we would expect the data generated by a correct model to give similar results as the actual data when summarized with a different ordering of the variables. Appendix G shows that this is indeed the case.

From the perspective of the model, the VAR shock ε_1 corresponds to combination of demand (v) and broad-base (ϵ) shocks. Because the actual variation in v_t is quite small, the shape of the IRF mainly reflects the impact of the broad-based shock. The other VAR shock, ε_2, corresponds to the residual supply shocks η, that is orthogonal to the broad-based shock ϵ. To better illustrate this point, we can construct an orthogonalized normal random variable w_t:

$$w_t = \eta_t - \frac{cov(\epsilon_t, \eta_t)}{\sigma_\epsilon^2}\epsilon_t = \eta_t - \frac{\varrho\sigma_\eta}{\sigma_\epsilon}\epsilon_t. \tag{45}$$

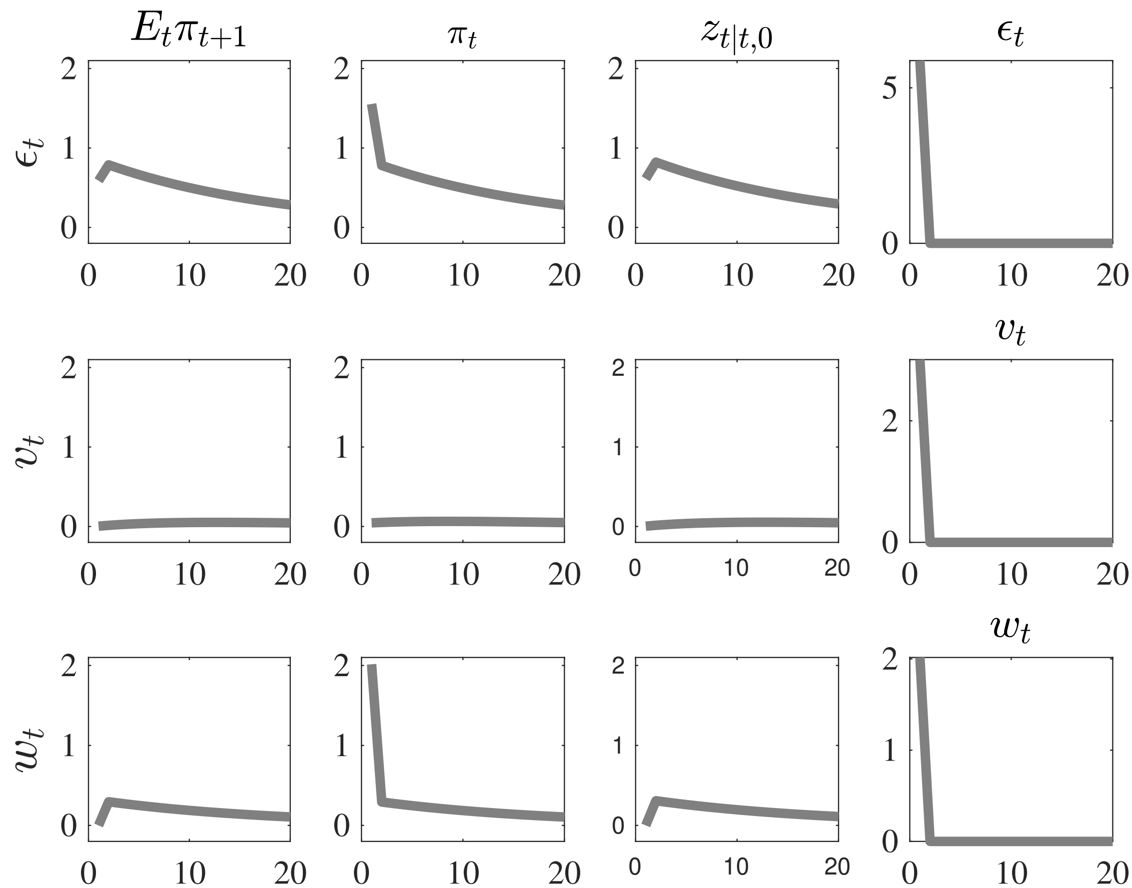
By construction, w_t is uncorrelated to ϵ_t. This constructed "shock" represents the supply shock that is not used to form the within period expectations. Since agents observe aggregate inflation only after having formed the expectations that they report, this shock affects measured inflation on impact but reported inflation expectations only with a lag. The VAR shock ε_2 then corresponds to the residual supply shock w_t.

In light of this, Figure 11 shows the implied model impulse responses to the three structural shocks: the broad-based supply shock ϵ_t, the demand shock v_t, and the residual supply shock w_t. We see why the broad-based supply shock ϵ_t plays a dominant role in shaping the response to ε_1 in Figure 10. This is because the variation in inflation created by v_t is minimal due to the flatness of Phillips Curve. Meanwhile, the response to ε_2 corresponds to a residual supply shock w_t. Moreover, the broad-based shock ϵ_t itself only increases inflation on impact by a factor ϱ. But the broad-based shock triggers a large increase of inflation expectations, which push up aggregate inflation. In comparison, the residual supply shock w_t increases aggregate inflation substantially on impact, but does not affect expectations right away. As a result, the response of aggregate inflation is more transitory.

It is worth noting that the three main parameters of the perceived law of motion, $\widetilde{\rho}_z$, $\widetilde{\sigma}_v$, and $\widetilde{\sigma}_\epsilon$, will be directly reflected in the process for inflation. For example, given the perceived law of motion, $\frac{cov(\pi_t,\pi_{t-2})}{cov(\pi_t,\pi_{t-1})}$ gives an estimate of $\widetilde{\rho}_z$. Similarly, $\widetilde{\sigma}_v$ could then be inferred using $\widetilde{\rho}_z$ and $cov(\pi_t, \pi_{t-1})$. Accordingly, if these moments for actual inflation were far from those implied by the agent's subjective model and if agents were learning, they would realize the discrepancy between their subjective model and reality, and will adjust their beliefs accordingly. Table 5 therefore reports the first three moments of the inflation auto-covariance function generated by the model and those implied by the beliefs of agents. As we can see, they are very close implying that agents would have little reason to change their beliefs had we have allowed them to learn.

Finally, in our estimation procedure we did not directly use the gap as one observable and did not try to match the empirical tri-variate VAR $(\pi_t, \pi_{t+1}^e, \text{gap}_t)$ that we estimated in

Figure 11: Impulse Responses to Structural Shocks in the Estimated Model



Notes: These impulse responses to structural shocks are computed from the model solution (34) to (37) with estimated parameters.

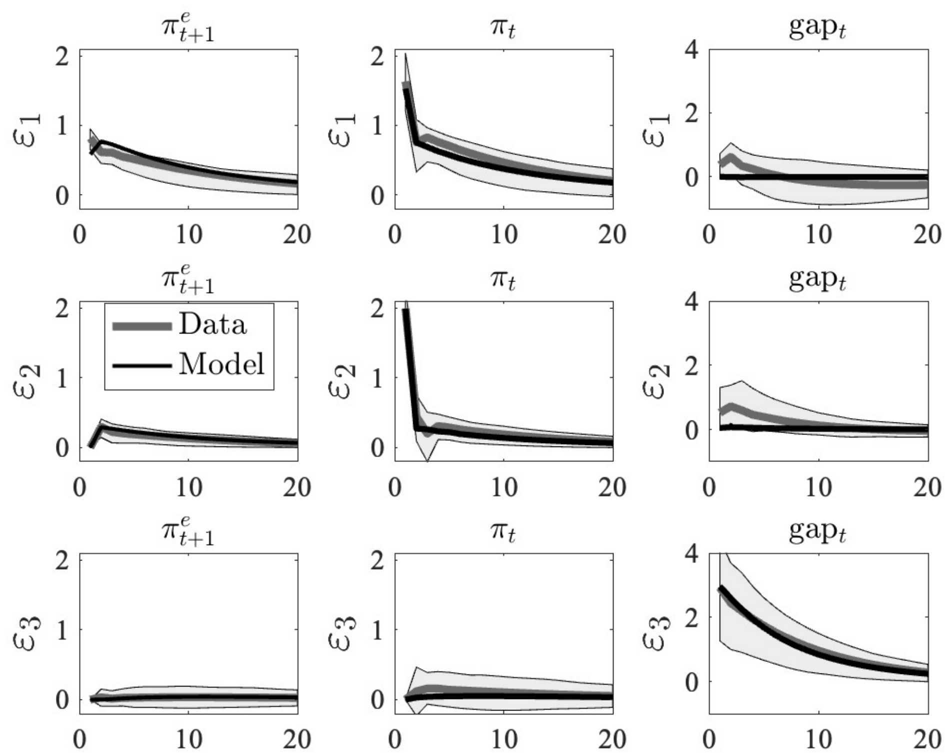
Table 5: Variance and Auto-covariance of Inflation

	Data	PLM	Model
$var(\pi_t)$	12.52	11.34	11.34
$cov(\pi_t, \pi_{t-1})$	7.70	6.18	6.28
$cov(\pi_t, \pi_{t-2})$	6.99	5.92	5.82

Notes: PLM stands for "Perceived Law of Motion". "The PLM and "Model" moments are the average moment across 200 random samples. In the estimations, we penalize distance between data and model 2-VAR responses as well as distance between these three moments in the Perceived Law of Motion and in the model.

Section 2. In figure 12 we compare these estimated impulse responses with those obtained from the same tri-variate VAR estimated with 200 random samples of our estimated model. The match is indeed very good. Note that the model responses of gap_t to the first two shocks ε_1 and ε_2 are zero by construction as gap_t is modelled as an exogenous shock.

Figure 12: Trivariate VAR IRF from data and model simulation



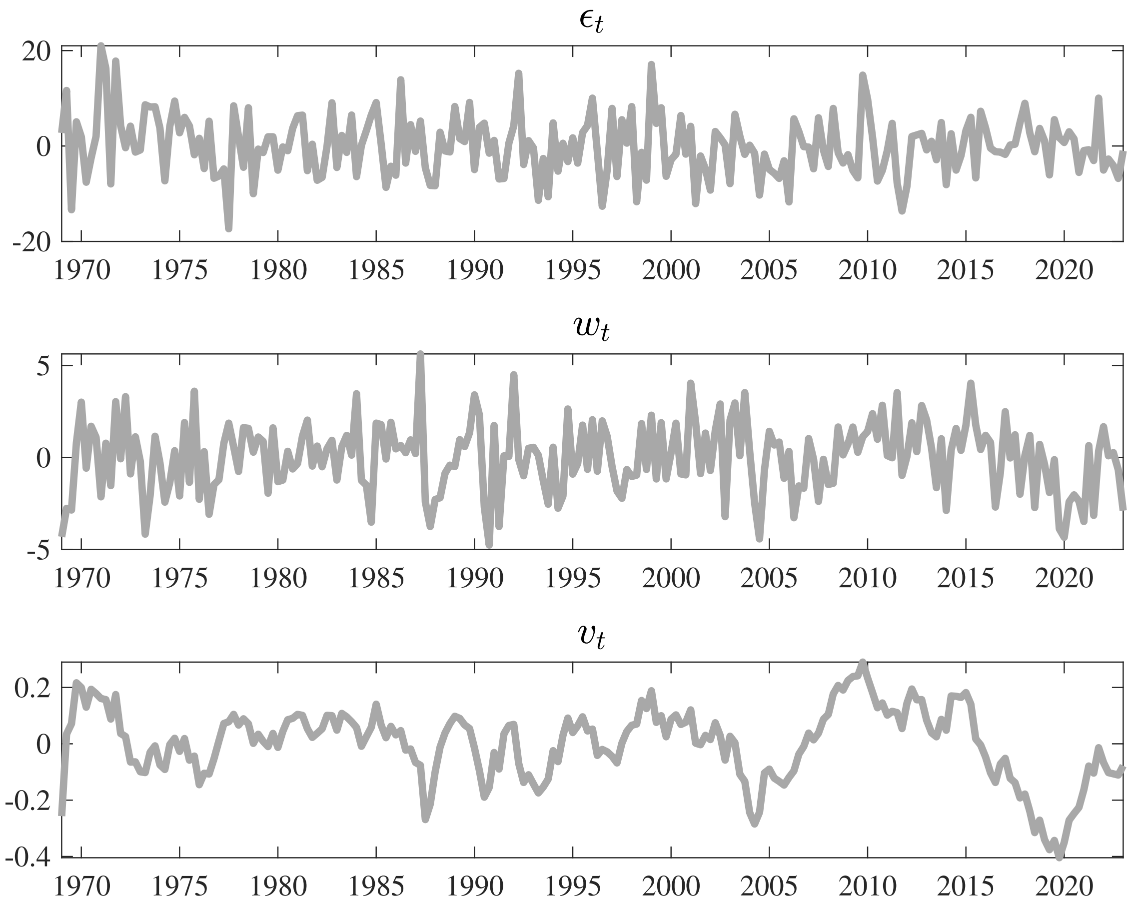
Notes: thick gray line is IRF from 3-VAR with actual data; black line is average IRF from simulated data across 200 random samples. The shaded area represents the 95% confidence interval.

4.4 Filtered Shocks and Counterfactuals

Using the state-space representation of our estimated model, we can use Kalman filter smoothing to form projections of the latent states and the shocks implied by the model. The latent variable $\widetilde{z}_{t+1|t,1}$ corresponds to the common component $\mathcal{C}_t$ the agents perceived using the disaggregate signals. The three shocks to the economy are the demand shock v_t, the broad-based supply shock ϵ_t and the residual supply shock w_t. Shocks are displayed in Figure 13. Because the Phillips Curve is quite flat, the data on inflation and expected inflation do not depend much on the gap, and therefore contain little information on it. As a result, the recovered series on v_t is imprecise but mainly irrelevant. This can be partly seen from the fact that the variance of recovered v_t is merely 0.04 whereas the true variance of v_t should be 3.01.

On the contrary, the recovered series for ϵ_t and w_t are quite trustworthy because the observables contain lots of information on these two shocks. We illustrate these points using simulated data in Appendix H.

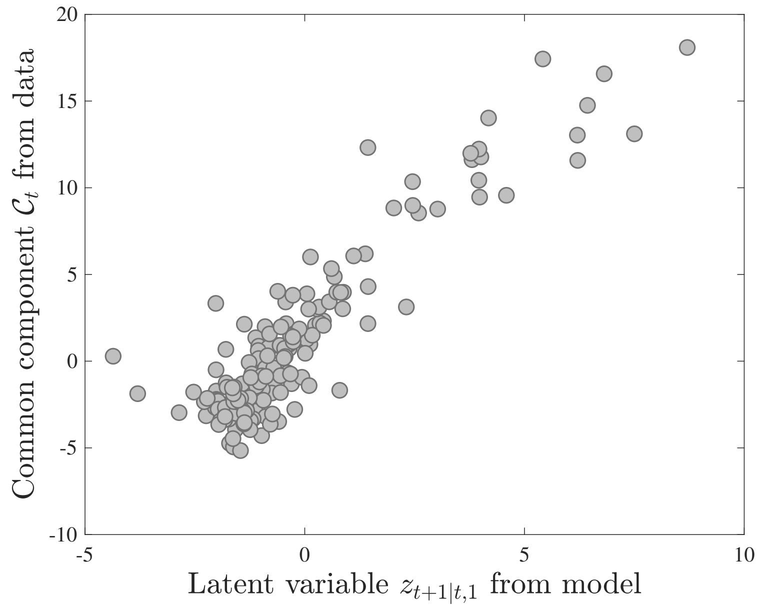
Figure 13: Filtered shocks from the estimated model



Notes: The demand shock v_t, the broad-based supply shock ϵ_t and the residual supply shock w_t are recovered by Kalman filter smoothing using the estimated model.

In our model, the system of inflation and expected inflation is driven by a latent variable $\widetilde{z}_{t+1|t,1}$, which is the common component the agents perceived using the disaggregate signals. This variable can be obtained using the Kalman Filter Smoothing approach mentioned above. In figure 14, we scatter the backed-out $\widetilde{z}_{t+1|t,1}$ against the common component $\mathcal{C}_t$ we extracted using the actual CPI disaggregated inflation series in Section 3. The correlation between these two series is 0.91, suggesting that the perceived common component in our model contents similar information to the common component directly extracted from disaggregated data, although only aggregate data are used in the model.

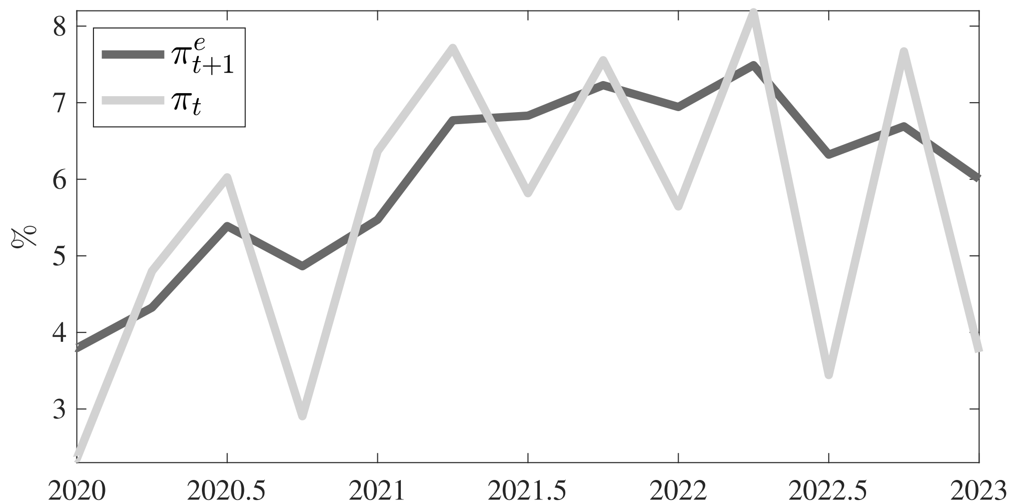
Figure 14: Filtered shocks from the estimated model



Notes: On this Figure, the latent variable $\tilde{z}_{t+1|t,1}$ is the "common component" the agents perceived using the disaggregate signals, as obtained from Kalman filter smoothing using the estimated model. It is plotted agains the common component $\mathcal{C}_t$ extracted from the CPI components in Section 3.

We then perform a counterfactual simulation starting in 2020. We simulate the economy with the broad-based shock ϵ_t. The resulting paths for Headline inflation and inflation expectations gives the contribution of that shock to realized Headline inflation and inflation expectations . Figure 15 shows these counterfactual paths. Inflation is much more volatile than expectations, because in the model it is a filtered version of the broad-based shock that determines expectations. For example, in 2022Q1, the broad-based shock puts Headline inflation at 8.2%, whereas it would have been 4.1% absent of shocks and was actually 10.6%. More than half of the deviation of Headline inflation from its unconditional mean is therefore attributable to the broad-based shock. In the same quarter, inflation expectations are 7.5% with the broad-based shock, instead of 4.7% unconditionally and 7.9% in the data. Virtually all the deviation of expectations from their unconditional mean is explained by the broad-based supply shock.

Figure 15: Counterfactual Simulations with Broad-Based Shock Only



Notes: the plain lines correspond to counterfactual simulations of the estimated model, when the economy is hit only by the broad-based shock ϵ_t for both expected inflation and actual inflation. Note that in the model the unconditional mean of inflation is 4.1%.

5 Conclusion

This paper began by presenting evidence in support of the pre-pandemic view that the Phillips curve is quasi-flat and that supply shock are mainly temporary. We also presented VAR evidence indicating that inflation expectations and one-step ahead inflation move very closely together, suggesting that either agents are very good at predicting inflation or that inflation expectations are potentially an important driver of inflation. After failing to explain the VAR patterns using rational expectations, we proposed an explanation to the observed inflation dynamics based on bounded rationality.

In our bounded rational framework, persistent inflation arises when the economy is simultaneously hit by supply shocks across many sectors, even if the shocks themselves are not persistent. When agents see prices rise in many sectors, they revise their view about the common force that drives inflation. This revision leads to higher expected inflation, which makes the initially temporary disturbances propagate through time. In contrast, when agents see prices rises in only a small set of sectors, they recognize that this is not an aggregate phenomena and therefore do not revise their expectations much. This results in sector specific shocks causing only minor movements in inflation. Both these features are needed to match the joint dynamics of inflation and inflation expectations as captured by a bivariate VAR. In support of our model's main assumption, we showed that the common component extracted from a large set of disaggregate data correlates closely with measured inflation expectations. We also showed that the this common component extracted from disaggregated data correlates closely with the model counterpart inferred only from aggregate data. Together, we believe these elements support a theory of persistent bouts of inflation driven primary by temporary but broad based supply shocks that get transmitted to expectations.

References

Ball, L. M., D. Leigh, and P. Mishra (2022): "Understanding U.S. Inflation During the COVID Era," Working Paper 30613, National Bureau of Economic Research.

Barnichon, R., and G. Mesters (2020): "Identifying Modern Macro Equations with Old Shocks," *The Quarterly Journal of Economics*, 135(4), 2255–2298.

Beaudry, P., T. J. Carter, and A. Lahiri (2023): "The Central Bank's Dilemma: Look Through Supply Shocks or Control Inflation Expectations?," Working Paper 31741, National Bureau of Economic Research.

Beaudry, P., C. Hou, and F. Portier (2023): "Monetary Policy when the Phillips Curve is Quite Flat," *American Economic Journal: Macroeconomics (forthcoming)*.

Benigno, P., and G. B. Eggertsson (2023): "It's Baaack: The Surge in Inflation in the 2020s and the Return of the Non-Linear Phillips Curve," Working Paper 31197, National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110(9), 2748–82.

Coibion, O., and Y. Gorodnichenko (2015a): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105(8), 2644–78.

——— (2015b): "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, 7(1), 197–232.

Coibion, O., Y. Gorodnichenko, and R. Kamdar (2018): "The Formation of Expectations, Inflation, and the Phillips Curve," *Journal of Economic Literature*, 56(4), 1447–91.

D'Acunto, F., U. Malmendier, J. Ospina, and M. Weber (2020): "Exposure to Grocery Prices and Inflation Expectations," *Journal of Political Economy*, 129(5), 1615–1639.

Gabaix, X. (2014): " A Sparsity-Based Model of Bounded Rationality *," *The Quarterly Journal of Economics*, 129(4), 1661–1710.

Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.

Hazell, J., J. Herreño, E. Nakamura, and J. Steinsson (2022): "The Slope of the Phillips Curve: Evidence from U.S. States," *The Quarterly Journal of Economics*, 137(3), 1299–1344.

Lansing, K. (2009): "Time Varying U.S. Inflation Dynamics and the New Keynesian Phillips Curve," *Review of Economic Dynamics*, 12(2), 304–326.

Lorenzoni, G. (2009): "A Theory of Demand Shocks," *American Economic Review*, 99(5), 2050–84.

Michaillat, P., and E. Saez (2022): "$u^{\star} = \sqrt{uv}$," NBER Working Papers 30211, National Bureau of Economic Research, Inc.

Reis, R. (2023): "Four Mistakes in the Use of Measures of Expected Inflation," *AEA Papers and Proceedings*, 113, 47–51.

Romer, C. D., and D. H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94(4), 1055–1084.

Sims, C. A. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50(3), 665 – 690, Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

Stock, J. H., and M. W. Watson (2011): "35 Dynamic Factor Models," in *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

Wieland, J., and M.-J. Yang (2020): "Financial Dampening," *Journal of Money, Credit and Banking*, 52(1), 79–113.

Appendix to
The Dominant Role of Expectations and Broad-Based Supply Shocks in Driving Inflation

Paul Beaudry, Chenyu Hou and Franck Portier

January 2024
version 1.8

A Relation to Benigno and Eggertsson [2023]

In this appendix, we repeat the analysis of section 1 when we use $\log(V/U)$ as a measure of labour market tightness and either the Survey of Professional Forecasters (SPF) or the Michigan Survey of Consumers (MSC) as a measure of inflation expectations. We first stay with our baseline sample 1969-2007 where we use the expected GDP Deflator from the Professional Forecaster as our measure of expected inflation. We then move to the sample used in Benigno and Eggertsson [2023] (but stopping in 2007 to perform our out-of-sample analysis).

A.1 Sample 1969-2007

Table A.1: Estimated Phillips Curves using $\log(V/U)$ as a Measure of the Gap, 1969-2007

	Using SPF	Using MSC
β	1.26*	0.99*
	(0.047)	(0.048)
γ_g	0.28*	0.10
	(0.062)	(0.069)
γ_r	-0.42*	0.23*
	(0.098)	(0.040)
Observations	144	144
J Test	10.538	10.700
(jp)	(0.987)	(0.986)
Weak ID Test	29.469	10.201

Notes: this table reports estimates of the augmented Phillips curve (2). All results are using IV-GMM procedure, Newey-West HAC standard errors with six lags are reported in parentheses. The constant term is omitted from the table. The measure of inflation is Core CPI and the gap is measured with $\log(V/U)$ *. All regressors are instrumented using six lags of Romer and Romer's [2004] shocks (as extended by Wieland and Yang [2020]) and their squares as instruments. A* $\star$ *indicates significance at 5%. Sample is 1969Q1-2007Q4.*

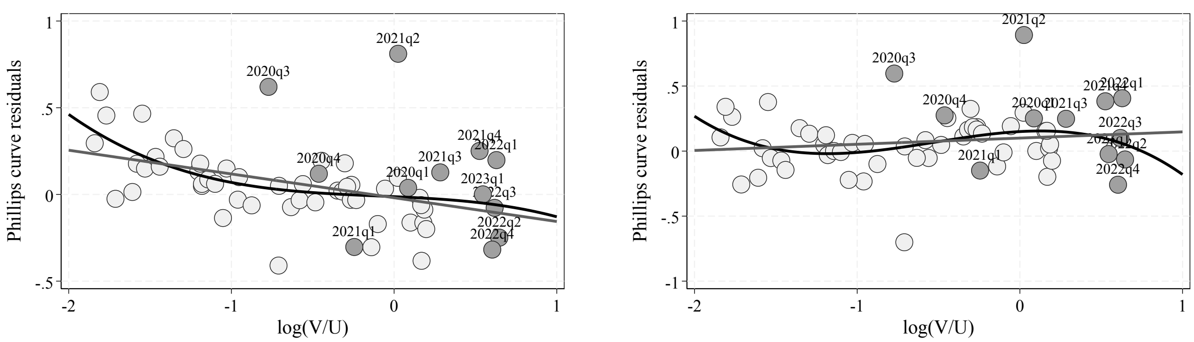
Table A.2: Projection of the Philips Curve Residuals ϵ on the Gap $\log(V/U)$, 2008-2023

	Using SPF		Using MSC	
	linear	nonlinear	linear	nonlinear
$\log(V/U)$	-0.14*	-0.04	0.05	0.07
	(0.038)	(0.076)	(0.042)	(0.083)
$\log(V/U)^2$		-0.02		-0.25
		(0.149)		(0.162)
$\log(V/U)^3$		-0.06		-0.15*
		(0.082)		(0.089)
N	60	60	60	60
σ_ϵ 60-07	.302		.291	
σ_ϵ 08-23	.232		.232	

Notes: the Phillips curve residuals are obtained from the augmented Phillips curve (2) estimated over the sample 1969Q1-2007Q4, using $\log(V/U)$ *as a measure of the gap and either Survey of Professional Forecasters or Michigan Survey of Consumers as a measure of expectations. The standard errors of estimated coefficients are between parentheses.*

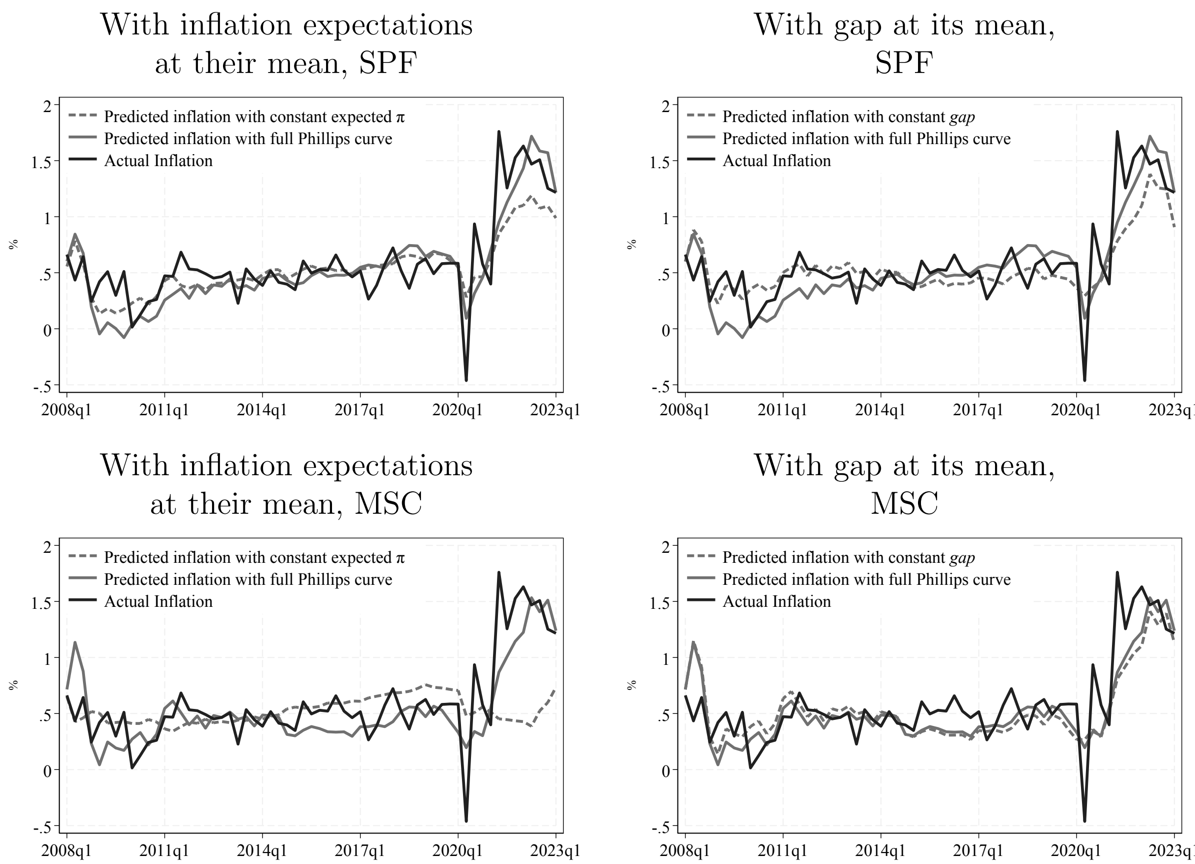
Figure A.1: Out-of-Sample Residuals from Phillips Curve, using $\log(V/U)$ as a Measure of the Gap

(a) Using Survey of Professional Forecasters (b) Using Michigan Survey of Consumers



Notes: Panels (a) and (b) of this figure plots the out-of-sample residuals of the estimated Phillips curve (2) against $\log(V/U)$ *as the measure of labor market tightness, for two measures of inflation expectations. The gray lines show the estimated linear or cubic relation between residuals and labor market tightness (see Table A.2 for the estimated coefficients). Light dots correspond to pre-2020 observations and dark ones to post-2020. We exclude 2020q2 from this graph.*

Figure A.2: Counterfactual Simulations from the Estimated Phillips Curve



Notes: These counterfactual simulations are done using the estimated Phillips curve (2), using $\log(V/U)$ *as the measure of labor market tightness.*

A.2 Sample 1981-2007

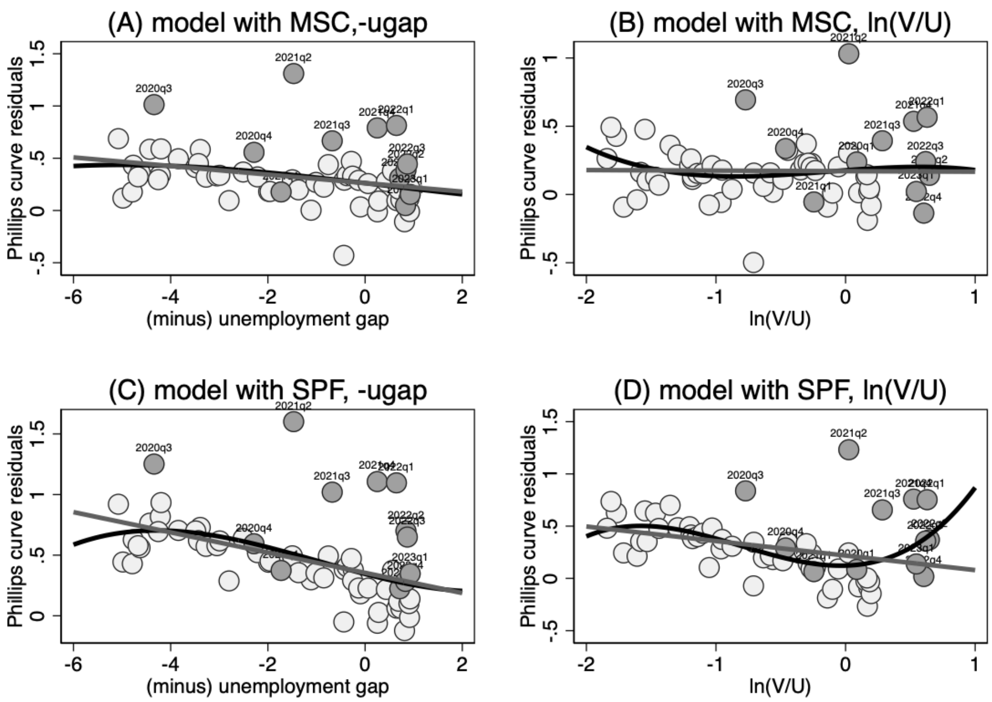
We then move to use the expected CPI series from the Survey of Professional Forecasters. Because that series starts only after 1981 quarter 3, we use the sample 1981q3-2007q4 for our analysis. In Table A.3 we present the estimation results of our Phillips Curve. For comparison, we also include the estimates using *minus* unemployment gap as measure of labor market tightness in this shorter sample.

Table A.3: Estimated Phillips Curves using $\log(V/U)$ as a Measure of the Gap, 1981-2007

Expectations:	Using MSC		Using SPF	
Labor Market Tightness:	(1) -ugap	(2) $\ln V/U$	(3) -ugap	(4) $\ln V/U$
β	0.90*	0.90*	0.97*	0.95*
	(0.033)	(0.033)	(0.048)	(0.045)
γ_g	0.07*	0.17*	0.10*	0.29*
	(0.019)	(0.071)	(0.017)	(0.070)
γ_r	0.35*	0.29*	0.33*	0.26*
	(0.052)	(0.049)	(0.065)	(0.067)
Observations	106	106	106	106
J Test	10.687	10.763	10.952	10.872
(jp)	(0.986)	(0.985)	(0.984)	(0.984)
Weak ID Test	163.460	60.640	121.279	89.040

Notes: this table reports estimates of the augmented Phillips curve (2). All results are using IV-GMM procedure, Newey-West HAC standard errors with six lags are reported in parentheses. The constant term is omitted from the table. The measure of inflation is Core CPI and the gap is measured with $\log(V/U)$ *. All regressors are instrumented using six lags of Romer and Romer's [2004] shocks (as extended by Wieland and Yang [2020]) and their squares as instruments. A $\star$ indicates significance at 5%. Sample is 1981Q3-2007Q4.*

Figure A.3: Out-of-Sample Residuals from Phillips Curve, 1981-2007



Notes: Panels (a) and (c) of this figure plots the out-of-sample residuals of the estimated Phillips curve (2) against minus *unemployment gap as the measure of labor market tightness, for two measures of inflation expectations. Panels (b) and (d) of this figure plots the out-of-sample residuals against* $\log(V/U)$ *as the measure of labor market tightness, for two measures of inflation expectations. The gray lines show the estimated linear or cubic relation between residuals and labor market tightness (see Table A.4 for the estimated coefficients). Light dots correspond to pre-2020 observations and dark ones to post-2020. We exclude 2020q2 from this graph.*

Table A.4: Projection of the Philips Curve Residuals ϵ on measures of gap, 2008-2023

Expectations:	MSC				SPF			
Gap:	-ugap		$\ln V/U$		-ugap		$\ln V/U$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
gap_t	-0.04*	-0.05	-0.00	0.07	-0.08*	-0.10	-0.14*	0.04
	(0.017)	(0.057)	(0.042)	(0.084)	(0.019)	(0.065)	(0.048)	(0.091)
gap_t^2		-0.00		-0.02		0.01		0.50*
		(0.044)		(0.165)		(0.050)		(0.179)
gap_t^3		0.00		-0.05		0.00		0.21*
		(0.007)		(0.090)		(0.008)		(0.098)
Observations	60	60	60	60	60	60	60	60

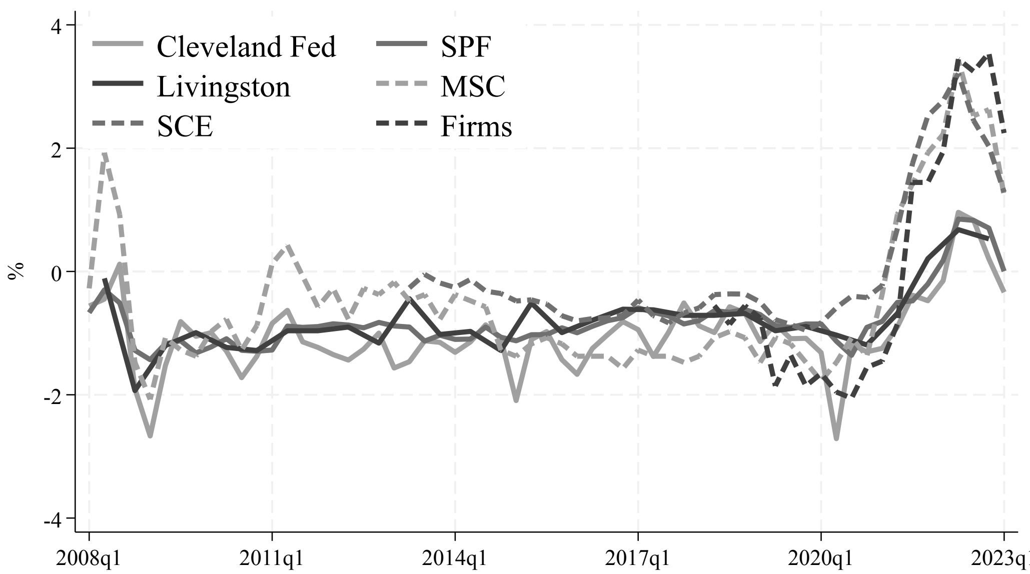
Notes: the Phillips curve residuals are obtained from the augmented Phillips curve (2) estimated over the sample 1981Q3-2007Q4, using $-ugap$ or $\log(V/U)$ as a measure of the gap and either Survey of Professional Forecasters or Michigan Survey of Consumers as a measure of expectations. The standard errors of estimated coefficients are between parentheses.

From the above analysis, we see only one specification suggests the Phillips Curve appears to become steeper in recent years. That is when using $\ln V/U$ as a measure of labor market tightness, SPF as the measure of expected inflation and sample 1981Q3-2007Q4. Once we use the Michigan Survey of Consumers as the measure of expected inflation, such a pattern disappears. This is not surprising given the differences between the expected inflations from the survey of professionals and central banks, and those from the households and firms. In Appendix B we show the differences between various surveys of expectations and repeat our analysis with them.

B Baseline Phillips Curve with Various Measures of Inflation Expectations

We first show the differences between various surveys of expectations. Figure B.1 depicts the demeaned series of expected inflation from different surveys. A stark pattern is that households and firms have similar expectations of inflation that are drastically different from those of professional and central banks. In particular, the households and firms expect much higher inflation after 2020.

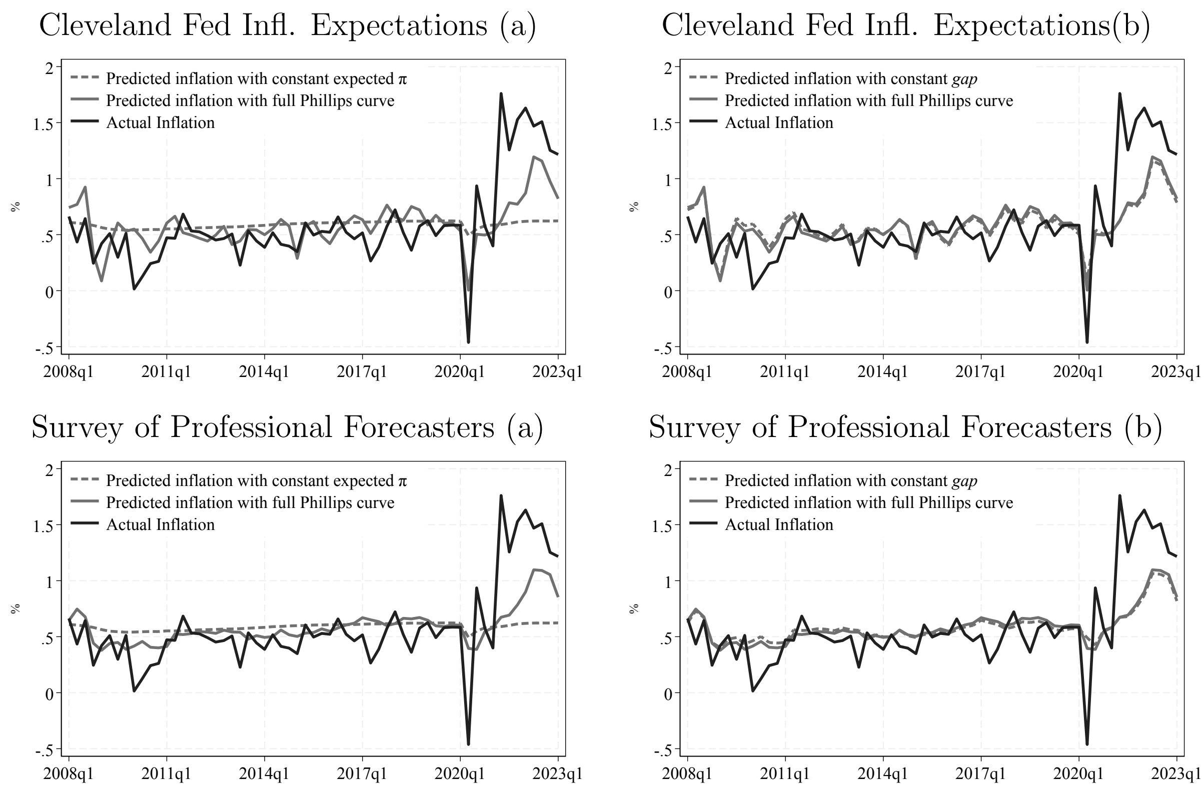
Figure B.1: Different measures of inflation expectations



Notes: Demeaned average expected inflation from various surveys. Solid lines are from the surveys of professionals and central banks. The dashed lines are from surveys of households and firms.

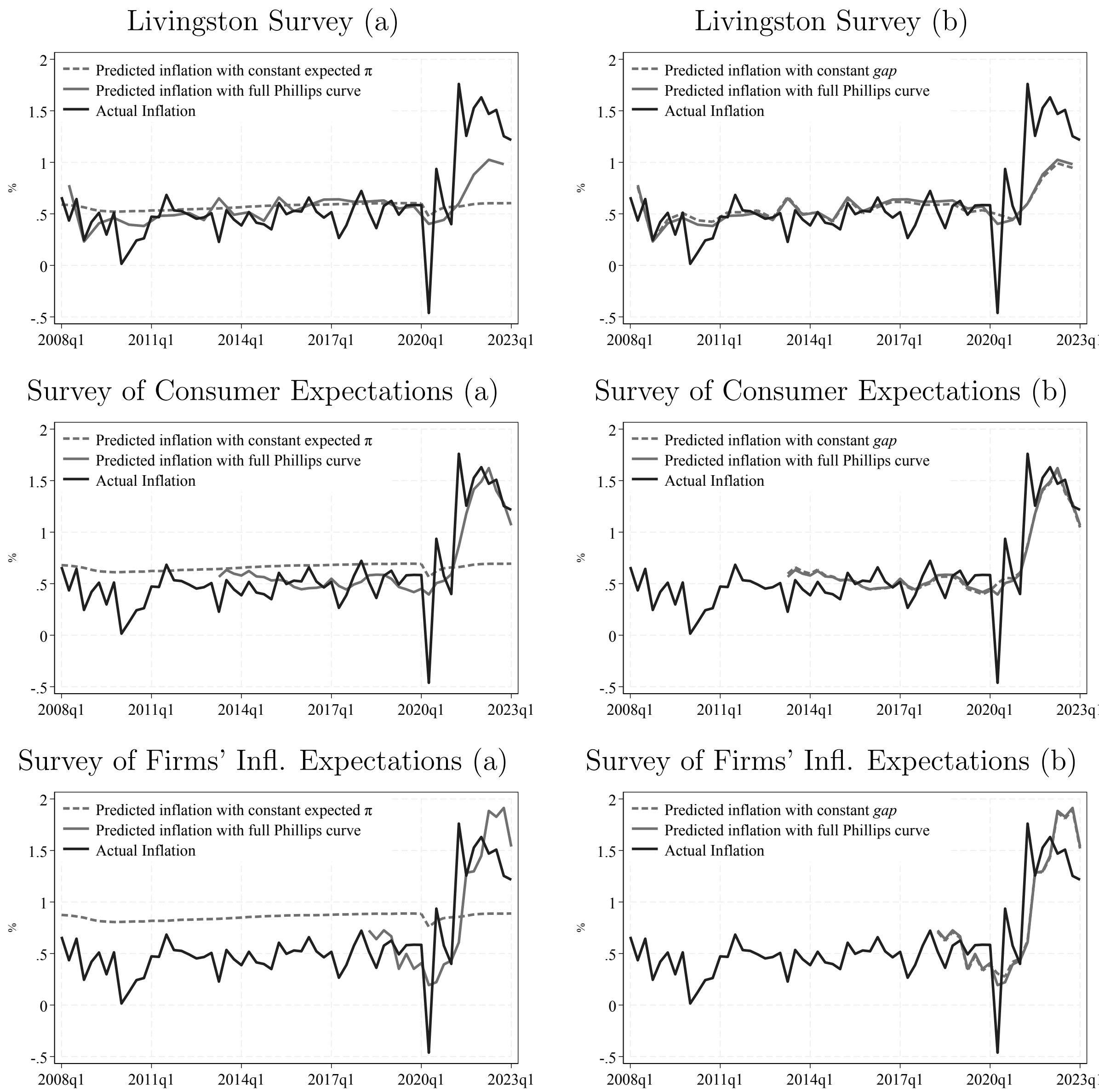
Here we repeat the exercise of Section 1 with various measures of expectations. Key result is that the run-up of inflation expectations can account for inflation increase post-COVID-19 when households or firms survey are used (Survey of Consumer Expectations (New York Fed) or Survey of Firms' Inflation Expectations (Cleveland Fed) or Michigan Survey of Consumers in the main text). The professional forecasts (Cleveland Fed Inflation Expectations, Survey of Professional Forecasters (Philadelphia Fed) or Livingston Survey (Philadelphia Fed)) increase less post-COVID-19, and fail to contribute fully to the increase in actual headline inflation. Note the the baseline Phillips curve (1) is used to do the counterfactual simulations.

Figure B.2: Counterfactual Simulations from the Estimated Phillips Curve, Using Various Measure of Inflation Expectations (a) With inflation expectations at their mean, (b) With gap at its mean



Notes: These counterfactual simulations are done using the baseline Phillips curve (1), using either Cleveland Fed Inflation Expectations or the Survey of Professional Forecasters (Philadelphia Fed) as a measure of expectations.

Figure B.3: Counterfactual Simulations from the Estimated Phillips Curve, Using Various Measure of Inflation Expectations (a) With inflation expectations at their mean, (b) With gap at its mean



Notes: These counterfactual simulations are done using the baseline Phillips curve (1), using either Livingston Survey (Philadelphia Fed), Survey of Consumer Expectations (New York Fed) or Survey of Firms' Inflation Expectations (Cleveland Fed) as a measure of expectations.

Finally, we do a "horse-race" regression to estimate our augmented Phillips curve (2) including both Survey of Professional Forecasters and Michigan Survey of Consumers as potential measures of expected inflation. The results are reported in Table B.1. In column (1) we use *minus* unemployment gap as the measure of the gap and freely estimate it. In column (2) we fix the slope of the negative unemployment gap at the estimate of Hazell, Herreño, Nakamura, and Steinsson [2022] and in column (3) we use the $\ln V/U$ as the measure of the gap. The results suggest that the Michigan Survey of Consumers is the more relevant measure for expected inflation in the Phillips Curve.[30]

Table B.1: Horse Race using SPF and MSC

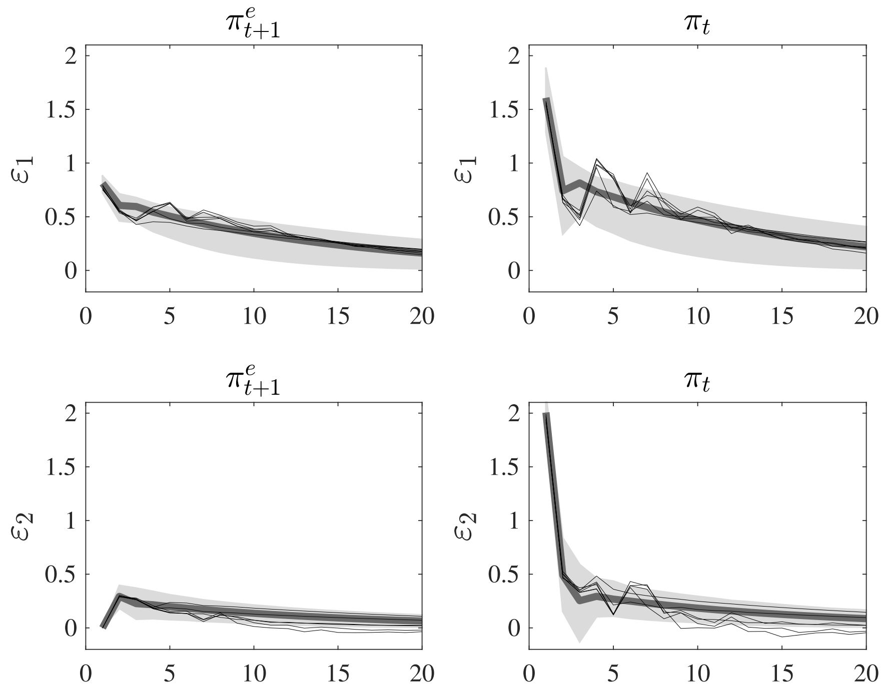
Gap	-ugap or $\ln V/U$		
π	Sample 1981q3-2007q4		
	(1) -ugap free slope	(2) -ugap fixed slope	(3) $\ln V/U$
MSC	0.40*	0.62*	0.47*
	(0.177)	(0.161)	(0.173)
SPF	0.29	-0.02	0.25
	(0.272)	(0.233)	(0.267)
γ_g	0.08*	0.0138	0.19*
	(0.019)	(-)	(0.071)
γ_r	0.30*	0.28*	0.24*
	(0.082)	(0.054)	(0.080)
Observations	106	106	106
J Test	10.655	11.939	10.636
(jp)	(0.979)	(0.971)	(0.980)
Weak ID Test	135.594	121.041	39.712

Notes: this table reports estimates of the augmented Phillips curve (2). All results are using IV-GMM procedure, Newey-West HAC standard errors with six lags are reported in parentheses. The constant term is omitted from the table. The measure of inflation is Core CPI and the gap is measured with minus *unemployment gap or* $\log(V/U)$ *. All regressors are instrumented using six lags of Romer and Romer's [2004] shocks (as extended by Wieland and Yang [2020]) and their squares as instruments. A* $\star$ *indicates significance at 5%. Sample is 1981Q3-2007Q4.*

[30] These findings are in line with the results from Coibion, Gorodnichenko, and Kamdar [2018]. In Coibion and Gorodnichenko [2015b] the authors find that households' expectations from the Michigan Survey help to explain the missing disinflation puzzle around the 2008 Financial Crisis.

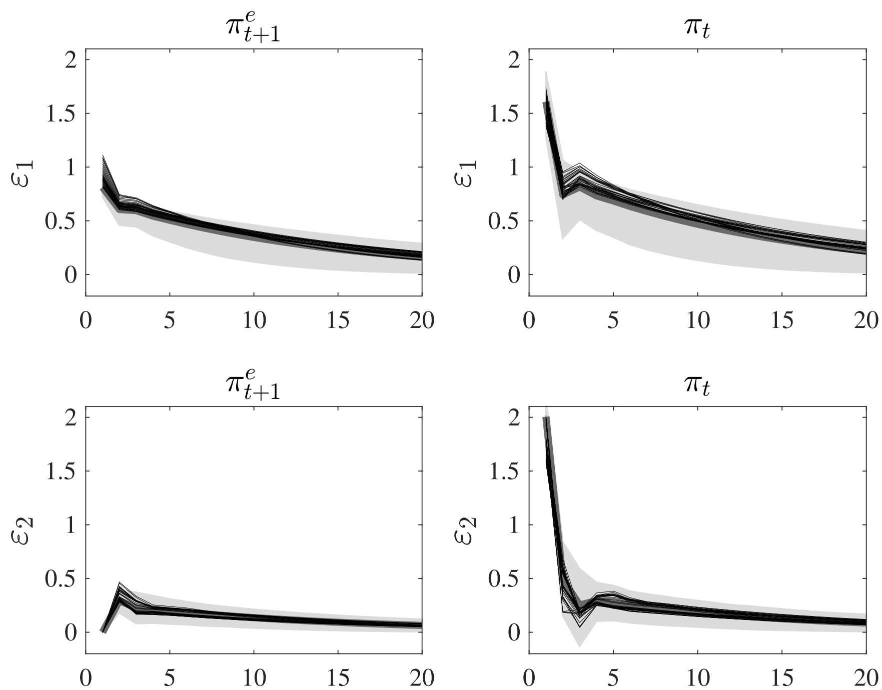
C Robustness of the 2-VAR to Lags and Sample

Figure C.1: Impulse Responses in the 2-VAR (π_t, π^e_{t+1}), with 2 to 8 Lags



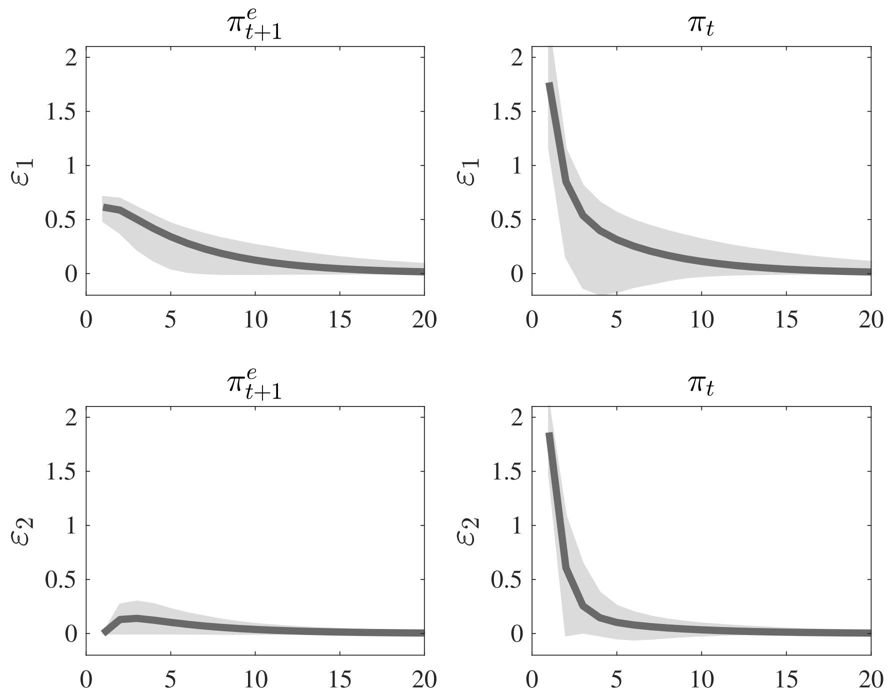
Notes: this Figure plots the impulse responses to a one standard deviation shocks ε_1 and ε_2. These shocks are obtained from a Choleski orthogonalization. The estimated VAR with two lags of Headline CPI inflations and the Michigan Survey of Consumers inflation expectations are in gray. Sample is 1969Q1-2023Q1. Shaded area represents the 95% confidence band. The black lines corresponds to an estimation with 3 to 8 lags.

Figure C.2: Impulse Responses in the 2-VAR (π_t, π^e_{t+1}), Starting with First 20 Years and Adding Years One by One



Notes: this Figure plots the impulse responses to a one standard deviation shocks ε_1 and ε_2. These shocks are obtained from a Choleski orthogonalization. The estimated VAR with two lags of Headline CPI inflations and the Michigan Survey of Consumers inflation expectations are in gray. Sample is 1969Q1-2023Q1. Shaded area represents the 95% confidence band. The black lines corresponds to an estimation samples 1969Q1-1989Q1, 1969Q1-1990Q1, 1969Q1-1991Q1,... etc.

Figure C.3: Impulse Responses in the 2-VAR (π_t, π^e_{t+1}) estimated over 2008–2023



Notes: this Figure plots the impulse responses to a one standard deviation shocks ε_1 and ε_2. These shocks are obtained from a Choleski orthogonalization. The estimated VAR with two lags of Headline CPI inflations and the Michigan Survey of Consumers inflation expectations are in gray. Sample is 2008Q1-2023Q1. Shaded area represents the 95% confidence band. The VAR is etsimated with one lag.

D Using Disaggregated Prices

Table D.1 display the 25 expenditures categories we use, as obtained from BLS and used to compute the CPI. Figure D.1 plots sectoral inflations and Table D.2 displays the estimated parameters α_i, σ_i and ρ when we estimate the model

$$\pi_{i,t} = \alpha_i \, \mathcal{C}_t + e_{i,t},$$

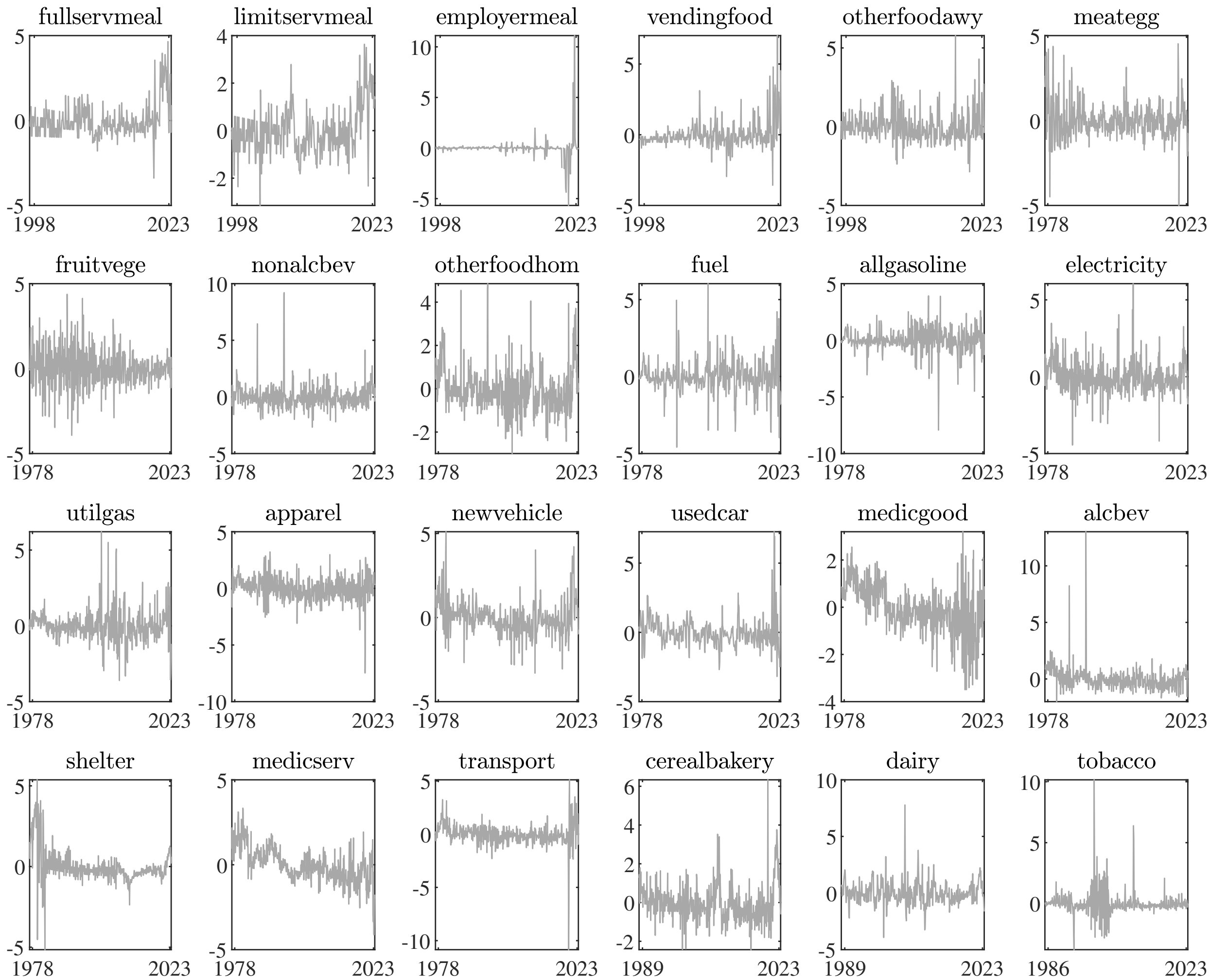
$$\mathcal{C}_t = \rho_{\mathcal{C}} \, \mathcal{C}_{t-1} + v_t.$$

Table D.1: Summary of Categories

Categories and items	Available sample	Categories and items	Available sample
Food at Home		**Food away from Home**	
Cereals and bakery products	1990m1-2023m5	Full service meals and snacks	1999m1-2023m5
Meats, poultry, fish, and eggs	1978m1-2023m5	Limited service meals and snacks	1998m12-2023m5
Dairy and related products	1990m1-2023m5	Food at employee sites and schools	1999m1-2023m5
Fruits and vegetables	1978m1-2023m5	Food from vending machines and mobile vendors	1998m12-2023m5
Nonalcoholic beverages and beverage materials	1978m1-2023m5	Other food away from home	1999m1-2023m5
Other food at home	1978m1-2023m5		
Energy commodities		**Energy services**	
Fuel oil	1978m1-2023m5	Electricity	1978m1-2023m5
Motor fuel (gasoline)	1978m1-2023m5	Utility gas	1978m1-2023m5
less food and energy commodity		**less energy services**	
Apparel	1978m1-2023m5	Shelter	1978m1-2023m5
New vehicles	1978m1-2023m5	Medical care service	1978m1-2023m5
Used cars and trucks	1978m1-2023m5	Transportation service	1978m1-2023m5
Medical care commodities	1978m1-2023m5		
Alcoholic beverages	1978m1-2023m5		
Tobacco and smoking products	1987m1-2023m5		

Notes: the main sample starts from 1978m1 due to the availability of the monthly Michigan Survey of Consumers. The disaggregated data are level 3 categories from BLS.

Figure D.1: Inflation for the Components in the CPI Basket



Notes: this Figure displays the time series of the 25 expenditures categories we use, as obtained from BLS and used to compute the CPI.

Table D.2: Estimated parameters

Parameter	Estimate		Parameter	Estimate	
ρ	0.98				
	α_i	σ_i		α_i	σ_i
fullservmeal	0.19	0.73	limitservmeal	0.15	0.84
employermeal	0.05	0.99	vending	0.09	0.94
otherfoodawy	0.06	0.98	meategg	0.04	0.98
fruitvege	0.02	0.99	nonalcohol	0.05	0.97
otherfoodhome	0.10	0.86	fuel	0.03	0.99
gasoline	0.02	1.00	electricity	0.08	0.93
utility gas	0.03	0.99	apparel	0.06	0.96
new vehicle	0.09	0.89	used car	0.03	0.99
medical good	0.09	0.90	alcoholbev	0.08	0.92
shelter	0.13	0.75	medical service	0.11	0.85
transportation	0.08	0.92	cereal bakery	0.16	0.83
dairy product	0.08	0.96	tobacco	0.01	1.00

Notes: we normalize $\sigma_v = 1$, and we input series that are normalized (demeaned and standard deviations normalized to 1) because the price series have different volatilities. But the normalization doesn't qualitatively change the results.

E Average Signal from Disaggregate Price Indices

We show that the signal-extraction problem with multiple disaggregate price indices is equivalent to one with an average signal across these disaggregate indices. The disaggregate signals the agent faces are given by (23). Without loss of generality, consider the m disaggregate signals the agent uses to form expectations:

$$X_t \equiv \begin{pmatrix} \pi_{1,t} \\ \pi_{2,t} \\ \cdots \\ \pi_{m,t} \end{pmatrix} = \ell_m \widetilde{z}_t + \begin{pmatrix} \widetilde{e}_{1,t} \\ \widetilde{e}_{2,t} \\ \cdots \\ \widetilde{e}_{m,t} \end{pmatrix}, \tag{E.1}$$

with $\widetilde{e}_{j,t} \sim N(0, \widetilde{\sigma}_j^2)$, and where we denote the vector of signals as X_t and ℓ_m is an $m \times 1$ vector of ones. Denote the prior of $\widetilde{z}_t$ in $t-1$ as:

$$\widetilde{z}_t^{prior} \sim N(\widetilde{z}_{t|t-1}, \sigma^2), \tag{E.2}$$

where $\widetilde{z}_{t|t-1}$ denotes the prior mean and σ^2 the stationary prior variances. The posterior mean of the nowcast is:

$$\widetilde{z}_{t|t} = \widetilde{z}_{t|t-1} + \kappa_z(X_t - \ell_m \widetilde{z}_{t|t-1}), \tag{E.3}$$

where κ_z is the Kalman Gain:

$$\kappa_z = \sigma^2 \ell_m' (\sigma^2 \ell_m \ell_m + V)^{-1}, \tag{E.4}$$

where V is a diagonal matrix with entries $\widetilde{\sigma}_j^2$ on the main diagonal. The Kalman Gain can then be explicitly written as a function of σ^2 and $\widetilde{\sigma}_j^2$'s:

$$\begin{aligned}
\kappa_z &= \sigma^2 \ell_m' \left(V^{-1} - \frac{V^{-1}\sigma^2 \ell_m \ell_m' V^{-1}}{1 + \sigma^2 \ell_m' V \ell_m} \right) \\
&= \sigma^2 \left[\begin{pmatrix} \frac{1}{\widetilde{\sigma}_1^2} & \frac{1}{\widetilde{\sigma}_2^2} & \cdots & \frac{1}{\widetilde{\sigma}_m^2} \end{pmatrix} - \begin{pmatrix} \frac{1}{\widetilde{\sigma}_1^2} \sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2} & \frac{1}{\widetilde{\sigma}_2^2} \sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2} & \cdots & \frac{1}{\widetilde{\sigma}_m^2} \sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2} \end{pmatrix} \frac{1}{\frac{1}{\sigma^2} + \sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2}} \right] \\
&= \begin{pmatrix} \frac{1}{\widetilde{\sigma}_1^2} & \frac{1}{\widetilde{\sigma}_2^2} & \cdots & \frac{1}{\widetilde{\sigma}_m^2} \end{pmatrix} \frac{1}{\frac{1}{\sigma^2} + \sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2}}.
\end{aligned} \tag{E.5}$$

Now consider an average signal:

$$x_t = \widetilde{z}_t + \underbrace{\frac{\sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2} \widetilde{e}_{j,t}}{\sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2}}}_{\equiv \epsilon_t}, \tag{E.6}$$

$$\epsilon_t \sim N(0, \underbrace{\frac{1}{\sum_{j=1}^m \frac{1}{\widetilde{\sigma}_j^2}}}_{\equiv \sigma_\epsilon^2}). \tag{E.7}$$

With the same prior, the Kalman Gain is given by:

$$\hat{\kappa}_z = \frac{\frac{1}{\sigma_\epsilon^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_\epsilon^2}}. \tag{E.8}$$

Using (E.5) and (E.8), it is then straightforward to show that the posterior mean $\widetilde{z}_{t|t}$ is the same when the agent uses disaggregate signals (E.1) with the aggregate signal (E.6).

Moreover, the stationary posterior variances are the same as well. As a result, the stationary posterior belief formed by observing multiple signals (E.1) is equivalent to one formed using an average signal (E.6).

Now consider the actual disaggregate signals from (19) and the perceived disaggregate signals from (23). Apply our results above, the average signals are given by:

$$s_t = \sum_{j \in S} \underbrace{\frac{1/\widetilde{\sigma}_j^2}{\sum_{j \in S} 1/\widetilde{\sigma}_j^2}}_{\equiv \chi_j} \pi_{j,t} = \beta \pi_{t+1}^e + \gamma_g \text{gap}_t + \underbrace{\sum_{j \in S} \chi_j e_{j,t}}_{\equiv \epsilon_t} \qquad \text{(Average Signal ALM)}$$

$$s_t = \sum_{j \in S} \underbrace{\frac{1/\widetilde{\sigma}_j^2}{\sum_{j \in S} 1/\widetilde{\sigma}_j^2}}_{\equiv \chi_j} \pi_{j,t} = \widetilde{z}_t + \underbrace{\sum_{j \in S} \chi_j \widetilde{e}_{j,t}}_{\equiv \widetilde{\epsilon}_t} \qquad \text{(Average Signal PLM)}$$

These give the first equations in (27) and (26), where

$$\sigma_\epsilon^2 = \sum_{j \in S} \chi_j^2 \sigma_j^2, \quad \widetilde{\sigma}_\epsilon^2 = \sum_{j \in S} \chi_j^2 \widetilde{\sigma}_j^2 = \frac{1}{\sum_{j \in S} \frac{1}{\widetilde{\sigma}_j^2}} \qquad \text{(E.9)}$$

Given the actual and perceived aggregate supply shocks $\eta_t = \sum_{j=1}^{N} \omega_j e_{j,t}$ and $\widetilde{\eta}_t = \sum_{j=1}^{N} \omega_j \widetilde{e}_{j,t}$, we can also derive the actual and perceived correlation between ϵ_t and η_t ($\widetilde{\epsilon}_t$ and $\widetilde{\eta}_t$):

$$\varrho \equiv corr(\epsilon_t, \eta_t) = \frac{\sum_{j \in S} \chi_j \omega_j \sigma_j^2}{\sigma_\epsilon \sigma_\eta}, \quad \widetilde{\varrho} \equiv corr(\widetilde{\epsilon}_t, \widetilde{\eta}_t) = \frac{\sum_{j \in S} \chi_j \omega_j \widetilde{\sigma}_j^2}{\widetilde{\sigma}_\epsilon \widetilde{\sigma}_\eta} \qquad \text{(E.10)}$$

F Models with Rational Expectations

F.1 Incomplete Information Rational Expectations

Take the IIRE model described by (9)-(12). The agent forms expectations about gap_t. Denoting the nowcasts of which as $\text{gap}_{t|t}$, expectations er given by:

$$\begin{aligned} E_t[\pi_{t+1}] &= E_t[\beta E_{t+1}\pi_{t+2} + \gamma_g \text{gap}_{t+1} + e_{t+1}] \\ &= \frac{\gamma_g \rho}{1-\beta\rho}\text{gap}_{t|t}. \end{aligned} \tag{F.1}$$

The nowcast $\text{gap}_{t|t}$ is given by:[31]

$$\text{gap}_{t|t} = \text{gap}_{t|t-1} + k_i(s_t - \gamma_g \text{gap}_{t|t-1}), \tag{F.2}$$

$$\mathcal{K} = \sigma_k^2 \gamma_g (\gamma_g^2 \sigma_k^2 + \sigma_\epsilon^2)^{-1}, \tag{F.3}$$

$$\sigma_k^2 = \rho^2(\sigma_k^2 - \mathcal{K}\gamma_g \sigma_k^2) + \sigma_v^2, \tag{F.4}$$

where $\mathcal{K}$ is the stationary Kalman Gain and σ_k^2 is the stationary posterior variance. This leads to equations (13) and (14) in the main text. Note that as $\sigma_\epsilon \to 0$, $\mathcal{K} \to 1/\gamma_g$ and $w_t \to e_t$. The IIRE case converges to the FIRE case.

F.2 Hybrid Phillips Curve or Adaptive Expectations

One related question is whether a hybrid Phillips Curve or adaptive expectations can help to explain the joint dynamics between expected and actual inflation. We consider a Phillips Curve taking the following hybrid form:

$$\pi_t = \beta \underbrace{\left(\tau\pi_{t-1} + (1-\tau)E_t^{FIRE}\pi_{t+1}\right)}_{\text{observed in MSC}} + \gamma_g \text{gap}_t + e_t, \tag{F.5}$$

$$\text{gap}_t = \rho \text{gap}_{t-1} + v_t, \tag{F.6}$$

[31]Note in (F.3) the agent uses correct σ_ϵ because when the agent is rational, she can easily back-out the correct ρ, $\gamma_g \sigma_v$ and σ_ϵ with the variance-covariance structure of s_t. As a result, the agent's information will not support any subjective $\widetilde{\sigma}_\epsilon \neq \sigma_\epsilon$.

where τ can represents the indexation, the fraction of people using adaptive expectations, or motivated by k-level thinking as in Beaudry, Carter, and Lahiri [2023]. The expectations formed under FIRE takes into account that there are backward looking component in inflation. It is easy to show that inflation takes the following form using undetermined coefficients:

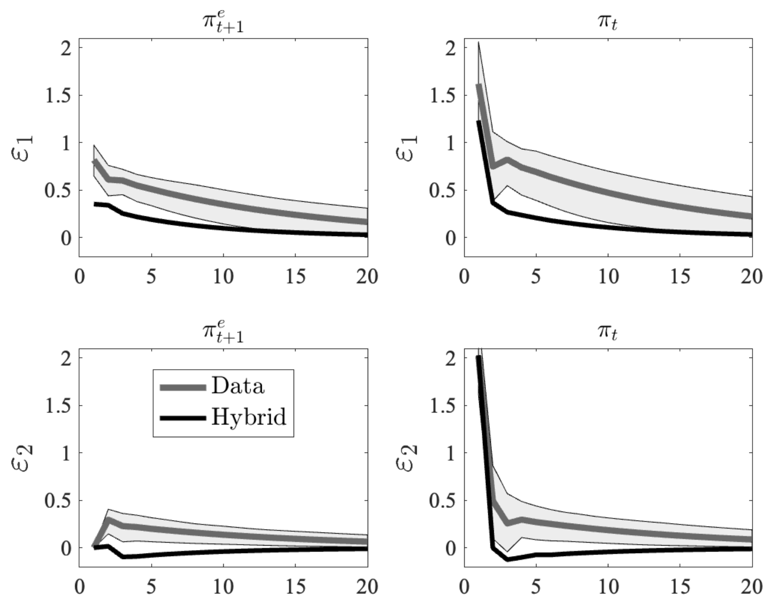
$$\pi_t = a\pi_{t-1} + b\text{gap}_t + ce_t, \tag{F.7}$$

with

$$\begin{cases} a &= \frac{\beta\tau}{1-a\beta(1-\tau)}, \\ b &= \frac{\beta(1-\tau)b\rho+\gamma_g}{1-a\beta(1-\tau)}, \\ c &= \frac{1}{1-a\beta(1-\tau)}. \end{cases}$$

Following our approach in section 2.2, we set β, ρ, σ_v and γ_g at values consistent with our Phillips Curve, and we estimate the free parameter τ to match the IRFs from the bivariate VAR(2). Our estimate is $\hat{\tau} = 0.05$, and the above model cannot match the empirical IRFs well:

Figure F.1: The Joint Process of π and π^e_{t+1} in the Data and Under Models with Rational Expectations

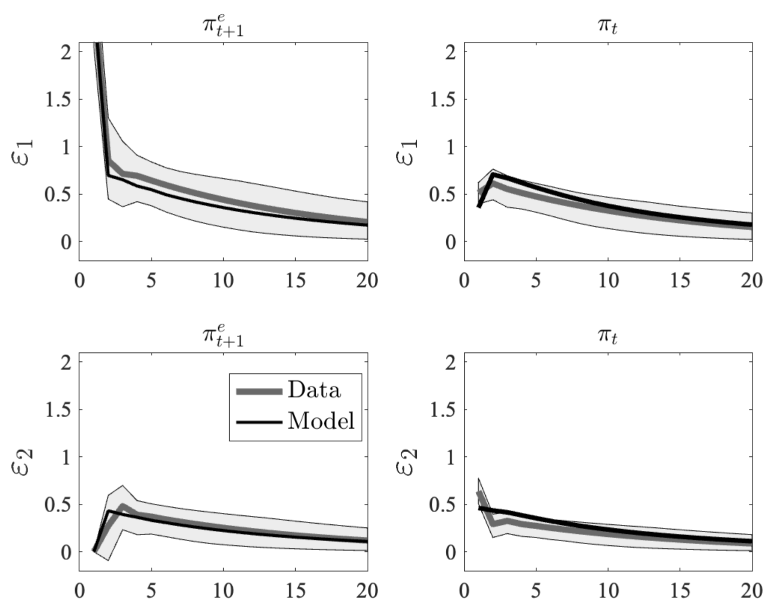


Notes: on this Figure, the grey line plots the impulse responses to a one standard deviation shocks ε_1 and ε_2 estimated with data from the Michigan Survey of Consumers and Headline CPI. Sample is 1969Q1-2023Q1. Shaded area represents the 95% confidence band. The black line plot the average impulse responses (over 200 simulations of length 216) obtained from the same VAR estimated on simulated data, when the Data Generating Process is the estimated hybrid model (F.5).

G Alternative Order of Cholesky VAR

We present the IRFs from Cholesky VAR(2) where we order headline inflation first and expected inflation the second. This change of ordering reflects an alternative identification assumption that the first underlying shock affects π_t and π^e_{t+1} simultaneously and the second shock affects π^e_{t+1} on impact then π_t with a lag. Note our Cholesky VAR is just method to summarize joint dynamics of inflation and expectations. If our model is close to the true data generating process for inflation and expected inflation, we would expect IRF from this VAR with alternative (but possibly incorrect) ordering to have similar result. Figure G.1 shows the results from this exercise. Our model matches very well with the VAR using alternative ordering.

Figure G.1: IRF from data and model simulation



Notes: thick gray line is IRF from bi-VAR with actual data; marked black line is average IRF from simulated data across 200 random samples. The estimated VAR is VAR(2) with Cholesky Decomposition ordering π_t - π^e_{t+1}. The shaded area represents 95% CI.

H State-space Representation of Model

H.1 Model

For easy exposition, we derive the state-space representation of our model with the orthogonalized residual supply shock w_t defined in section 4.3. To save notations, we define

$$\omega = \varrho \frac{\sigma_\eta}{\sigma_\epsilon} \tag{H.1}$$

According to (45), w_t is given by:

$$w_t = \eta_t - \omega \epsilon_t \tag{H.2}$$

The two observables $\pi_{t+1|t,0}$ and π_t can be written as system of equations of latent states $\widetilde{z}_{t|t-1,1}$ and gap_t.

$$\begin{aligned}
\pi_{t+1|t,0} &= \widetilde{\rho}_z \widetilde{z}_{t|t,0} = \widetilde{\rho}_z((1-K)\widetilde{z}_{t|t-1,1} + K s_t) \\
&= \widetilde{\rho}_z((1-K)\widetilde{z}_{t|t-1,1} + K(\beta \pi_{t+1|t,0} + \gamma_g \text{gap}_t + \epsilon_t)) \\
&= \frac{\widetilde{\rho}_z(1-K)}{1-K\beta\widetilde{\rho}_z} \widetilde{z}_{t|t-1,1} + \frac{\widetilde{\rho}_z K \gamma_g}{1-K\beta\widetilde{\rho}_z} \text{gap}_t + \frac{\widetilde{\rho}_z K}{1-K\beta\widetilde{\rho}_z} \epsilon_t.
\end{aligned} \tag{H.3}$$

From (27):

$$\begin{aligned}
s_t &= \pi_{t+1|t,0} + \gamma_g \text{gap}_t + \epsilon_t \\
&= \frac{\beta\widetilde{\rho}_z(1-K)}{1-K\beta\widetilde{\rho}_z} \widetilde{z}_{t|t-1,1} + \frac{\gamma_g}{1-K\beta\widetilde{\rho}_z} \text{gap}_t + \frac{1}{1-K\beta\widetilde{\rho}_z} \epsilon_t.
\end{aligned} \tag{H.4}$$

Inflation is given by:

$$\begin{aligned}
\pi_t &= \beta \pi_{t+1|t,0} + \gamma_g \text{gap}_t + \omega \epsilon_t + w_t \\
&= \frac{\beta\widetilde{\rho}_z(1-K)}{1-K\beta\widetilde{\rho}_z} \widetilde{z}_{t|t-1,1} + \frac{\gamma_g}{1-K\beta\widetilde{\rho}_z} \text{gap}_t + \left(\frac{K\beta\widetilde{\rho}_z}{1-K\beta\widetilde{\rho}_z} + \omega \right) \epsilon_t + w_t.
\end{aligned} \tag{H.5}$$

From (33) and expression of s_t, we get recursion of $\widetilde{z}_{t|t-1,1}$:

$$\begin{aligned}\widetilde{z}_{t+1|t,1} &= \widetilde{\rho}_z \widetilde{z}_{t|t,1} = \widetilde{\rho}_z(1-K-k)\widetilde{z}_{t|t-1,1} + \widetilde{\rho}_z K s_t + \widetilde{\rho}_z k \pi_t \\ &= \widetilde{\rho}_z \left(1-(K+k)\frac{1-\beta\widetilde{\rho}_z}{1-K\beta\widetilde{\rho}_z}\right)\widetilde{z}_{t|t-1,1} + \widetilde{\rho}_z \frac{(K+k)\gamma_g}{1-K\beta\widetilde{\rho}_z}\mathrm{gap}_t \\ &+ \widetilde{\rho}_z \left(\frac{K+\widetilde{\rho}_z Kk\beta}{1-K\beta\widetilde{\rho}_z} + \omega k\right)\epsilon_t + \widetilde{\rho}_z k w_t. \end{aligned} \tag{H.6}$$

We could write the state-space representation as:

$$X_t \equiv \begin{pmatrix} \pi_{t+1|t,0} \\ \pi_t \\ \widetilde{z}_{t+1|t,1} \\ \mathrm{gap}_t \end{pmatrix} = F X_{t-1} + B \begin{pmatrix} \epsilon_t \\ w_t \\ v_t \end{pmatrix}, \tag{H.7}$$

where

$$F = \begin{pmatrix} 0 & 0 & \frac{\widetilde{\rho}_z(1-K)}{1-K\beta\widetilde{\rho}_z} & \rho\frac{\widetilde{\rho}_z K\gamma_g}{1-K\beta\widetilde{\rho}_z} \\ 0 & 0 & \frac{\beta\widetilde{\rho}_z(1-K)}{1-K\beta\widetilde{\rho}_z} & \rho\frac{\gamma_g}{1-K\beta\widetilde{\rho}_z} \\ 0 & 0 & \widetilde{\rho}_z\left(1-(K+k)\frac{1-\beta\widetilde{\rho}_z}{1-K\beta\widetilde{\rho}_z}\right) & \widetilde{\rho}_z\frac{(K+k)\gamma_g}{1-K\beta\widetilde{\rho}_z}\rho \\ 0 & 0 & 0 & \rho \end{pmatrix}, \tag{H.8}$$

$$B = \begin{pmatrix} \frac{\widetilde{\rho}_z K}{1-K\beta\widetilde{\rho}_z} & 0 & \frac{\widetilde{\rho}_z K\gamma_g}{1-K\beta\widetilde{\rho}_z} \\ \frac{K\beta\widetilde{\rho}_z}{1-K\beta\widetilde{\rho}_z} + \omega & 1 & \frac{\gamma_g}{1-K\beta\widetilde{\rho}_z} \\ \left(\frac{K+\widetilde{\rho}_z Kk\beta}{1-K\beta\widetilde{\rho}_z} + \omega k\right)\widetilde{\rho}_z & \widetilde{\rho}_z k & \widetilde{\rho}_z\frac{(K+k)\gamma_g}{1-K\beta\widetilde{\rho}_z} \\ 0 & 0 & 1 \end{pmatrix}. \tag{H.9}$$

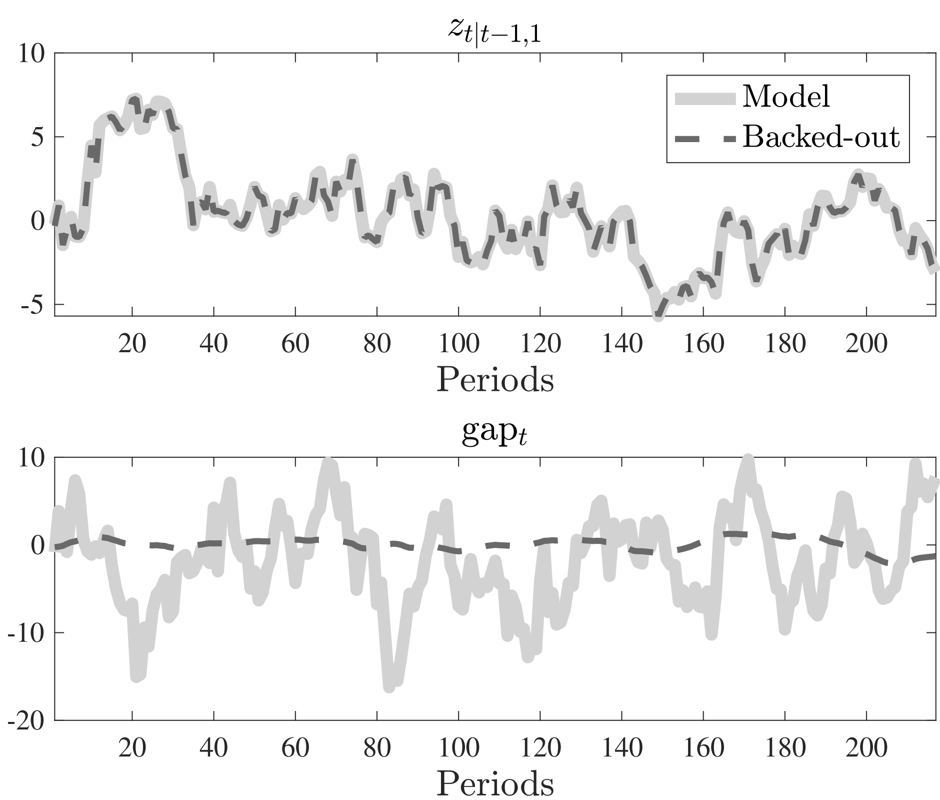
The observational equation is given by:

$$O_t = \begin{pmatrix} I_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \end{pmatrix} X_t. \tag{H.10}$$

We can then estimate the hidden states $\widetilde{z}_{t|t-1,1}$ and gap_t using Kalman Filter Smoothing, then get the corresponding shocks implied by the state-space representation.

The precision of the estimated states and shocks depend on the parameters in F and B. To illustrate the performance of this approach in the context of our estimated model, we simulate data using parameters from Table 4. We then plot the time series of the estimated states $E_t\pi_{t+1}$, π_t, $\widetilde{z}_{t|t-1,1}$ and gap_t in Figure H.1. The blue solid lines are the actual hidden states from simulated data and the red dash lines are the estimated ones from the above approach. Not surprisingly, the hidden state $\widetilde{z}_{t|t-1,1}$ is very well recovered where the estimated values are almost identical to the actual ones. Whereas the gap_t is very illy recovered. This is because the observables $\{E_t\pi_{t+1}, \pi_t\}$ contain a lot more information for $\widetilde{z}_{t|t-1,1}$ and almost no information about gap_t due to the fact γ_g is very small in the actual DGP. As a result, the Kalman Smoothing algorithm puts low weights on the observables when making predictions about this hidden state. Figure H.2 depicts the backed-out shocks and compares them with

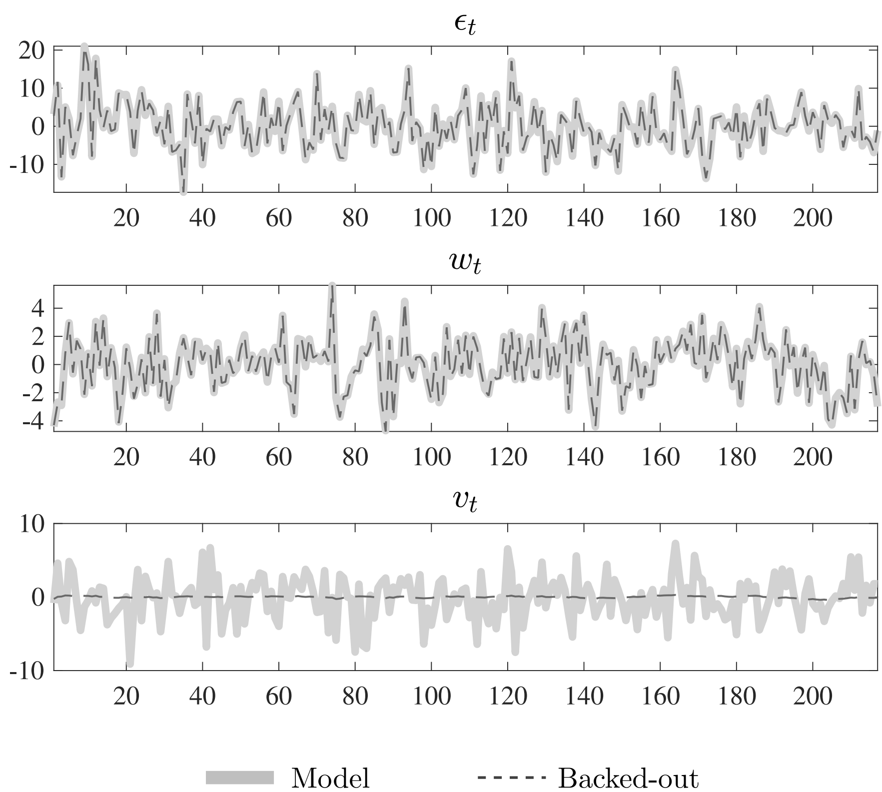
Figure H.1: Estimated v.s. actual latent states $\widetilde{z}_{t|t-1,1}$ and gap_t from simulated data



Notes: the blue solid lines are actual latent states from simulated data. The red dash lines are the recovered latent states using the Kalman Smoothing.

actual shock series simulated. For the same reason as the latent states, the observations are quite informative about the broad-based shocks ϵ_t and the aggregate shock w_t, but they are not informative about the common shock v_t. As a result, the recovered v_t series are much less volatile and not really comparable to the actual shock series.

Figure H.2: Estimated v.s. actual latent states $\widetilde{z}_{t|t-1,1}$ and gap_t from simulated data

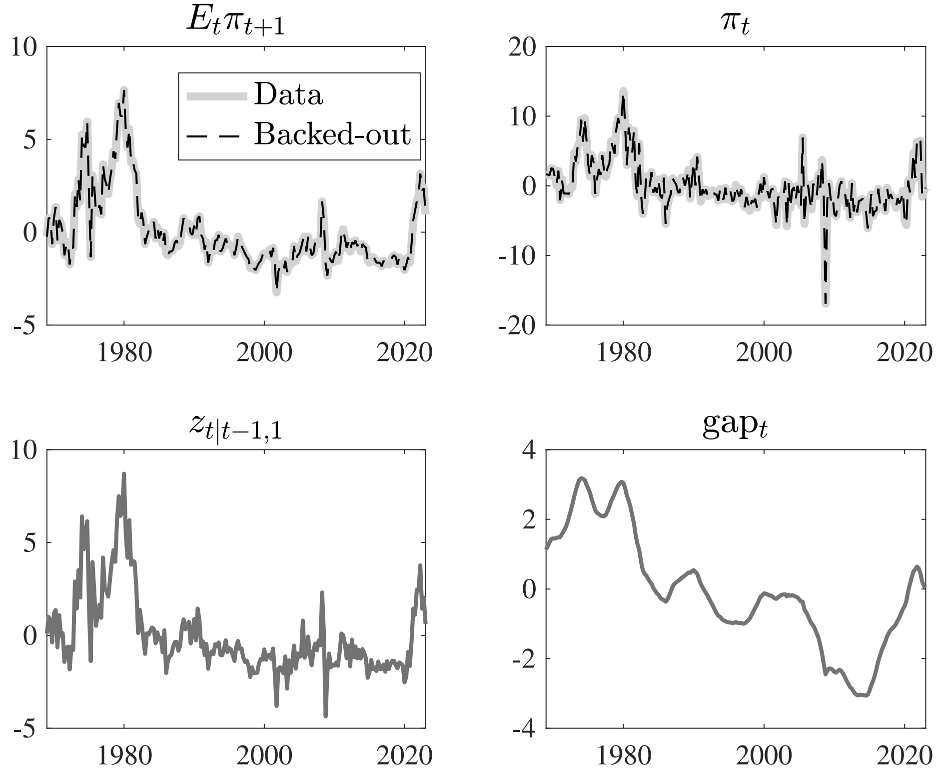


Notes: the dark gray lines are model simulated latent states. The light gray lines are the recovered latent states using the Kalman Smoothing.

H.2 Backed-out Series from Real Data

The figure H.3 depicts the recovered states $\{E_t\pi_{t+1}, \pi_t, \widetilde{z}_{t|t-1,1}, \text{gap}_t\}$ applying Kalman Smoothing. By construction, the expected inflation and aggregate inflation are observable so they coincide with each other. Because the actual values for $\widetilde{z}_{t|t-1,1}$ and gap_t are treated as unobserved, there are no corresponding actual values plotted for these two variables.

Figure H.3: Estimated v.s. actual states from actual data



Notes: the states π_t and $E_t\pi_{t+1}$ are observables. In the two upper panels, the thick grey line plots actual states and the black dashed line is the recovered latent states using the Kalman Smoothing. The two lines coincide by construction as these two states are observable. Because the actual values for $\widetilde{z}_{t|t-1,1}$ and gap_t are treated as unobserved, there are no corresponding actual values plotted for these two variables.

I The Baseline Model without the Timing Restriction

In this appendix, we reestimate the model under the assumption that inflation expectations are reported *after* observing current inflation, meaning that $\pi^e_{t+1} = \pi_{t+1|t,1}$. The estimated parameters become:

Table I.1: Estimated parameters

From the Phillips Curve			
β	0.99	ρ	0.89
σ_v	3.02	γ_g	0.0138
σ_η	2.24		
From Minimum Distance			
$\widetilde{\sigma}_z$	0.45	σ_ϵ	2.80
$\widetilde{\rho}_z$	0.98	$\widetilde{\sigma}_w$	2.46
ϱ	0.08	$\widetilde{\sigma}_\epsilon$	1.58
$\widetilde{\varrho}$	0.00		

Notes: The Phillips curve estimates use the baseline estimation with Hazell, Herreño, Nakamura, and Steinsson's [2022] estimate of γ_g. σ_η is implied by the variance of the residual e from the Phillips curve.

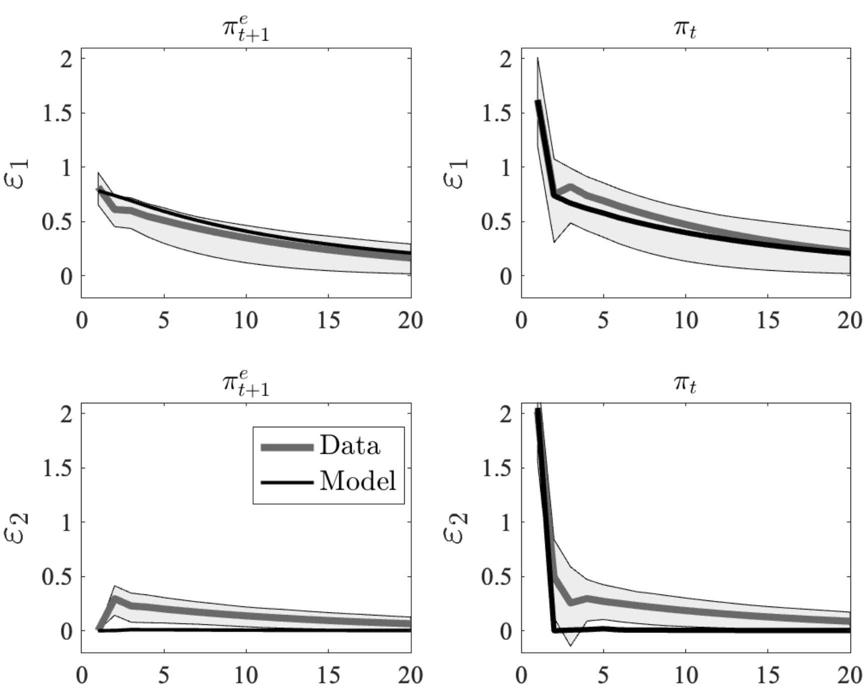
In Figure I.1, we see the model cannot match the IRF of the second shock in the Cholesky VAR because the shock w_t affecting π_t on impact will have no impact on expectations at all in the model without timing restriction. The variance-covariance structure from the model without timing restriction is:

Table I.2: Variance and Auto-covariance of Inflation

	Data	PLM	Model
$var(\pi_{t,h})$	12.52	11.14	11.14
$cov(\pi_{t,h}, \pi_{t-1,h})$	7.70	5.11	4.98
$cov(\pi_{t,h}, \pi_{t-2,h})$	6.99	4.74	4.87

Notes: PLM stands for "Perceived Law of Motion". "The PLM and "Model" moments are the average moment across 200 random samples. In the estimations, we penalize distance between data and model 2-VAR responses as well as distance between these three moments in the PLM and in the model.

Figure I.1: IRF from data and simulated data from model without timing restriction



Notes: The thick gray line is the IRF from the 2-VAR with actual data; the black line is the average IRF from simulated data across 200 random samples. The estimated VAR is a VAR(2) with Cholesky decomposition, ordering π^e_{t+1} *first. The shaded area represents the 95% confidence interval.*