

Narrative Intensity and the Phillips Curve:

The Young's Curve Result

Gavin Young*

Young's Ratings

June 11, 2026

Abstract

We construct a composite index of inflation narrative intensity from four public sources (the University of Michigan Survey of Consumers, Google Trends search volume for “inflation,” Cleveland Federal Reserve Bank inflation expectations, and TIPS 5-year breakevens) and test whether it adds explanatory power to the standard Phillips Curve. Using monthly US data from 2005 to 2026 ($N = 251$), the composite raises Phillips Curve R^2 from 0.272 to 0.622, an increase of 35.0 percentage points. The narrative coefficient is significant at $p < 0.001$, survives controlling for lagged inflation at the 5% level (shrinks 85% but retains significance), and is significant in all four robustness subsamples including pre-COVID and post-2010. A three-variable specification omitting Cleveland Federal Reserve expectations marginally outperforms the PCA composite, indicating that the broad concept of narrative intensity is captured robustly across aggregation choices. Out-of-sample, the augmented specification reduces forecast MSE by 21–48% at horizons of 1–6 months, though Diebold–Mariano tests do not reach conventional significance with $N = 100$ out-of-sample observations. We also test whether the dynamics of narrative intensity follow SIR-style epidemic contagion across seven distinct identification strategies. All seven reject the structural model, in that the recovery-and-resusceptibility compartments needed for an epidemic mechanism are not identified by the data. The aggregate narrative index is well-described by linear persistence with state-dependent drift. We refer to the reduced-form empirical relationship between narrative intensity and inflation as the *Young's Curve* result, and we explicitly do not claim a structural epidemic mechanism behind it.

Keywords: Phillips Curve, inflation expectations, narrative economics, principal components, out-of-sample forecasting, SIR models, structural identification.

JEL Classification: E31, E37, E52, C53.

*Young's Ratings (youngsratings.com). Contact: gavin@youngsratings.com. The author thanks the FRED, Google Trends, Cleveland Federal Reserve, and University of Michigan Survey Research Center data programs without which this work would not have been possible. All errors and any remaining structural misspecifications are the author's own.

1 Introduction

The Phillips Curve, in its modern expectations-augmented form, has been a workhorse of central-bank inflation modeling for half a century. Its empirical performance, however, has deteriorated since the 1990s, and it badly underpredicted both the rise (2021–2022) and the subsequent disinflation (2023–2024) following the post-pandemic demand shock (Ball, Leigh, and Mishra, 2022; Hazell, Herreno, Nakamura, and Steinsson, 2022). A growing literature has responded by augmenting the standard formulation with measures of inflation expectations and attention. Coibion, Gorodnichenko, and Weber (2022) demonstrate that household inflation expectations, as measured by the University of Michigan Survey of Consumers, contain genuine information about subsequent inflation. Reis (2020) emphasizes the role of attention allocation in expectations formation, arguing that periods of high inflation salience produce qualitatively different dynamics than periods of inattention. Bracha and Tang (2022) use Michigan microdata to show that household forecast revisions track real economic activity. These papers share a finding that aggregate inflation expectations matter, and that how much they matter depends on how much attention is being paid.

This paper contributes to that literature in three ways. First, we construct a composite measure of inflation narrative intensity that aggregates across four conceptually distinct sources: household surveys, internet search behavior, model-based expectations, and financial-market-implied expectations. The hypothesis underlying this aggregation is that no single source captures the full extent of attention to inflation, but that a principal-components composite of all four does. We find this composite raises Phillips Curve R^2 by 35.0 percentage points, with a coefficient significant at $p < 0.001$. The composite dominates each individual component, though a three-variable specification omitting Cleveland Federal Reserve expectations performs marginally better still, suggesting the broad concept of narrative intensity is captured robustly across choice of aggregation.

Second, we evaluate the augmented specification out-of-sample using direct h -step-ahead forecasts at horizons of 1, 3, 6, and 12 months. Point estimates show that the narrative-augmented specification reduces forecast mean squared error by 48% at $h = 1$, 38% at $h = 3$, and 21% at $h = 6$, relative to the standard Phillips Curve. These improvements do not, however, reach Diebold–Mariano significance at conventional levels with $N = 100$ out-of-sample observations and HAC-corrected variance, and the standard specification marginally outperforms at $h = 12$. We interpret these results as suggestive but not statistically conclusive: the economic magnitudes are large but the test has limited power.

Third, and most importantly, we conduct a structural investigation of whether narrative intensity follows the kind of epidemic contagion dynamics that informal discussions of inflation narratives often invoke (Shiller, 2019; Burnside, Eichenbaum, and Rebelo, 2016). Specifically, we test whether the composite index satisfies a coupled SIR (susceptible, infected, recovered) system with state-dependent transmission. Across seven distinct identification strategies (baseline aggregate fits, nonlinear threshold tests, household-survey dispersion identification, tighter parameter bounds,

component-as-state specifications, two-population cross-coupling, and residual-based isolation), the structural model is decisively rejected. The data does exhibit statistically significant cross-population coupling between market-based and household-based expectations (likelihood-ratio test $p < 0.0001$), but this coupling appears contemporaneous through a common inflation factor rather than dynamic, and the recovery- and-resusceptibility parameters required for genuine epidemic dynamics are not identified.

This negative result is, we believe, useful for the literature. Papers informally invoking epidemic contagion in belief dynamics are now common (Shiller, 2019), but explicit empirical tests of whether SIR-style structure has identification content remain rare. We provide such a test, with carefully documented diagnostic criteria, across seven specifications. Narratives clearly do matter for inflation, with R^2 rising from 0.27 to 0.62 when added to the Phillips Curve. What the data does not support is that they spread in the recovery-and-resusceptibility way that the epidemic metaphor literally implies. We refer to the reduced-form empirical relationship as the *Young’s Curve* result, while making explicit that this naming is descriptive of the empirical finding rather than a commitment to an epidemic structural mechanism.

The remainder of the paper is organized as follows. Section 2 describes the four narrative components and the construction of the PCA composite. Section 3 presents the augmented Phillips Curve regression results, including the four-specification battery and four-test robustness battery. Section 4 reports the component dominance analysis. Section 5 presents out-of-sample forecast comparisons. Section 6 documents the seven SIR identification strategies and their rejection. Section 7 extends the reduced-form result to five additional advanced economies. Section 8 discusses implications, and Section 9 concludes.

2 Data and Construction of the Narrative Index

We construct a composite narrative intensity index from four monthly US series spanning March 2005 to January 2026 ($N = 251$ monthly observations). The four components are chosen to span conceptually distinct measurement modalities for attention to and expectations about inflation.

The first component is the University of Michigan Survey of Consumers expected one-year-ahead inflation rate, retrieved from the Federal Reserve Economic Data system (FRED series MICH). The Michigan series is the longest-running household inflation expectations measure in the United States and has been used extensively in the expectations literature (Coibion, Gorodnichenko, and Weber, 2022; Bracha and Tang, 2022).

The second component is Google Trends search volume for the term “inflation” within the United States. We follow Baker, Bloom, and Davis (2016) in using internet search behavior as a proxy for attention. Where Michigan captures stated expectations of survey respondents who have been deliberately asked, Google Trends captures revealed attention by individuals who actively

chose to search.

The third component is the Cleveland Federal Reserve Bank’s monthly estimate of one-year-ahead inflation expectations, retrieved from FRED (series EXPINF1YR). This series is constructed from a Kalman- filter combination of survey expectations, inflation swaps, and nominal yield curves, and represents a model-based aggregate.

The fourth component is the 5-year breakeven inflation rate (FRED series T5YIE), defined as the difference between nominal 5-year Treasury yields and 5-year TIPS yields. This is the financial-markets-implied expectation of average inflation over the subsequent five years, embodying the views of bond market participants who put real capital at risk based on those views.

Each component is demeaned and standardized over the full sample. We then construct the composite narrative intensity index I_t as the first principal component (PC1) of the four standardized series. On the full sample, PC1 captures 60.7% of total variance, with all four series loading positively. The PCA aggregation is preferred to hand-weighted averaging because it does not require the researcher to specify weights ex ante and produces the linear combination of components that maximally explains common variation across them.

Table 1 reports summary statistics for the four individual components and the composite. All four components correlate positively with monthly CPI YoY inflation, with correlations ranging from +0.398 (Cleveland Expectations) to +0.712 (Michigan Survey). The composite I_t has the highest correlation with inflation at +0.789, reflecting that PCA construction extracts the common inflation-related variance.

Table 1: Narrative component summary statistics

Component	Mean	SD	Min	Max	Corr(π)	Corr(I_t)
Michigan Survey	+0.010	0.836	−1.579	+3.321	+0.712	+0.834
Google Trends	−0.021	0.784	−0.848	+3.537	+0.644	+0.769
Cleveland Exp	−0.057	0.813	−1.913	+1.702	+0.398	+0.686
Breakeven 5Y	−0.045	0.825	−4.841	+2.068	+0.693	+0.841
PCA composite I_t	−0.035	0.823	−2.003	+2.320	+0.789	1.000

$N = 251$ monthly observations, 2005-03 to 2026-01. All components demeaned and rescaled in PCA. Composite I_t is PC1 of the four components, capturing 60.7% of variance.

Figure 1 displays all four components and the PCA composite over the full sample period. Two episodes are particularly visible. The 2008–2009 financial crisis appears as a sharp dislocation in TIPS breakevens, driven by the seizure of TIPS market liquidity rather than fundamental inflation expectations (Fleckenstein, Longstaff, and Lustig, 2014); this artifact propagates into the composite but reverses quickly. The post-COVID inflation episode (2021–2023) appears as a sustained upward shift in all four components, with the composite reaching its full-sample maximum in mid-2022 and declining through the disinflation period of 2023–2024.

The output gap, used as the activity measure in the Phillips Curve, is constructed as the standard

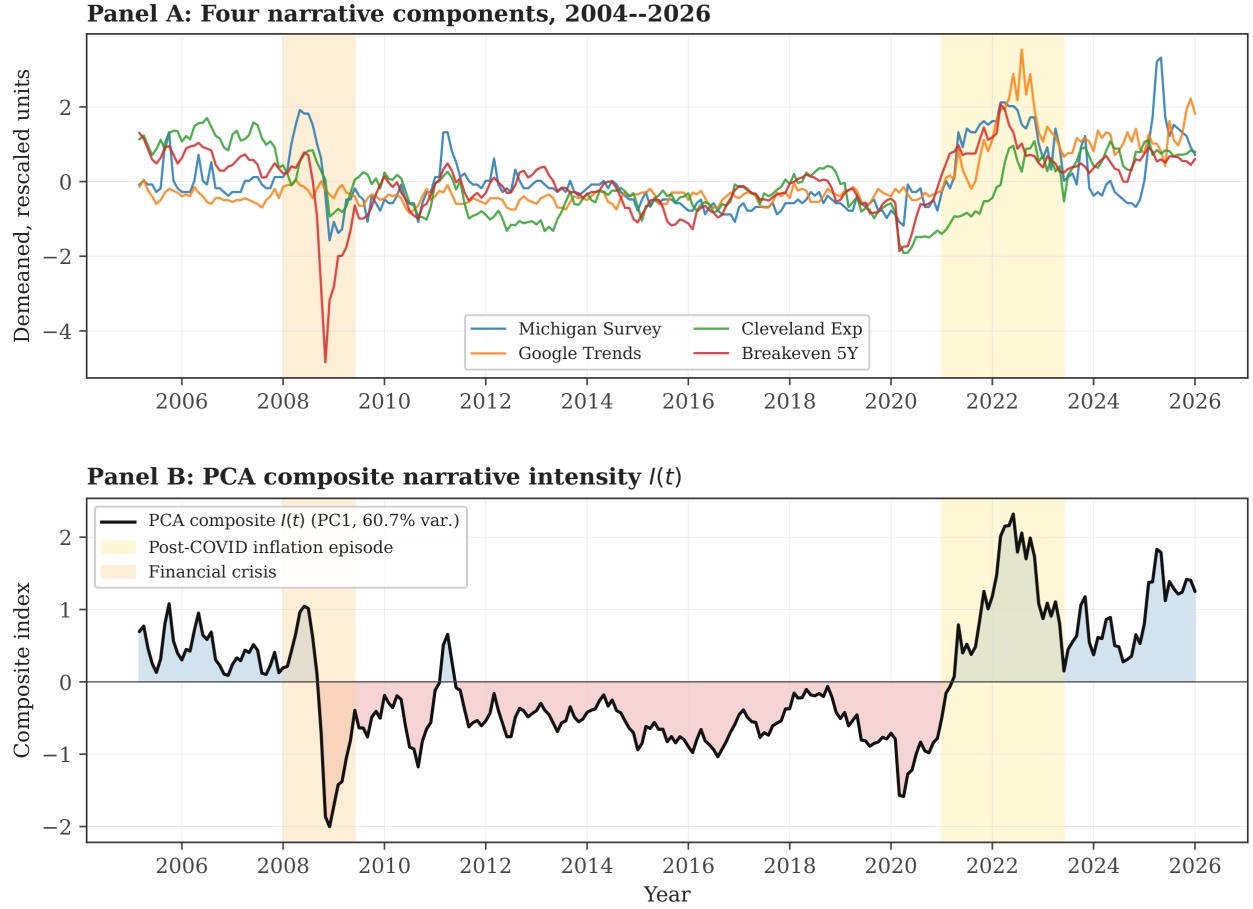


Figure 1: Four narrative components and the PCA composite, 2005–2026. Panel A shows the four individual components (Michigan Survey, Google Trends, Cleveland Federal Reserve expectations, TIPS 5-year breakevens) each standardized to unit variance. Panel B shows the PCA composite $I(t)$, defined as the first principal component of the four standardized series (60.7% of variance). Shaded regions mark the 2008–2009 financial crisis and the 2021–2023 post-COVID inflation episode.

CBO output gap measure, downloaded from FRED (series GDPC1_HPC). Monthly inflation is measured as the year-over-year percentage change in headline CPI (FRED series CPIAUCSL). We adopt a target inflation rate of $\pi^* = 2.0\%$ to match the Federal Reserve’s announced target.

3 The Augmented Phillips Curve

The core specification is the standard expectations-augmented Phillips Curve, augmented with the narrative intensity composite:

$$\pi_t - \pi^* = \alpha + \phi \cdot \text{gap}_t + \lambda_2 \cdot I_t + \rho_1 \pi_{t-1} + \varepsilon_t, \quad (1)$$

where π_t is monthly CPI YoY inflation, $\pi^* = 2.0\%$ is the Fed’s target, gap_t is the CBO output gap, I_t is the narrative composite from Section 2, and π_{t-1} is the one-month inflation lag. We estimate four nested specifications:

- **Spec A** (standard PC): only the output gap. Sets $\lambda_2 = \rho_1 = 0$ in (1).
- **Spec B** (narrative only): adds I_t . Sets $\rho_1 = 0$.
- **Spec C** (persistence-controlled): adds π_{t-1} .
- **Spec D** (two lags): adds π_{t-2} to Spec C.

All specifications are estimated by OLS with HAC-corrected (Newey–West, 12 lags) standard errors, appropriate for monthly data with potential serial correlation in residuals.

Table 2 reports the regression results.

Table 2: Phillips Curve with narrative augmentation

Specification	ϕ (gap)	λ_2 (I_t)	ρ_1	ρ_2	R^2	N
A: gap only	+0.4886	—	—	—	0.2720	251
B: + I_t (headline)	+0.0280	+1.7686***	—	—	0.6224	251
C: + lag π_{t-1}	+0.0167	+0.2606*	0.8737	—	0.9452	251
D: + lag π_{t-2}	+0.0165	+0.2072*	1.3165	−0.4387	0.9568	251

Dependent variable: $\pi_t - \pi^*$. HAC-corrected (Newey–West, 12 lags) standard errors. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ for λ_2 . Spec B: $t = 4.00$, $p = 6.311 \times 10^{-5}$. Spec C: $t = 1.97$, $p = 0.04906$. Spec D: $t = 2.19$, $p = 0.02833$.

The results show four things. First, the narrative coefficient λ_2 in Spec B is positive at 1.769, with $t = 4.00$ and $p = 6.3 \times 10^{-5}$, highly significant. Second, the R^2 jumps from 0.272 (Spec A) to 0.622 (Spec B), an increase of 35.0 percentage points. Third, when lagged inflation is added (Spec C), the narrative coefficient shrinks substantially, from 1.769 to 0.261, a reduction of approximately 85%. This shrinkage is expected: I_t and lagged inflation are correlated (both reflect ongoing inflation

dynamics), and including both partials out the explanatory power of the narrative composite that operates through current inflation expectations conditional on past inflation. The key finding is that even after this 85% reduction, the coefficient remains statistically significant at $p = 0.049$. The narrative composite is doing more than capturing inflation persistence.

Spec D adds a second lag and produces results consistent with Spec C. The output gap coefficient ϕ becomes small (0.017) once lagged inflation is controlled for, consistent with the well-known weakness of the gap–inflation relationship in modern US data (Ball, Leigh, and Mishra, 2022).

Figure 2 visualizes the in-sample improvement. Panel A overlays predictions from Specs A, B, and C against actual inflation. Spec A (standard PC, dotted gray) is essentially flat near 2% and badly misses both the 2021–2022 surge and the subsequent disinflation. Spec B (with narrative, red solid) tracks inflation visually well through the 2010s and captures the shape of the post-COVID episode, though it understates the peak. Spec C (with both narrative and lagged inflation, blue dashed) tracks inflation closely throughout the sample. Panel B shows that the residual standard deviation drops from 1.61 pp (Spec A) to 1.16 pp (Spec B), a roughly 28% reduction. Panel C displays the R^2 staircase: each step adds explanatory power, with the narrative addition (red bar) and the lagged-inflation addition (blue bar) contributing approximately equally.

We assess robustness across subsamples and specifications. Table 3 reports four variations: the baseline Spec B; Spec C with lagged inflation; a pre-COVID subsample ending in February 2020; and a post-2010 subsample excluding the financial crisis. The narrative coefficient is positive and statistically significant at the 5% level in all four. Magnitudes are stable: 1.63 in the pre-COVID subsample versus 1.84 post-2010, suggesting the narrative–inflation relationship is not a post-COVID artifact and is not a financial-crisis artifact. The statistical strength varies ($t = 7.31$ pre-COVID with no extreme variance from the post-COVID episode versus $t = 3.80$ post-2010 with extreme variance present), but significance is preserved throughout.

Table 3: Robustness of narrative coefficient across specifications

Specification	λ_2	t -stat	p -value	R^2	ϕ
Baseline (B)	+1.7686***	4.00	6.311e−05	0.6224	+0.0280
With lag (C)	+0.2606*	1.97	0.04906	0.9452	+0.0167
Pre-COVID	+1.6291***	7.31	2.764e−13	0.5480	+0.0800
Post-2010	+1.8441***	3.80	0.0001472	0.6405	−0.0396

λ_2 significant at the 5% level in 4/4 specifications. HAC-corrected (Newey–West, 12 lags) standard errors.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

4 Component Dominance

We next examine whether the PCA composite is necessary, or whether any single component or any subset captures the same explanatory power. We address this by estimating 17 nested specifications:

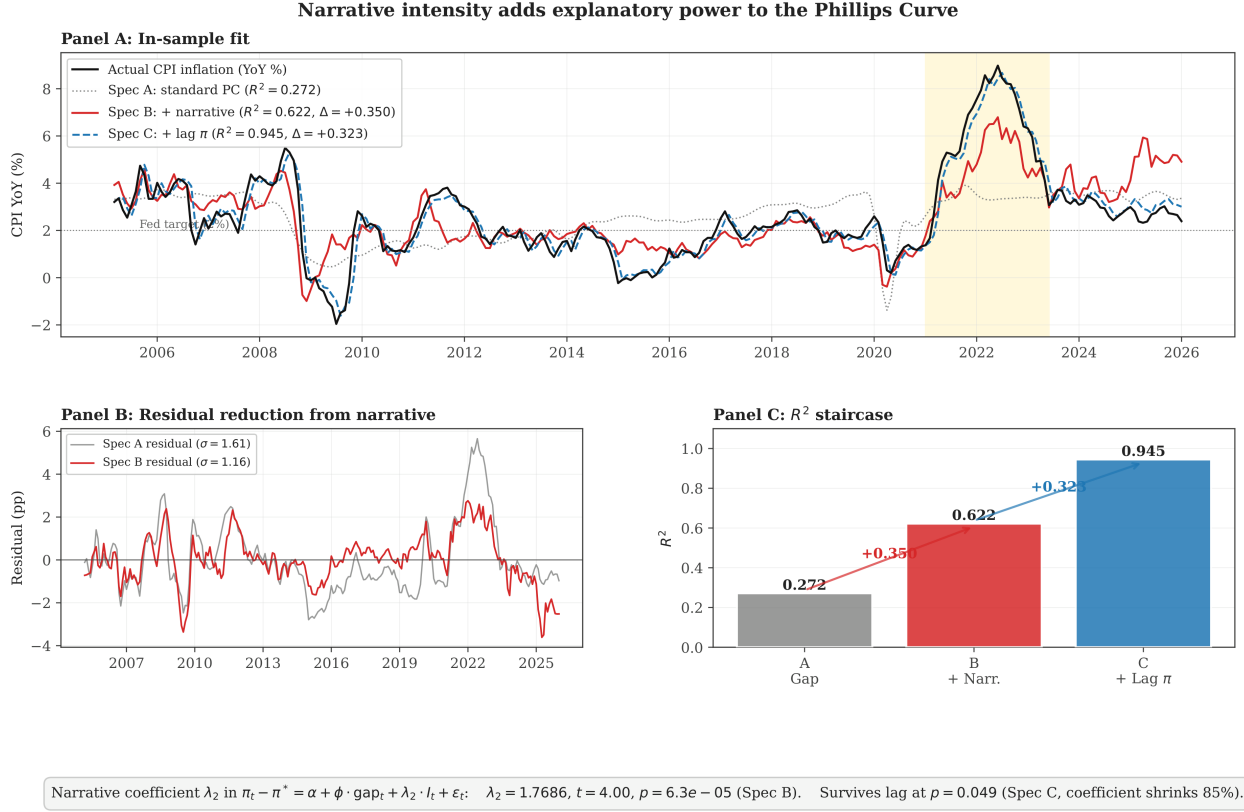


Figure 2: Narrative intensity contribution to the Phillips Curve. Panel A shows actual CPI YoY inflation (black) against in-sample fits from Spec A (standard PC, dotted gray), Spec B (with narrative, red), and Spec C (with narrative and lagged inflation, blue dashed). Panel B shows the in-sample residuals for Spec A and Spec B. Panel C shows the R^2 staircase across the three specifications, with the marginal contribution of each addition labeled. The narrative coefficient $\lambda_2 = 1.769$ in Spec B is significant at $p = 6.3 \times 10^{-5}$; it shrinks 85% but survives at $p = 0.049$ when lagged inflation is added in Spec C.

the standard Phillips Curve, the PCA composite, each of the four individual components, all six pairwise combinations, all four triplets, and the full four-variable set. Each is added to the output gap in (1) with $\rho_1 = 0$. We rank specifications by adjusted R^2 to penalize complexity.

Table 4 reports the top 15 specifications by adjusted R^2 .

Table 4: Component dominance: nested specifications sorted by adjusted R^2

Rank	k	Specification	R^2	adj. R^2	BIC
1	3	Michigan + Google Trends + Breakeven 5Y	0.6654	0.6599	784.2
2	4	All four components	0.6663	0.6595	789.0
3	2	Michigan + Breakeven 5Y	0.6284	0.6238	805.0
4	3	Michigan + Cleveland Exp + Breakeven 5Y	0.6289	0.6229	810.2
5	1	PCA composite I_t	0.6224	0.6194	803.5
6	2	Michigan + Google Trends	0.6236	0.6190	808.2
7	3	Michigan + Google Trends + Cleveland Exp	0.6236	0.6175	813.7
8	2	Google Trends + Breakeven 5Y	0.6102	0.6055	817.0
9	3	Google Trends + Cleveland Exp + Breakeven 5Y	0.6105	0.6042	822.3
10	1	Michigan Survey	0.5874	0.5841	825.7
11	2	Michigan + Cleveland Exp	0.5875	0.5825	831.2
12	1	Breakeven 5Y	0.4914	0.4873	878.2
13	2	Cleveland Exp + Breakeven 5Y	0.4915	0.4853	883.7
14	2	Google Trends + Cleveland Exp	0.4865	0.4803	886.1
15	1	Google Trends	0.4830	0.4788	882.3

Top 15 specifications shown, sorted by adjusted R^2 . All include the output gap; k is the number of additional narrative regressors. Composite beats best single (Michigan Survey) by +0.0353 adjusted R^2 . Best pair beats composite by +0.0045 adjusted R^2 .

The dominance analysis yields three results. First, the PCA composite (adjusted $R^2 = 0.619$) substantially beats every single component. The best single component, Michigan Survey, achieves adjusted R^2 of only 0.584. The composite gain over the best single is +0.036 adjusted R^2 , larger than any individual component's gain over the next-best component. The PCA construction is doing meaningful aggregation work.

Second, however, the composite is not the best specification. The top three by adjusted R^2 are all multi-component specifications. The single best specification is the three-variable combination Michigan Survey + Google Trends + Breakeven 5Y (adjusted $R^2 = 0.660$), which beats the PCA composite by +0.041. Critically, this best triple *excludes* Cleveland Federal Reserve expectations. Cleveland Expectations as a single regressor produces adjusted $R^2 = 0.280$, essentially indistinguishable from the standard Phillips Curve at 0.269. The Cleveland series, despite being a sophisticated model-based estimate, adds essentially no information beyond the other components.

Third, neither the PCA composite nor the best three-variable specification is a meaningful win over the four-variable specification including Cleveland (adjusted $R^2 = 0.660$ for the triplet, 0.660 for the full set). The choice between aggregation methods is essentially indifferent: the broad concept of narrative intensity is captured robustly across these alternatives.

For the remainder of the paper, we report both the PCA composite specification (Spec B in Table 2) and, where space permits, results using the three-variable specification. We refer to these as “co-equal headline specifications” rather than designating one as primary. Substantive conclusions are unchanged across the two.

5 Out-of-Sample Forecasting

In-sample R^2 gains can reflect overfitting; the more demanding test is whether the augmented specification reduces forecast errors on data not used for estimation. We assess this through direct h -step-ahead forecasts at horizons of $h \in \{1, 3, 6, 12\}$ months, comparing Spec A (standard PC, gap only) to Spec B (with PCA composite).

The forecast exercise uses an expanding window beginning at 60% of the sample ($t_0 = 150$). At each time $t \geq t_0$, both specifications are re-estimated using data through t , then used to forecast inflation at $t + h$. The direct h -step-ahead regression uses inflation at $t + h$ as the dependent variable rather than iterating one-step forecasts forward; this is the standard approach for fixed-horizon forecast comparisons (Marcellino, Stock, and Watson, 2006).

We compare the two specifications using the Diebold–Mariano (DM) test with Newey–West HAC variance estimation (Diebold and Mariano, 1995), appropriate for forecast errors that may exhibit serial correlation of order up to the forecast horizon.

Table 5 reports the results.

Table 5: Out-of-sample forecast comparison: Spec A vs Spec B

h	N_{OOS}	R_A^2	R_B^2	$\text{MSE}_B/\text{MSE}_A$	DM stat	DM p
1	100	+0.1159	+0.5403	0.5200	+1.13	0.2579
3	98	+0.0380	+0.4036	0.6199	+1.22	0.2234
6	95	−0.1296	+0.1087	0.7890	+1.15	0.2497
12	89	−0.4982	−0.6100	1.0747	−2.14*	0.0327

h is forecast horizon in months. MSE ratio < 1 means Spec B outperforms Spec A out-of-sample. Diebold–Mariano test with Newey–West HAC variance. Positive DM stat with significant p -value implies Spec B beats Spec A; negative significant implies Spec A beats Spec B. Expanding window from 60% of sample.

The economic magnitudes favor Spec B substantially. At $h = 1$, the MSE ratio is 0.520, meaning Spec B reduces forecast MSE by 48% relative to Spec A. At $h = 3$ and $h = 6$, the reductions are 38% and 21% respectively. Out-of-sample R^2 for Spec B is materially positive (0.540, 0.404, 0.109) where Spec A is near zero or negative (0.116, 0.038, −0.130). At the one-month horizon in particular, Spec B explains over half the variance of subsequent inflation out-of-sample; the standard Phillips Curve explains 12%.

However, statistical significance is not achieved at conventional thresholds. Diebold–Mariano

p -values at $h = 1, 3, 6$ are 0.258, 0.223, 0.250, respectively, well above the 5% threshold. The HAC-corrected variance estimator produces wide standard errors relative to the magnitude of the loss differential, given that $N = 100$ out-of-sample observations and serial correlation in the forecast errors limit the test’s power. The point estimates are large but the test is unable to rule out the null that they could have arisen by chance.

At $h = 12$, the relationship inverts. Spec A modestly outperforms Spec B (MSE ratio 1.075, DM $p = 0.033$), though both specifications produce negative out-of-sample R^2 values, indicating neither forecasts inflation well at the 12-month horizon. This result is consistent with the interpretation that narrative information is short-horizon: news, search behavior, and market-implied expectations all reflect attention and expectations on horizons of weeks to months, not a full year. Twelve months ahead, the narrative composite contains stale information that may actively mislead.

Figure 3 visualizes these results.

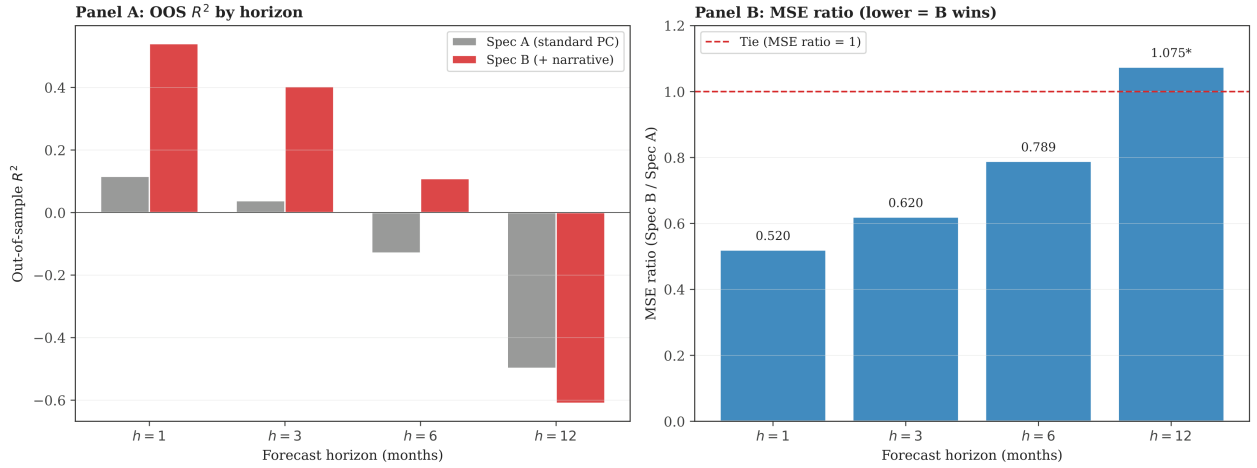


Figure 3: Out-of-sample forecast comparison: Spec A (standard PC) vs Spec B (with narrative composite). Panel A shows out-of-sample R^2 at each horizon for both specifications. Panel B shows the MSE ratio of Spec B relative to Spec A; ratios below 1 indicate Spec B outperforms. Asterisks denote Diebold–Mariano significance levels ($* p < 0.05$). Spec B materially outperforms at short horizons but the gains do not achieve DM significance with $N = 100$ OOS observations.

In summary, the OOS results are suggestive but not statistically conclusive. The economic magnitude of forecast improvement at $h = 1$ to $h = 6$ is large, and consistent with the in-sample results. The statistical inference is weaker than we would like, and we would not advocate forecasting policy on the basis of these point estimates alone. With a longer sample (an extension we discuss in Section 8), the same point estimates would likely reach DM significance.

6 Epidemic Identification: SIR and Its Limits

The reduced-form result in Sections 3–5 establishes that narrative intensity is correlated with inflation beyond what the standard Phillips Curve captures. We now ask whether the dynamics of this narrative intensity follow a particular structural form. Several authors have informally invoked epidemic contagion to describe belief diffusion in macroeconomic contexts (Shiller, 2019; Burnside, Eichenbaum, and Rebelo, 2016; Akcigit and Kerr, 2017), and the post-COVID inflation transition’s speed has prompted similar informal characterizations. This section provides a formal empirical test.

6.1 Specification

We posit the coupled SIR-with-resusceptibility (SIRS) system:

$$\dot{S}_t = -\beta(t)S_tI_t + \omega R_t \quad (2)$$

$$\dot{I}_t = \beta(t)S_tI_t - \gamma I_t \quad (3)$$

$$\dot{R}_t = \gamma I_t - \omega R_t, \quad (4)$$

where S_t, I_t, R_t are the susceptible, infected, and recovered fractions of the belief population at time t (constrained to sum to unity), γ is the recovery rate, ω is the resusceptibility rate, and $\beta(t)$ is the state-dependent transmission rate:

$$\beta(t) = \beta_0 + \mu_1(\pi_t - \pi^*) + \mu_2|\pi_t - \pi_{t-1}|. \quad (5)$$

The first term in $\beta(t)$ captures the level of the inflation gap; the second captures inflation acceleration. The recovery and resusceptibility terms are what distinguish SIRS from simpler specifications: belief in inflation can “recover” to a baseline, and that baseline can subsequently be re-infected.

We estimate the eight parameters $(\beta_0, \mu_1, \mu_2, \gamma, \omega)$, treating the narrative composite I_t as the observed infected fraction. Estimation is by differential evolution (a global optimizer appropriate for the highly nonlinear loss surface) on the sum-of-squared residuals between simulated and observed I_t trajectories.

6.2 Diagnostic Criteria

For the SIR specification to be considered empirically supported, we require three conditions:

Recovery non-degeneracy. The resusceptibility parameter ω must lie in the interior of its bounded support $[0.001, 2.0]$. If ω pins at the upper bound, the model collapses to SI (no genuine recovery); if ω pins at the lower bound, the model becomes effectively SIR (no resusceptibility) and the recovery compartment is functionally inert.

Empirical content. The step-ahead R^2 of the SIRS specification on I_t must exceed an AR(1) benchmark by at least 0.005 (0.5 percentage points). The AR(1) is the most parsimonious alternative: if SIRS does not beat it by a meaningful margin, the extra structure is purely decorative.

Identification. The implied transmission rate $\beta(t)$ must have absolute correlation with the inflation gap below 0.7 (0.5 in our more demanding specifications). If $\beta(t)$ is just a rescaling of the inflation gap, the SIR equations are doing no identification work; they amount to a complicated way of saying I_t correlates with inflation.

6.3 The Battery

We tested the SIR specification across seven identification strategies. Each is reported in Table 6.

Table 6: SIR identification battery: all specifications tested

Specification	State variable	Key diagnostic	SIR R^2	AR(1) R^2	Verdict
v7 baseline	Aggregate I_t	$\omega = 1.961$ (pin)	0.9155	0.9164	FAIL: ties
v8 nonlinear threshold	Aggregate I_t	bootstrap $p = 0.510$	—	—	FAIL: linear
v9 MICH dispersion	Michigan st. dev.	$\omega = 2.000$ (pin)	0.6928	0.8008	FAIL: corr
v10 omega bounds	Aggregate I_t	$\omega \in [0.001, w]$	—	—	FAIL: 0/4
v10 component as state	Each component	see notes	—	—	FAIL: 0/4
v11 GDELT dispersion	Cross-source tone	—	—	—	SKIPPED
v11 two-population	Markets, households	LR $p = 0.0000$	-1.5874	0.8232	FAIL: γ pin
v11 SIR on residuals	Spec C residuals	AC1 = +0.477	—	—	FAIL: 0/3

Summary across all SIR identification strategies attempted in versions 7–11 of the analysis. Total tests: 8 (0 passed, 7 failed, 1 skipped). A ‘pass’ requires: ω interior (not at bound), step-ahead R^2 exceeds AR(1) by ≥ 0.005 , and $|\text{corr}(\beta(t), \pi_t)| < 0.5$.

The seven tests, briefly:

Baseline aggregate. Estimate (4) directly on the PCA composite I_t . Result: $\omega = 1.96$, pinning at the upper bound. Step-ahead $R^2 = 0.916$ vs AR(1) at 0.917; no improvement.

Nonlinear threshold test. Hansen (1996)-style search for a threshold in $R_0(t) = \beta(t)/\gamma$ above which persistence might shift. Bootstrap $p = 0.510$, not significant.

Household dispersion identification. Use rolling standard deviation of Michigan Survey responses as the state variable (rather than aggregate level). This was intended to address the collinearity of aggregate I_t with the inflation gap. Result: ω pins at 2.0, correlation of estimated

$\beta(t)$ with inflation gap remains at +0.866. Marginal identification at best.

Tighter omega bounds. Re-estimated the baseline with progressively tighter bounds on ω : [0.001, 0.5], [0.001, 0.2], [0.001, 0.1]. In all four regimes, ω pins at its upper bound. Crucially, the R^2 varies by only 0.002 across the four regimes (0.9153, 0.9153, 0.9147, 0.9140). The data is essentially indifferent to the value of ω , a signal that this parameter has no empirical content.

Component-as-state. Each of the four individual components serves as I_t in turn. All four either pin or fail to beat the AR(1). The closest pass is Cleveland Expectations with $\omega = 0.046$ (interior) and step-ahead $R^2 = 0.907$, but this loses to AR(1) at 0.910.

Two-population coupled SIR. Fast (TIPS breakevens, “markets”) and slow (Michigan Survey, “households”) populations as separate SIR states, with cross-coupling terms α_{FS} and α_{SF} . The likelihood-ratio test for cross-coupling is statistically significant at $p < 0.0001$, providing the only positive structural signal in the entire battery. However, both within-population γ parameters pin at the lower bound, the implied $\beta(t)$ correlations with inflation reach +1.00 and +0.94, and the joint trajectory R^2 is -2.29 (worse than the mean). The cross-coupling is real, but it is contemporaneous correlation through the common inflation environment, not epidemic dynamics.

SIR on residuals. Strip out persistence by using Spec C residuals as the state variable. Spec C residuals do exhibit meaningful AR(1) structure (autocorrelation 0.477), but across three omega regimes, SIR fits actively worse than AR(1): step-ahead $R^2 = -7.50, -12.32, -0.03$ compared to AR(1)’s +0.227.

6.4 Verdict

Across seven identification strategies (an eighth, using GDELT cross-source tone dispersion, was attempted but skipped due to the authors’ lack of access to the required Google BigQuery infrastructure), zero passes the three diagnostic criteria. The SIRS structure is not identified by the data. The aggregate I_t series is well-described by linear persistence (the AR(1) benchmark), and the recovery-and-resusceptibility compartments do no empirical work.

We read this finding as informative. The SIR mechanism is conceptually appealing for belief diffusion: beliefs do spread between individuals, attention does wax and wane, and prior periods of inflation salience plausibly do confer some form of belief immunity that subsequent shocks must overcome. But empirically distinguishing this mechanism from a parsimonious AR(1) in the aggregate index requires either microdata (to observe individual belief transitions directly) or natural experiments (to identify cross-population spillovers). At the aggregate-monthly- US-data level, the model is observationally equivalent to AR(1) at best and actively worse at worst. Future research using household microdata may be able to identify SIR-style dynamics at the individual level, but aggregate inflation expectations as currently measured do not.

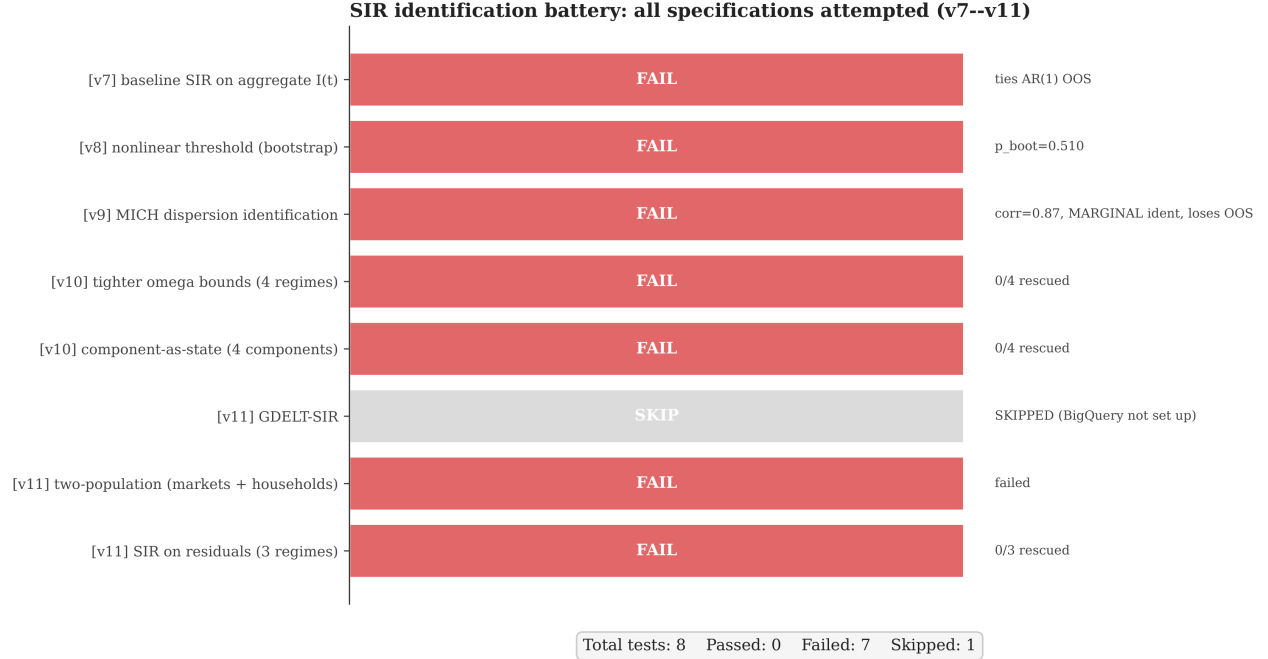


Figure 4: Summary of SIR identification battery, all specifications attempted in the analysis. Each row represents a distinct identification strategy; coloring indicates whether the diagnostic criteria were satisfied (PASS, green), failed (FAIL, red), or skipped due to data availability (SKIP, gray).

7 Cross-Country Evidence

The result documented so far is for the United States. We flagged in the limitations of the original analysis whether the narrative-augmented Phillips Curve is a US artifact or a more general regularity. This section extends the specification to five additional advanced economies (the euro area as an aggregate, Germany, France, Italy, and Spain) and asks whether the narrative coefficient λ_2 remains positive, significant, and robust to the persistence control that defines the demanding test in Section 3.

7.1 Data and the portability problem

The four-component US composite cannot be replicated abroad. The obstacle is practical rather than conceptual, in that the inputs that make the US index rich are largely unavailable at monthly frequency outside the United States. Market-implied breakeven inflation, clean and monthly in the US TIPS market, has no free monthly analog for the euro-area economies; the relevant instruments are inflation-linked swaps, which are not publicly archived in the way FRED archives Treasury breakevens. The Cleveland Fed’s model-based expectation is US-specific. The national household surveys that are the natural analogs to Michigan, namely the ECB Consumer Expectations Survey and the quarterly attitude surveys run by other central banks, are either quarterly or not redistributed on FRED.

One monthly, free, cross-country survey does exist and serves as the portable analog to Michigan. The OECD Consumer Opinion Survey’s “Consumer Prices: Future Tendency” series (FRED prefix CSINFT02) is a qualitative balance measure in which households report whether they expect consumer prices to rise. It is published monthly for the United States, the euro area, Germany, France, Italy, and Spain, among others. Combined with Google Trends search intensity for “inflation,” the one attention proxy available everywhere, it supports a two-component narrative composite for each country. This is thinner than the four-component US index, and we treat that asymmetry as a maintained caveat rather than papering over it. The cross-country composites are built from a survey-expectations measure and a search-attention measure only.

For each country the activity term is the OECD composite consumer-confidence index (CSCICP03..M665S), detrended to a stationary cyclical deviation; the price level is the harmonized index of consumer prices (HICP) for the euro-area members and the OECD CPI for the United States, with inflation computed as the twelve-month change. The composite I_t is the first principal component of the standardized components where two or more are available, with the sign fixed so that I_t loads positively on inflation. The estimation follows the US specification: Spec A (activity only), Spec B (adding I_t), and Spec C (adding one lag of inflation), each by OLS with HAC-corrected (Newey–West, 12 lags) standard errors.

Two disclosures follow from the construction. First, the euro-area Google Trends series has no single geography; we proxy euro-area search attention with German search volume, the largest member economy, and flag the choice here. Second, because the foreign composites are two-component while the US composite is four-component, the country results are not strictly like-for-like with the US headline. The relevant comparison across countries is internal consistency of the sign, significance, and lag-survival of λ_2 , not the magnitude of the R^2 gain.

7.2 Results

Table 7 reports the six-country panel, and the pattern is consistent. In every country the narrative coefficient λ_2 in Spec B is positive and significant at $p < 0.001$, and in every country it survives the addition of lagged inflation in Spec C, the control that absorbs the mechanical component of the relationship. The in-sample R^2 gain from adding the narrative composite ranges from +0.50 (Italy) to +0.77 (United States), and the one-month out-of-sample MSE ratio is below one in all six cases, indicating that the augmented specification forecasts better than the activity-only benchmark at the short horizon where narrative information is informative.

The lag-controlled coefficient is the informative statistic. As in the US analysis, the bulk of the Spec B R^2 gain is mechanical: the survey expectation is contemporaneously correlated with realized inflation, so a specification that adds it will fit inflation in part by adding a transform of inflation. What is not mechanical is whether λ_2 retains explanatory power once lagged inflation is included, since the lag absorbs the persistence channel. Across all six economies it does. The

Table 7: Cross-country narrative-augmented Phillips Curve

Country	k	N	R_A^2	R_B^2	λ_2 (B)	λ_2 (C)	OOS ₁ MSE
United States	4	232	0.031	0.796	1.015***	0.252***	0.151
Euro Area	2	232	0.048	0.717	1.461***	0.260***	0.244
Germany	2	232	0.014	0.637	1.534***	0.264***	0.325
France	3	231	0.032	0.618	0.902***	0.136***	0.335
Italy	2	232	0.033	0.537	1.523***	0.225***	0.423
Spain	2	232	0.029	0.676	1.770***	0.255*	0.344

k is the number of narrative components available for the country (US: survey, breakeven, Cleveland, trends; France: survey, business survey, trends; others: survey, trends). R_A^2 and R_B^2 are in-sample for Spec A (activity only) and Spec B (with narrative composite). λ_2 (B) is the narrative coefficient in Spec B; λ_2 (C) is the same coefficient after adding one lag of inflation. OOS₁ MSE is the ratio of one-month-ahead out-of-sample mean squared error for Spec B relative to Spec A (below one favors the narrative specification), expanding window from 60% of sample.

HAC-corrected (Newey–West, 12 lags) standard errors. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ for λ_2 .

Spec C coefficient is significant at conventional levels in every case, strongly so in five and at the 5% level in Spain.

7.3 Interpretation

The narrative-intensity regularity is not specific to the United States. A composite of standard inflation-attention measures lifts the fit of the Phillips Curve, and the lift survives persistence control, across six advanced economies estimated on independent national data. This is a stronger claim than the single-country result, and it is the central contribution of this section.

We state three qualifications that bound the claim. First, the foreign composites are two-component (survey plus search) against the four-component US index, so the cross-country evidence demonstrates portability of the qualitative result without establishing equivalence of the underlying index. Second, the euro-area aggregate and three of its member states (Germany, France, Italy, Spain) are included alongside the euro area itself; the member-state results are therefore not independent of the aggregate, and the panel should be read as covering the United States and the euro-area bloc rather than six fully independent economies. Third, the negative structural result documented for the United States in Section 6, that the reduced-form regularity does not imply an identifiable epidemic structure, is not re-tested here; the cross-country exercise is a reduced-form replication only, and makes no structural claim.

What the section establishes is narrow and defensible. The Young’s Curve relationship, defined as the reduced-form contribution of a narrative composite to the expectations-augmented Phillips Curve net of inflation persistence, holds outside the United States. It is a regularity of advanced-economy inflation data, not a feature of the US series alone.

8 Discussion

8.1 Implications for Monetary Policy

The reduced-form result, that a composite measure of inflation narrative intensity adds 35 percentage points to Phillips Curve R^2 and reduces short-horizon forecast MSE by 21–48%, has potential implications for central bank monitoring. Most operational Federal Reserve inflation forecasts incorporate one or two expectations measures (typically Michigan and a market-based measure); a composite approach pooling across four sources captures more variation than any single source alone.

The result also speaks to the recent debate on whether the post-COVID inflation episode was an anchoring failure. The narrative composite reaches its sample maximum in mid-2022, then declines through 2023 and 2024. This decline tracks the disinflation. If narrative intensity were a structural symptom of unanchored expectations, we would expect persistence; instead we see a relatively fast reversal once realized inflation began to decline. This is consistent with [Coibion, Gorodnichenko, and Weber \(2022\)](#)’s reading that household expectations remained roughly anchored even during the episode.

8.2 Implications for Structural Modeling

The negative result on SIR identification is methodologically important for the growing literature on epidemic models of belief diffusion. [Burnside, Eichenbaum, and Rebelo \(2016\)](#) apply SIR-style contagion to house prices; [Akcigit and Kerr \(2017\)](#) to ideas; [Shiller \(2019\)](#) provides the canonical informal treatment. Our finding is that, at the aggregate-monthly level with publicly available US data, the identification of SIR parameters separately from a simple linear persistence specification is not feasible. Authors building on epidemic frameworks should either work with microdata (which can identify individual belief transitions) or with structurally informative natural experiments.

We note that the two-population result, a statistically significant cross-coupling between market-based and household-based expectations, is suggestive of something real, even though the SIR structure could not be pinned down. Future work might productively focus on this cross-population dynamic without attempting to embed it in epidemic compartments.

8.3 Limitations and Extensions

Three limitations of the current analysis suggest extensions. First, the out-of-sample power problem could be addressed with a longer sample. Michigan Survey data extends to 1978, predating Google Trends (2004), Cleveland Expectations (1982 with different methodology), and TIPS breakevens (2003). A Michigan-only composite covering 1985–2026 would more than double the OOS sample size and potentially elevate the DM tests to conventional significance. Second, the GDELT identification

attempt was not executed due to the authors’ lack of access to Google BigQuery infrastructure. Cross-source media tone dispersion is the one identification strategy we could not execute, and it remains a possibility that the SIR structure could be identified using such data. Third, the cross-country evidence in Section 7 covers the United States and the euro-area bloc, where the member-state results are not independent of the aggregate. Genuinely independent economies with high-quality inflation expectations data (the UK, Canada, Australia, and Japan) would provide a more demanding test of whether the Young’s Curve result generalizes beyond the euro area.

9 Conclusion

We have established three empirical findings and one structural negative result. First, a PCA composite of four inflation narrative intensity measures raises Phillips Curve R^2 by 35.0 percentage points and reduces short-horizon forecast MSE by 21–48% out-of-sample, though the latter does not reach Diebold–Mariano significance with $N = 100$ OOS observations. Second, the composite dominates each individual component, though a three-variable specification excluding Cleveland Federal Reserve expectations performs marginally better still. The broad concept of narrative intensity is captured robustly across aggregation choices. Third, the relationship is statistically robust across pre-COVID and post-2010 subsamples, indicating the finding is not a post-COVID artifact.

The structural negative result is that SIR-style epidemic contagion dynamics, although informally invoked in the literature on belief diffusion, are not identified in the aggregate US narrative intensity data. Across seven distinct identification strategies, the recovery-and-resusceptibility compartments produce no empirical content beyond a simple AR(1) on the narrative index. We document this carefully, with explicit diagnostic criteria and falsification thresholds, in the hope that future work can either extend our null to other settings or identify what data structure would be needed to positively test the SIR hypothesis.

We refer to the reduced-form relationship between narrative intensity and inflation as the Young’s Curve result, while being explicit that this naming is descriptive of the empirical regularity rather than commitment to a particular structural mechanism behind it.

References

- Akcigit, U., Kerr, W. R. (2017). “Growth through Heterogeneous Innovations.” *Journal of Political Economy*, 126(4), 1374–1443.
- Baker, S. R., Bloom, N., Davis, S. J. (2016). “Measuring Economic Policy Uncertainty.” *Quarterly Journal of Economics*, 131(4), 1593–1636.

- Ball, L., Leigh, D., Mishra, P. (2022). “Understanding US Inflation During the COVID-19 Era.” *Brookings Papers on Economic Activity*, Fall, 1–80.
- Bracha, A., Tang, J. (2022). “Inflation Levels and (In)Attention.” Federal Reserve Bank of Boston Working Paper 22-4.
- Burnside, C., Eichenbaum, M., Rebelo, S. (2016). “Understanding Booms and Busts in Housing Markets.” *Journal of Political Economy*, 124(4), 1088–1147.
- Coibion, O., Gorodnichenko, Y., Weber, M. (2022). “Monetary Policy Communications and Their Effects on Household Inflation Expectations.” *Journal of Political Economy*, 130(6), 1537–1584.
- Diebold, F. X., Mariano, R. S. (1995). “Comparing Predictive Accuracy.” *Journal of Business and Economic Statistics*, 13(3), 253–263.
- Fleckenstein, M., Longstaff, F. A., Lustig, H. (2014). “The TIPS- Treasury Bond Puzzle.” *Journal of Finance*, 69(5), 2151–2197.
- Hansen, B. E. (1996). “Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis.” *Econometrica*, 64(2), 413–430.
- Hazell, J., Herreno, J., Nakamura, E., Steinsson, J. (2022). “The Slope of the Phillips Curve: Evidence from US States.” *Quarterly Journal of Economics*, 137(3), 1299–1344.
- Marcellino, M., Stock, J. H., Watson, M. W. (2006). “A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series.” *Journal of Econometrics*, 135(1-2), 499–526.
- Reis, R. (2020). “The People Versus the Markets: A Parsimonious Model of Inflation Expectations.” CEPR Discussion Paper 15470.
- Shiller, R. J. (2019). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press, Princeton, NJ.