Geneva Graduate Institute (IHEID)

Topics in Econometrics (EI137)

Term Paper

The Incremental Predictive Power of Consumer Sentiment in Macroeconomic Forecasting

Evidence from a Hierarchical Bayesian VAR and Forecast-Revision Diagnostics

Jingle Fu

Professor: Marko Mlikota

Spring 2025

January 17, 2026

Abstract

This paper asks whether consumer sentiment adds incremental predictive content for U.S. inflation and industrial production once standard macro aggregates and financial prices are already included, and whether sentiment alters forecast-revision patterns consistent with informational frictions. Using monthly data (1985M1–2019M12) and an expanding-window pseudo out-of-sample design (origins 2001M1–2019M11), we estimate hierarchical Bayesian VARs under three nested information sets: *Small* (core macro), *Medium* (+ financial variables and oil), and *Full* (+ sentiment). Forecast accuracy is evaluated by RMSFE at horizons $h \in \{1, 3, 12\}$, and forecast revisions are assessed via the Coibion and Gorodnichenko (2015) error-on-revision regression.

Two results summarize the evidence. First, all BVAR specifications substantially improve on a no-change benchmark for inflation, but sentiment delivers little incremental reduction in inflation RMSFE once financial variables are included: the best long-horizon inflation performance is attained by the baseline macro specification. Second, the revision diagnostic shows sizeable short-horizon underreaction and long-horizon overreaction for inflation; richer information sets move long-horizon coefficients toward the rational-expectations benchmark, while short-horizon underreaction remains economically meaningful. Because the information sets are nested, statistical inference on accuracy differences relies on nested-model-robust procedures; we report Clark–West adjusted MSPE tests as robustness and emphasize the magnitude and stability of RMSFE differences as primary evidence.

Keywords: Bayesian VAR; hierarchical shrinkage; forecasting; consumer sentiment; forecast revisions.

JEL codes: C11; C53; E37.

1 Introduction

This paper studies a practical forecasting question: does consumer sentiment add incremental predictive content for inflation and industrial production once conventional macro aggregates and financial prices are already included? I address this question using a transparent horse race across three nested information sets within a hierarchical Bayesian VAR, complemented by a revision-based diagnostic of forecast updating following Coibion and Gorodnichenko (2015).

The analysis uses monthly U.S. data and a recursive pseudo out-of-sample design. Forecast accuracy is summarized by RMSFE at a small set of horizons, and revision dynamics are summarized by the error-on-revision regression with Newey–West HAC standard errors.

The headline findings are disciplined. For inflation, all BVAR specifications improve markedly on a no-change benchmark, but sentiment adds little incremental reduction in RMSFE once the medium information set is included (and the smallest model attains the best long-horizon inflation RMSFE). For industrial production, adding financial variables improves short-horizon forecasts, while long-horizon accuracy remains difficult. Because the information sets are nested, I treat standard equal-accuracy tests as suggestive and report nested-model-robust Clark–West adjusted MSPE tests as robustness.

2 Data

The dataset comprises monthly U.S. series ending before the COVID-19 period. The information sets are nested: *Small* includes industrial production, CPI, unemployment, and the federal funds rate; *Medium* adds the 10-year Treasury yield, the S&P 500, and oil prices; *Full* further adds the University of Michigan sentiment index. The empirical design is constructed so that differences between Medium and Full isolate the incremental role of sentiment.

Following the standard BVAR forecasting literature, the model is estimated in levels or log-levels Giannone, Lenza, and Primiceri (2015); Sims (1980). Forecasts are evaluated on annualized cumulative growth rates (constructed from the same origin date as the forecast), ensuring comparability across horizons. Full transformation and implementation details are reported in the extended version.

3 Empirical design

3.1 Model and prior selection

For each information set, I estimate a reduced-form VAR with monthly lags under a Minnesota-style prior with hierarchical prior selection, as implemented in `BVAR` Kuschnig and Vashold (2021). The overall tightness parameter λ is learned from the data following Giannone et al. (2015); as the information set expands, hierarchical selection tightens shrinkage to mitigate overfitting.

3.2 Pseudo out-of-sample evaluation and diagnostics

I use an expanding-window pseudo out-of-sample design. Accuracy is summarized by RMSFE on annualized cumulative growth targets; forecasts are compared to no-change and AR(1) benchmarks. To assess internal updating, I estimate the Coibion and Gorodnichenko (2015) regression of forecast errors on forecast revisions by horizon using Newey–West HAC standard errors.

4 Results

This section reports the core evidence on the incremental role of sentiment: a forecast-accuracy horse race and a forecast-revision diagnostic.

4.1 Forecast accuracy

Table 1 summarizes RMSFEs for inflation and industrial production across the three nested information sets. The main pattern is that expanding the information set improves some short-horizon forecasts (especially for industrial production), but sentiment adds little incremental accuracy once financial variables are included; for inflation at longer horizons, the baseline macro specification attains the lowest RMSFE.

Table 1: Root Mean Squared Forecast Errors

model	variable	h1	h3	h12
Small	CPI	3.468	2.643	1.305
Small	INDPRO	7.649	5.558	4.998
Medium	CPI	2.982	2.500	1.349
Medium	INDPRO	7.315	4.966	4.371
Full	CPI	3.128	2.538	1.330
Full	INDPRO	7.424	5.087	4.387

Notes: RMSFE (in percentage points for inflation and growth rates).
Sample period: 2001M1–2019M12 (230 forecast origins).
Forecast horizons: $h = \{1, 3, 12\}$ months ahead.
Models: Small (INDPRO, CPI, UNRATE, FEDFUNDS), Medium (+ GS10, SP500), Full (+ UMCSENT).

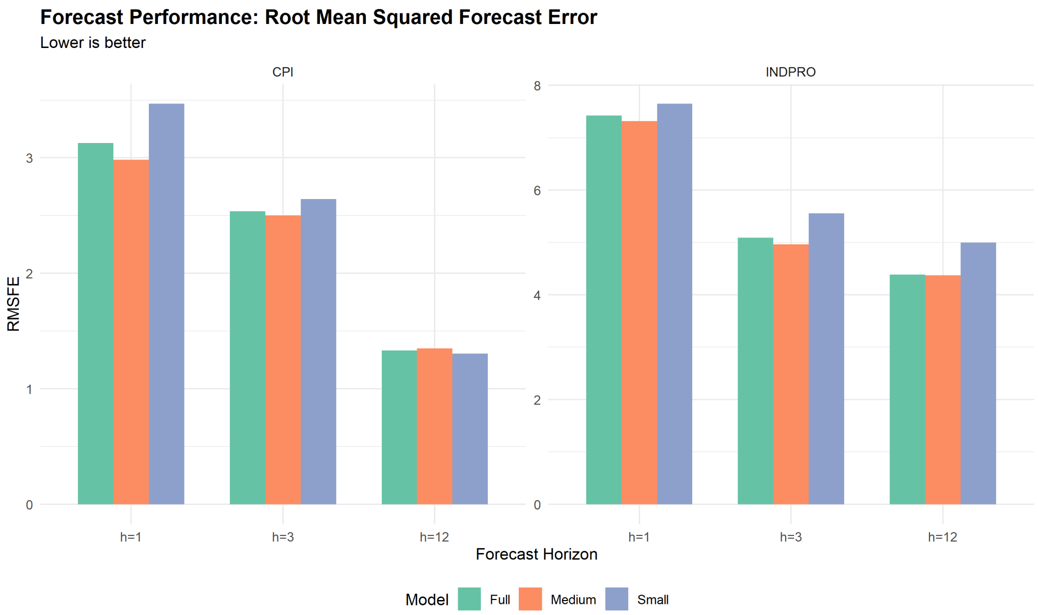

Figure 1: Forecast accuracy by horizon (RMSFE; lower is better)
Notes: Bars report RMSFEs (evaluation scale) for each information set and horizon; values correspond to Table 1. Source: `results/tables/rmsfe_results.csv`.

4.2 Forecast revisions and the CG diagnostic

Table 2 reports the Coibion and Gorodnichenko (2015) error-on-revision coefficients. For inflation, short-horizon coefficients are positive (underreaction) while long-horizon coefficients are near zero or slightly negative (overreaction), and richer information sets move long-horizon coefficients closer to the rational-expectations benchmark. For industrial production, coefficients are small and statistically weak across horizons.

Table 2: Coibion–Gorodnichenko Regression Results

term	estimate	std.error	statistic	p.value
Small CPI h=1	2.2608	1.1942	1.8931	0.0597
Small CPI h=3	0.6917	0.7987	0.8661	0.3874
Small CPI h=12	-0.5178	0.3213	-1.6115	0.1086
Small INDPRO h=1	0.7184	0.5612	1.2801	0.2019
Small INDPRO h=3	0.8923	0.4816	1.8528	0.0653
Small INDPRO h=12	0.1449	0.4423	0.3276	0.7436
Medium CPI h=1	0.7086	0.2840	2.4951	0.0134
Medium CPI h=3	0.5602	0.3204	1.7485	0.0818
Medium CPI h=12	-0.0841	0.2065	-0.4073	0.6842
Medium INDPRO h=1	0.2663	0.4916	0.5416	0.5886
Medium INDPRO h=3	0.5983	0.3637	1.6449	0.1015
Medium INDPRO h=12	0.3184	0.5292	0.6017	0.5480
Full CPI h=1	0.9257	0.3188	2.9040	0.0041
Full CPI h=3	0.6894	0.3729	1.8488	0.0659
Full CPI h=12	-0.0272	0.2111	-0.1291	0.8974
Full INDPRO h=1	0.1078	0.3465	0.3110	0.7561
Full INDPRO h=3	0.1889	0.3630	0.5204	0.6033
Full INDPRO h=12	0.2237	0.4925	0.4543	0.6501

Notes: OLS regression of forecast errors on forecast revisions: $(y_{t+h} - \hat{y}_{t+h|t}) = \alpha_h + \beta_h(\hat{y}_{t+h|t} - \hat{y}_{t+h|t-1}) + \varepsilon_{t+h}$.
Standard errors are Newey–West HAC-robust with lag truncation parameter equal to the forecast horizon.
Under rational expectations, $\beta_h = 0$. Positive values indicate under-reaction (sticky information), while negative values suggest over-reaction consistent with diagnostic expectations.
Sample: 2001M1–2019M12. Variables: CPI (annualized inflation), INDPRO (industrial production growth).

4.3 Nested-model inference caveat

Because the information sets are nested (Small $\subset$ Medium $\subset$ Full), standard equal-accuracy tests can have nonstandard behavior under the null Clark and McCracken (2001). I therefore treat Diebold–Mariano comparisons across nested models as suggestive and report Clark–West MSPE-adjusted tests as robustness Clark and West (2007) (Appendix Table 3). The main interpretation emphasizes magnitudes and stability of RMSFE differences.

5 Conclusion

I study whether sentiment adds incremental predictive content in a hierarchical BVAR once standard macro aggregates and financial variables are already included. The core evidence is that sentiment adds little incremental improvement in point-forecast accuracy, while revision diagnostics indicate economically meaningful short-horizon underreaction and long-horizon coefficients closer to (or slightly below) zero as the information set expands.

Because the models are nested and the exercise is descriptive (pseudo out-of-sample using final vintages), I emphasize magnitudes and stability of RMSFE differences and report Clark–West MSPE-adjusted tests as robustness. A natural next step is to evaluate the same design with real-time data vintages and alternative sentiment measures.

References

Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, *105*(1), 85–110. doi: 10.1016/S0304-4076(01)00083-9

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311. doi: 10.1016/j.jeconom.2006.05.023

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*(8), 2644–2678. doi: 10.1257/aer.20110306

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, *97*(2), 436–451. doi: 10.1162/REST_a_00483

Kuschnig, N., & Vashold, L. (2021). Bvar: Bayesian vector autoregressions with hierarchical prior selection in r. *Journal of Statistical Software*, *100*(14), 1–27. doi: 10.18637/jss.v100.i14

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48.

A Additional figures and robustness

Nested-model forecast accuracy: Clark–West tests. Table 3 reports Clark–West MSPE-adjusted tests for nested model comparisons (Small vs. Medium; Medium vs. Full) at horizons $h \in \{1, 3, 12\}$; one-sided p-values correspond to the alternative that the larger model improves MSPE. This robustness addresses the nonstandard behavior of standard equal-accuracy tests under nesting.

Table 3: Clark–West (2007) MSPE-Adjusted Tests for Nested Models

Smaller	Larger	variable	horizon	t-stat	p-value	N	NW lag
Small	Medium	CPI	h=1	3.312***	0.001	227.000	1.000
Small	Medium	CPI	h=3	2.405***	0.008	225.000	3.000
Small	Medium	CPI	h=12	-0.063	0.525	216.000	12.000
Small	Medium	INDPRO	h=1	3.211***	0.001	227.000	1.000
Small	Medium	INDPRO	h=3	2.387***	0.009	225.000	3.000
Small	Medium	INDPRO	h=12	2.452***	0.008	216.000	12.000
Medium	Full	CPI	h=1	-1.146	0.874	227.000	1.000
Medium	Full	CPI	h=3	-0.325	0.627	225.000	3.000
Medium	Full	CPI	h=12	0.742	0.230	216.000	12.000
Medium	Full	INDPRO	h=1	0.107	0.458	227.000	1.000
Medium	Full	INDPRO	h=3	0.057	0.477	225.000	3.000
Medium	Full	INDPRO	h=12	0.253	0.400	216.000	12.000

Notes: Clark–West (2007) MSPE-adjusted test for equal forecast accuracy in nested models. For smaller-model forecast error $e_{1t} = y_t - f_{1t}$ and larger-model error $e_{2t} = y_t - f_{2t}$, the adjusted loss differential is
$d_t = e_{1t}^2 - \left(e_{2t}^2 - (f_{2t} - f_{1t})^2\right)$. The test regresses d_t on a constant.
Newey–West HAC standard errors use lag truncation equal to the forecast horizon (overlap adjustment).
One-sided p-values reported for the alternative that the larger model improves MSPE.
*** p<0.01, ** p<0.05, * p<0.1

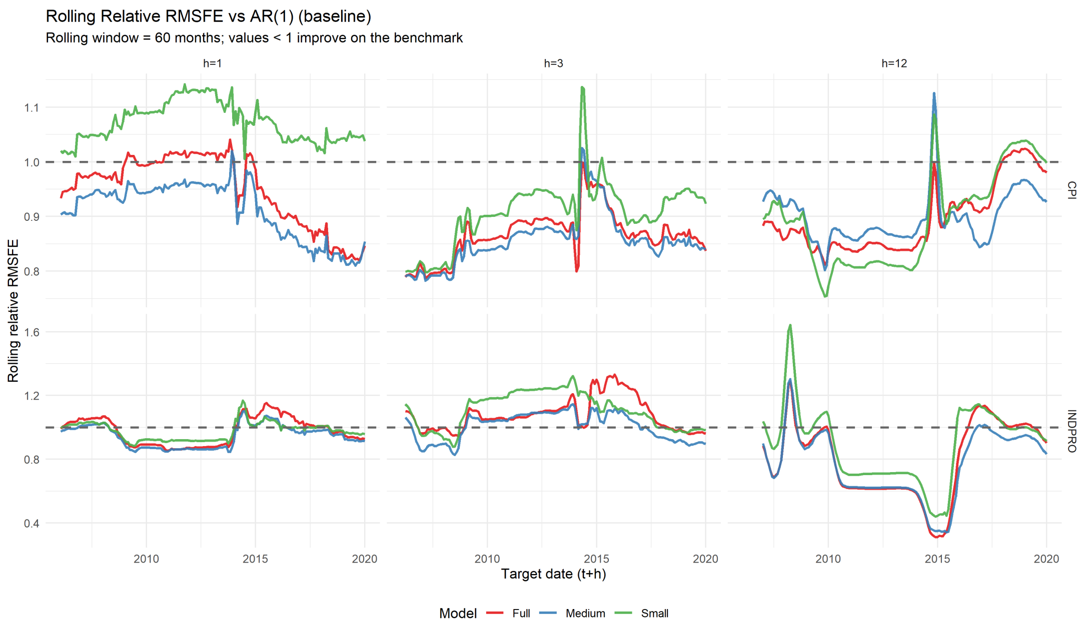

Figure 2: Rolling relative RMSFE versus AR(1) benchmark

Notes: Rolling-window relative RMSFEs (window length 60 months). Values below one indicate improvement over the recursively estimated AR(1) benchmark.

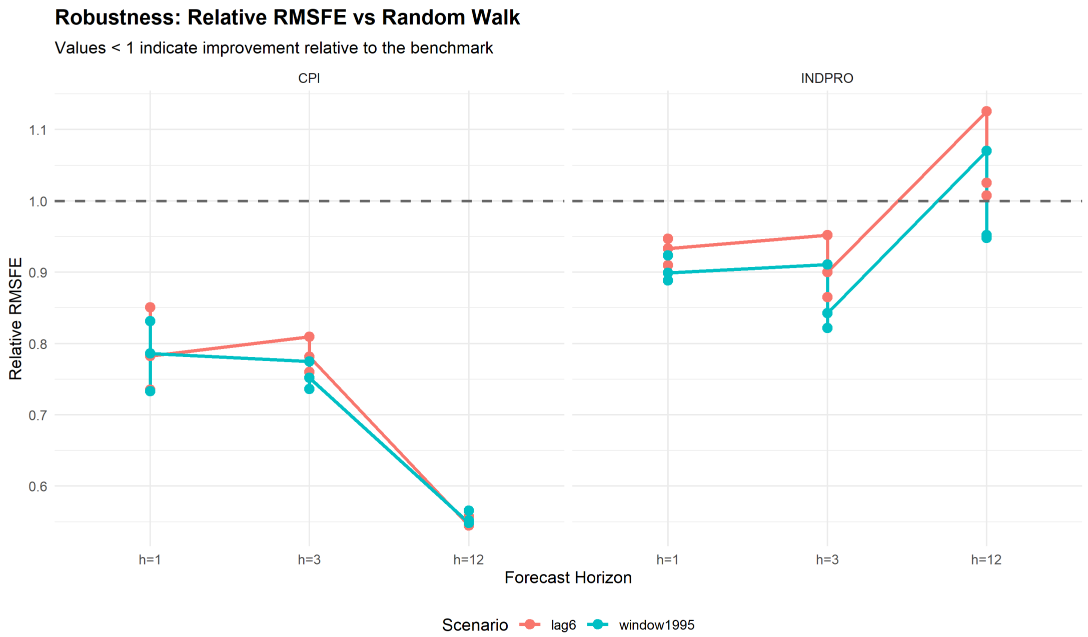

Figure 3: Robustness: relative RMSFE versus no-change benchmark

Notes: Relative RMSFEs under alternative lag length ($p = 6$) and an earlier initial training window end date (1995M12). Values below one indicate improvement over the no-change benchmark.

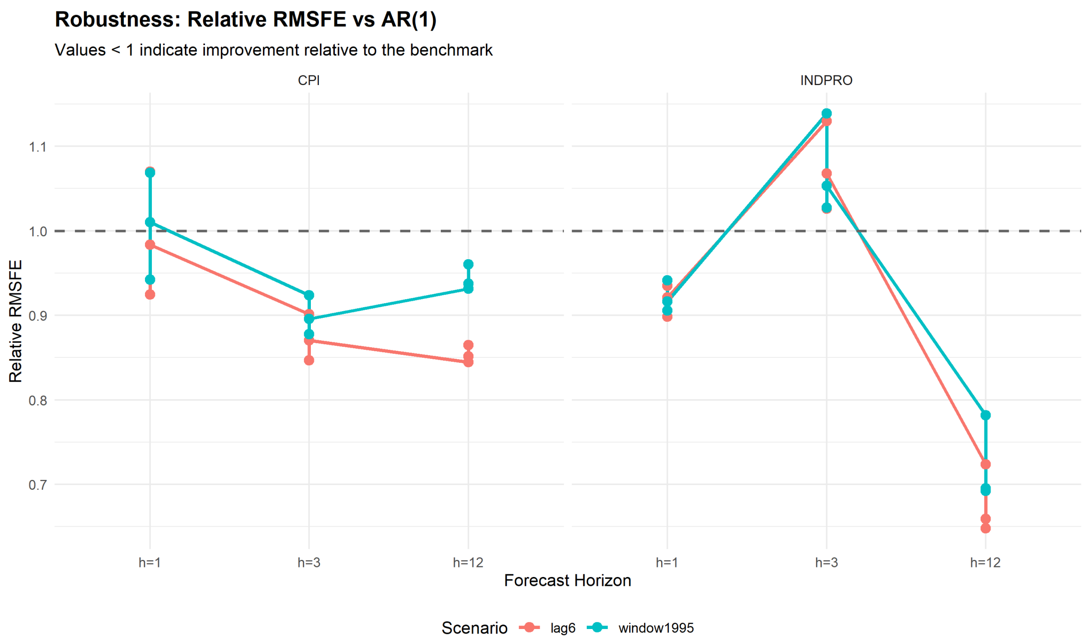

Figure 4: Robustness: relative RMSFE versus AR(1) benchmark

Notes: Relative RMSFEs under robustness scenarios, reported against the recursively estimated AR(1) benchmark.

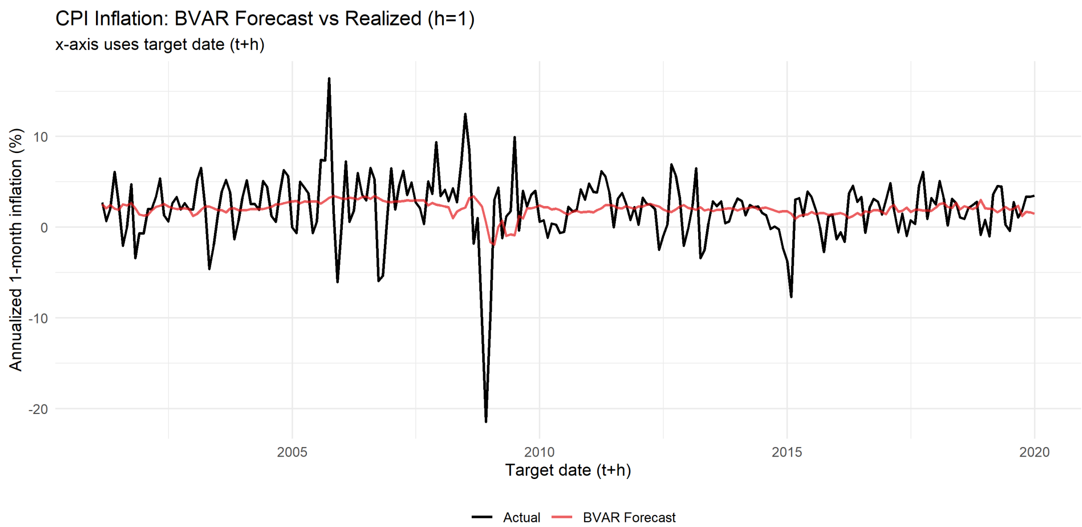

Figure 5: CPI inflation: BVAR forecast versus realized ($h = 1$)

Notes: The x-axis uses the target date ($t + h$). The plotted forecast is the model-implied predictive mean from the baseline specification.

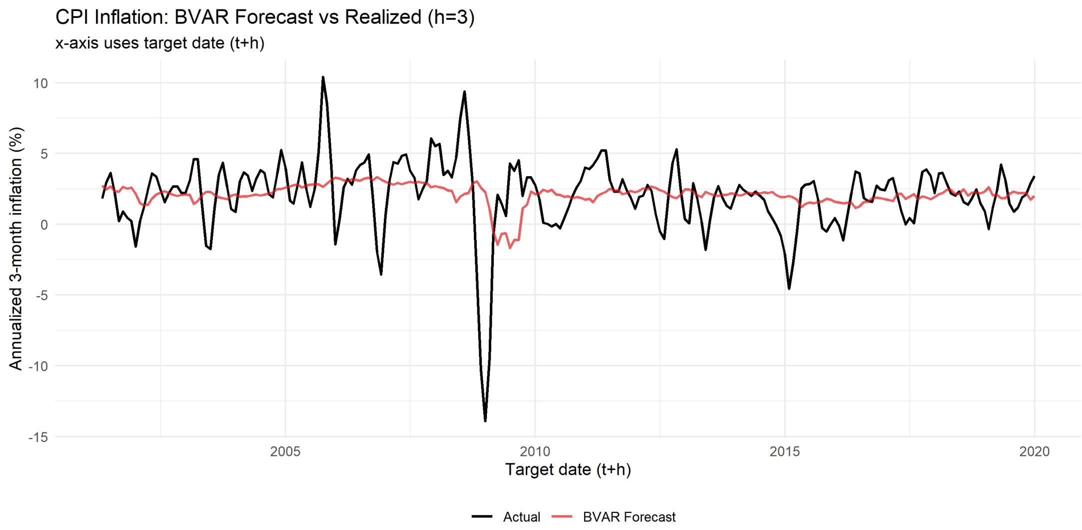

Figure 6: CPI inflation: BVAR forecast versus realized ($h = 3$)

Notes: The x-axis uses the target date $(t + h)$. The plotted forecast is the model-implied predictive mean from the baseline specification.

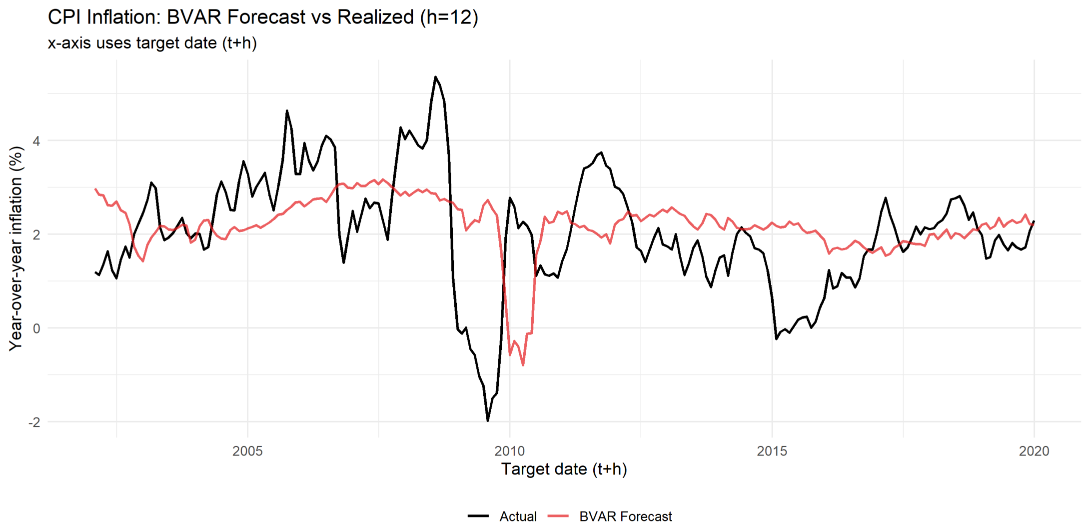

Figure 7: CPI inflation: BVAR forecast versus realized ($h = 12$)

Notes: The x-axis uses the target date $(t + h)$. The plotted forecast is the model-implied predictive mean from the baseline specification.

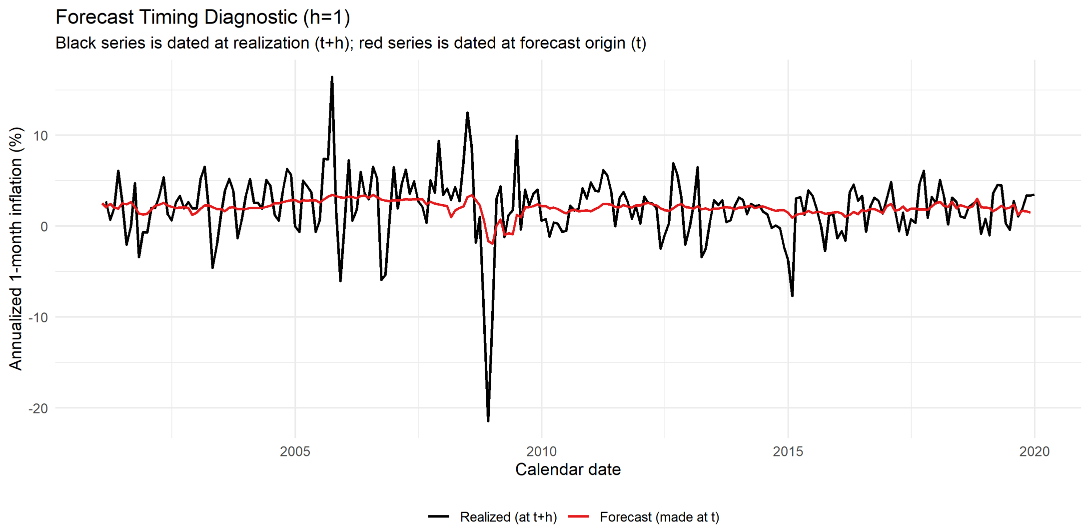

Figure 8: Forecast timing diagnostic ($h = 1$)

Notes: The black series is dated at the realization ($t + h$); the red forecast series is dated at the forecast origin (t).

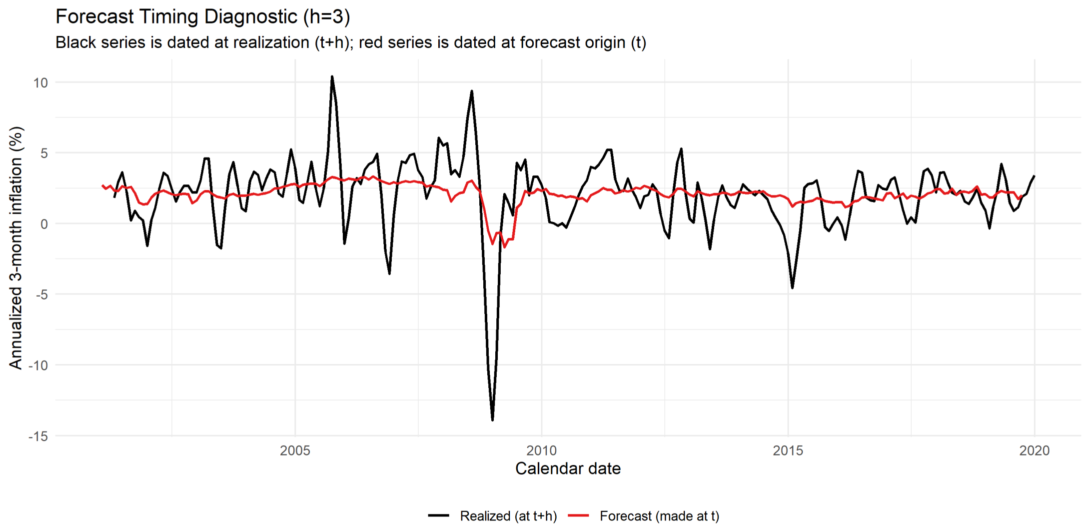

Figure 9: Forecast timing diagnostic ($h = 3$)

Notes: The black series is dated at the realization ($t + h$); the red forecast series is dated at the forecast origin (t).

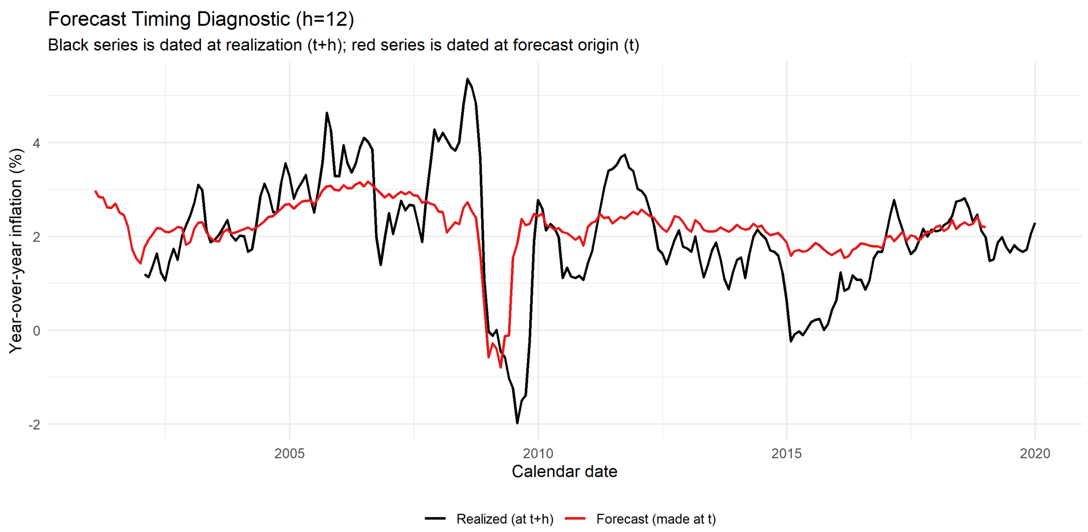

Figure 10: Forecast timing diagnostic ($h = 12$)

Notes: The black series is dated at the realization ($t + h$); the red forecast series is dated at the forecast origin (t).

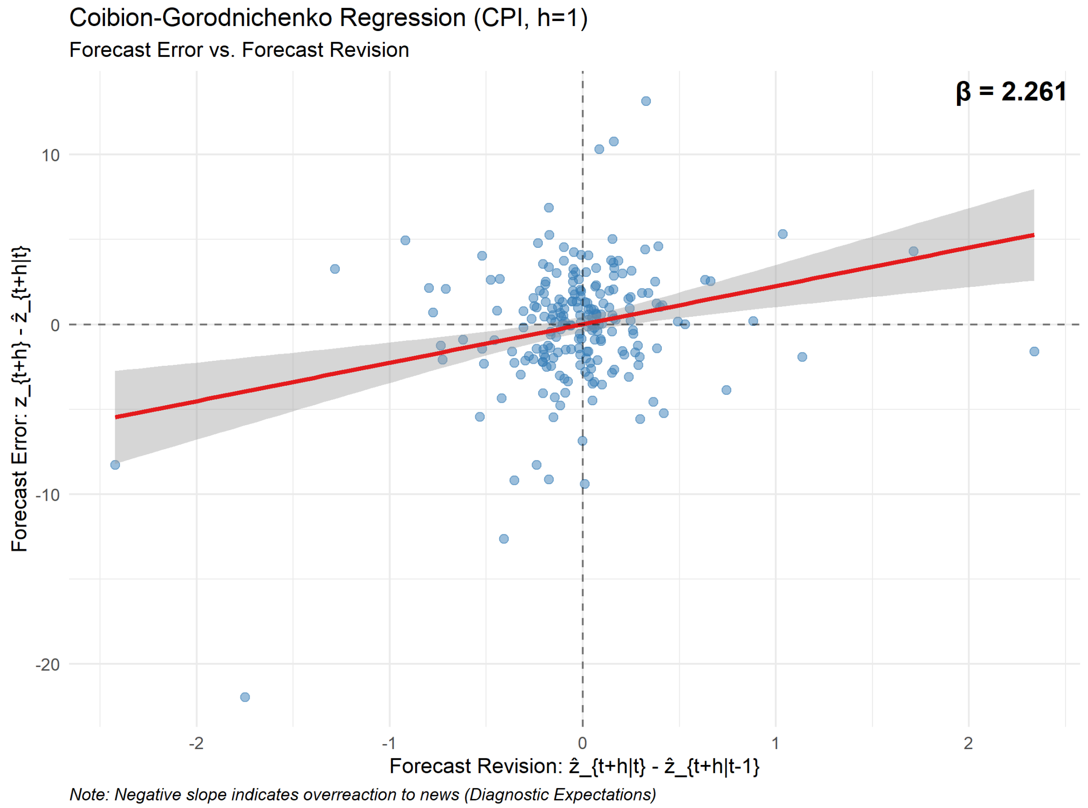

Figure 11: Revision diagnostic scatter: CPI ($h = 1$)

Notes: Scatter of forecast errors against forecast revisions for CPI inflation in the baseline design. The fitted line corresponds to the Coibion and Gorodnichenko (2015) regression.

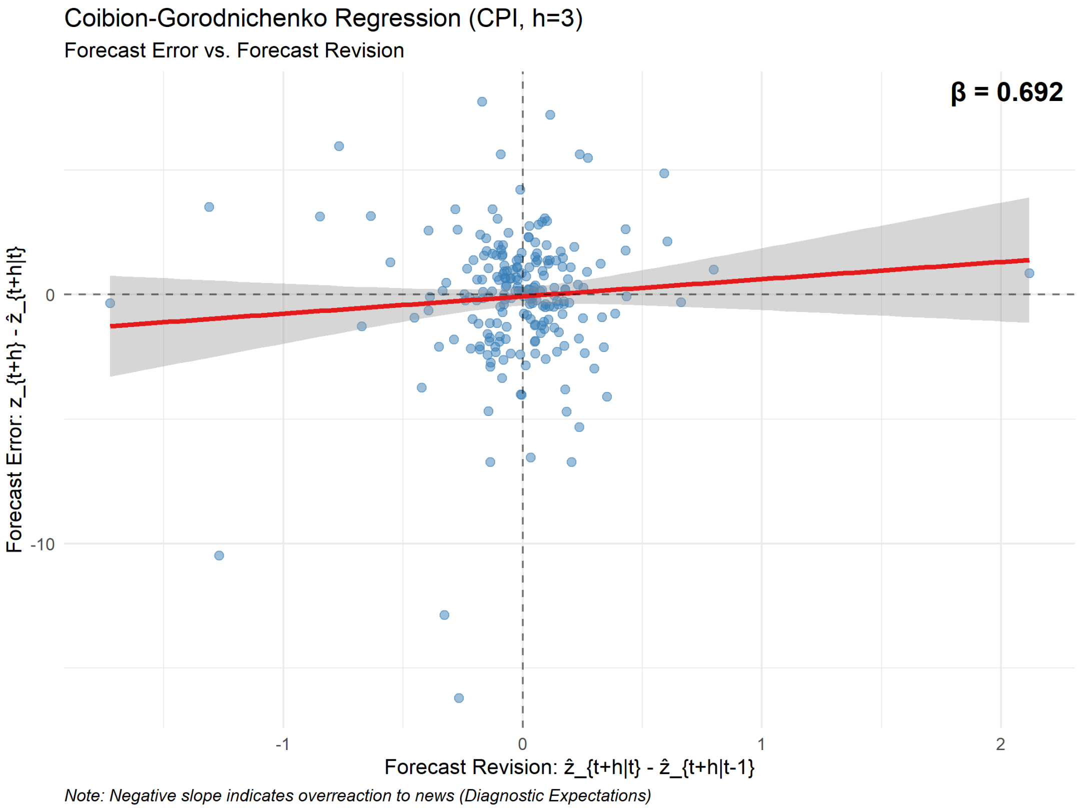

Figure 12: Revision diagnostic scatter: CPI ($h = 3$)

Notes: Scatter of forecast errors against forecast revisions for CPI inflation in the baseline design. The fitted line corresponds to the Coibion and Gorodnichenko (2015) regression.

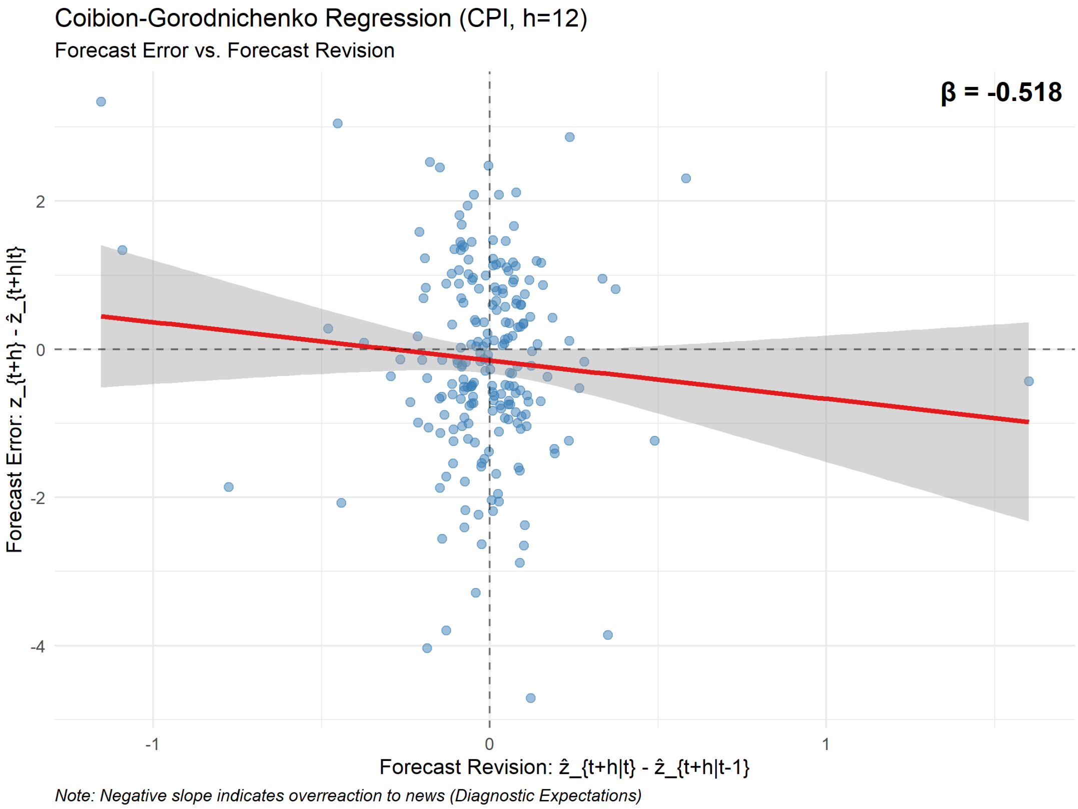

Figure 13: Revision diagnostic scatter: CPI ($h = 12$)

Notes: Scatter of forecast errors against forecast revisions for CPI inflation in the baseline design. The fitted line corresponds to the Coibion and Gorodnichenko (2015) regression.

Table 4: Robustness: RMSFEs under alternative lag length and training window

Scenario	Model	Target	$h = 1$	$h = 3$	$h = 12$
$p = 6$	Small	CPI	3.451	2.639	1.291
$p = 6$	Medium	CPI	3.445	2.641	1.294
$p = 6$	Full	CPI	3.529	2.673	1.342
$p = 6$	Small	INDPRO	7.582	5.410	4.825
$p = 6$	Medium	INDPRO	7.336	5.033	4.621
$p = 6$	Full	INDPRO	7.494	5.173	4.572
Initial window ends 1995M12	Small	CPI	3.231	2.448	1.323
Initial window ends 1995M12	Medium	CPI	3.216	2.446	1.325
Initial window ends 1995M12	Full	CPI	3.267	2.431	1.286
Initial window ends 1995M12	Small	INDPRO	7.329	5.204	4.802
Initial window ends 1995M12	Medium	INDPRO	7.077	4.803	4.590
Initial window ends 1995M12	Full	INDPRO	7.218	4.930	4.453

Notes: Values are taken from `results/robustness/lag6/tables/rmsfe_results.csv` and `results/robustness/window1995/tables/rmsfe_results.csv`.