

Geneva Graduate Institute (IHEID)

Topics in Econometrics

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

Abstract

Does consumer sentiment add predictive content for inflation and real activity once standard macro aggregates and financial prices are already included? I answer this question with a horse race across nested information sets in a hierarchical Bayesian VAR in a recursive pseudo out-of-sample design and a revision-based diagnostic of forecast updating following Coibion and Gorodnichenko (2015). The point-forecast evidence indicates limited incremental accuracy from adding sentiment once financial prices are included (Table 1; Figure 1). The revision diagnostic shows that information sets can shift systematic updating patterns even when point accuracy changes little (Table 2). Throughout, revision-regression coefficients are interpreted as diagnostics of a regularized forecasting system rather than causal evidence about beliefs.

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

1 Introduction

This paper studies a practical forecasting question: does consumer sentiment add incremental predictive content for inflation and real activity once conventional macro aggregates and financial prices are already included? I organize the answer around two objects that matter to forecasting practice: *forecast accuracy* (how close point forecasts are to realizations) and *forecast discipline* (whether the forecasting system revises in an internally coherent way, rather than exhibiting systematic updating patterns).

Contributions and headline evidence. The paper makes two contributions. First, it provides a transparent comparison across nested information sets in a hierarchical BVAR that learns overall shrinkage, allowing dimensional changes without ad hoc tuning (Ba  bura, Giannone, & Reichlin, 2010; Giannone, Lenza, & Primiceri, 2015; Kuschnig & Vashold, 2021). Second, it separates forecast accuracy from forecast-discipline diagnostics: sentiment adds limited incremental accuracy beyond financial prices (Table 1; Figure 1), while revision diagnostics reveal systematic updating patterns that shift with the information set (Table 2). Because the specifications are nested, equal-accuracy tests are treated as suggestive and nested-robust adjustments are reported in the appendix (Clark & McCracken, 2001; Clark & West, 2007). Applied to model-implied forecasts, the Coibion and Gorodnichenko (2015) regression is a diagnostic for a regularized forecasting system rather than a behavioral claim.

Related literature. The revision diagnostic builds on Coibion and Gorodnichenko (2015) and relates to work on belief distortions and diagnostic expectations that emphasizes systematic underreaction and overreaction patterns in expectations; here, the regression is used as a model diagnostic rather than a structural test (Bordalo, Gennaioli, & Shleifer, 2018, 2020). Evidence on whether confidence or sentiment contains incremental forecasting information is mixed once other indicators are included, and real-time evaluations often find limited or unstable gains (Bram & Ludvigson, 1998; Carroll, Fuhrer, & Wilcox, 1994; Croushore, 2005; Ludvigson, 2004). The inflation-forecasting literature emphasizes that parsimonious benchmarks can be difficult to beat and that forecasting relationships shift, motivating cautious interpretation of small differences (Atkeson & Ohanian, 2001; Stock & Watson, 2007). Hierarchical BVAR shrinkage provides a disciplined way to compare information sets of different dimensions without ad hoc tuning, which is central to the design here (Ba  bura et al., 2010; Giannone et al., 2015; Kuschnig & Vashold, 2021).

Roadmap. Section 2 describes the nested information sets and the evaluation setup. Section 3 presents the forecasting system, the role of hierarchical shrinkage, and the accuracy and revision diagnostics. Section 4 reports the evidence, and Section 5 concludes.

Empirical focus. The empirical goal is descriptive: quantify the incremental predictive content of sentiment conditional on macro aggregates and financial prices, and summarize forecast-updating patterns using the revision diagnostic. I avoid structural interpretations of revision-regression coefficients and treat formal comparisons under nesting cautiously. This framing separates incremental information content from any belief-distortion interpretation.

2 Data

This section defines the nested information sets and the evaluation scale used in the forecasting comparison. The dataset combines macro aggregates, a set of widely used financial prices, and a survey-based measure of consumer sentiment. The information sets are nested to isolate incremental information content: *Small* contains macro aggregates, *Medium* adds financial prices, and *Full* further adds consumer sentiment. The comparison between Medium and Full therefore targets whether sentiment contributes beyond information already summarized in market prices.

Following standard BVAR practice, the model is estimated in levels or log-levels (Giannone et al., 2015; Sims, 1980). Forecasts are evaluated on a common growth-rate scale constructed from model-implied level forecasts, using the same transformation as the code pipeline. The output-to-manuscript mapping is audited in `INTERNAL_MAPPING.md`. This structure isolates the incremental role of sentiment while keeping the evaluation scale consistent across information sets.

3 Empirical design

This section describes the forecasting system, the pseudo out-of-sample design, and the diagnostics used to compare nested information sets under hierarchical shrinkage.

3.1 Forecasting system and nested information sets

For each information set, I estimate a reduced-form VAR forecasting system and only change the information set. This isolates the incremental role of forward-

looking prices and sentiment within a common estimation and prediction rule. The forecasting system is

$$y_t = c + \sum_{\ell=1}^p B_\ell y_{t-\ell} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma), \quad (1)$$

where y_t collects the variables in the information set. Because adding variables increases parameter uncertainty even when predictive content is present, I regularize the system with Minnesota-style shrinkage and learn the overall tightness from the data using the hierarchical prior-selection approach of Giannone et al. (2015), as implemented in Kuschnig and Vashold (2021). Importantly, shrinkage is interpreted as statistical regularization of the forecasting system rather than as a proxy for economic frictions.

3.2 Pseudo out-of-sample evaluation

I evaluate performance in a recursive pseudo out-of-sample design with expanding estimation windows. At each forecast origin, the system is re-estimated using all data available up to that origin and then produces point forecasts at the horizons reported in the main accuracy table. This recursion mirrors a real-time workflow while remaining descriptive because it uses revised data rather than real-time vintages.

3.3 Forecast accuracy and nested-model inference

Forecast accuracy is summarized by RMSFE on the common evaluation scale described in Section 2. For target i and horizon h ,

$$\text{RMSFE}_{i,h} = \left(P^{-1} \sum_{t \in \mathcal{T}} (y_{i,t+h} - \hat{y}_{i,t+h|t})^2 \right)^{1/2}.$$

I report RMSFEs (Table 1) and relative RMSFEs versus a parsimonious benchmark (Figure 1). Because the information sets are nested, standard equal-accuracy tests can have nonstandard behavior (Clark & McCracken, 2001); I therefore emphasize magnitudes and stability and report a nested-model-robust adjustment as a robustness check (Clark & West, 2007) (Appendix Table 3).

3.4 Forecast discipline: revision-based diagnostic

To assess whether forecast updates are systematically related to subsequent forecast errors, I use the error-on-revision regression framework of Coibion and Gorodnichenko

(2015) applied to model-implied forecasts:

$$(z_{t,h} - \hat{z}_{t,h|t}^{(m)}) = \alpha_h + \beta_h r_{t,h}^{(m)} + \varepsilon_{t,h}, \quad (2)$$

where $r_{t,h}^{(m)}$ is the revision to the forecast for the same target date made one period apart. In this paper, the regression is used as a diagnostic of the forecasting system's updating rule: it measures whether revisions are followed by predictable errors, indicating systematic patterns in updating. Because the forecasts are produced under shrinkage, such patterns can partly reflect prior-induced conservatism, misspecification, or instability rather than an economic mechanism.

3.5 Implementation map and audit trail

All quantitative evidence in the paper is drawn from existing outputs under `results/` and is mapped to manuscript tables and figures in `INTERNAL_MAPPING.md`. The mapping also records where the hierarchical shrinkage settings and hyperparameter extraction are defined in code, ensuring that the narrative aligns with the audited pipeline without rerunning estimation.

4 Results

This section synthesizes evidence from forecast accuracy and revision diagnostics, then interprets the patterns and limitations within a signal-extraction framework. The accuracy results show that adding financial prices can improve point forecasts relative to a macro-only information set, while adding sentiment on top of prices yields limited incremental gains (Table 1; Figure 1). The revision diagnostic indicates systematic updating patterns that shift with the information set even when point accuracy changes little (Table 2; Figure 2). Because the specifications are nested, these comparisons are read as descriptive magnitudes, with nested-robust checks reported in the appendix (Table 3).

4.1 Forecast accuracy

Table 1 reports RMSFEs by information set, and Figure 1 summarizes the same comparison in relative terms versus a parsimonious benchmark. The comparison between Medium and Full isolates sentiment's incremental contribution conditional on prices, while the comparison between Small and Medium captures the incremental content of financial prices. The evidence indicates limited incremental gains from

sentiment once prices are included, consistent with information overlap across forward-looking indicators. *Notes:* The model labels correspond to nested information sets defined

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

in Section 2. Small includes macro aggregates; Medium adds a small set of financial prices; Full further adds consumer sentiment. All evaluation-scale and pseudo out-of-sample implementation details follow the audited pipeline summarized in `INTERNAL_MAPPING.md`.

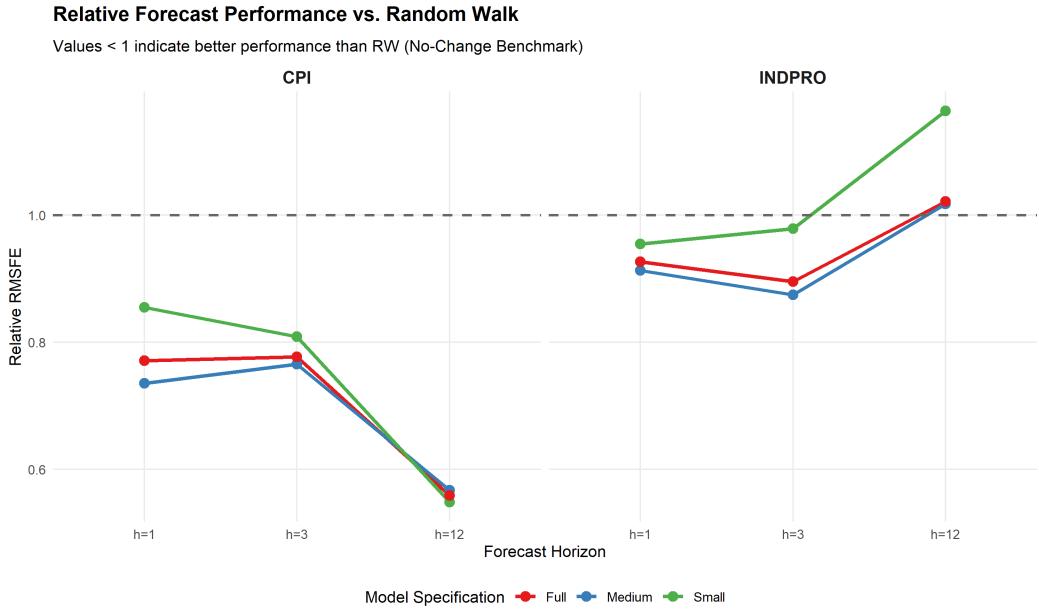


Figure 1: Relative forecast accuracy versus a parsimonious benchmark
Notes: The figure reports RMSFEs for each information set relative to a parsimonious benchmark, using the same evaluation scale as Table 1. See `INTERNAL_MAPPING.md` for source files.

4.2 Forecast discipline: revision-based diagnostic

Table 2 reports error-on-revision coefficients from the Coibion and Gorodnichenko (2015) diagnostic applied to model-implied forecasts, and Figure 2 visualizes the same patterns. For inflation, the coefficients display horizon-dependent sign patterns consistent with underreaction at shorter horizons and overreaction at longer horizons; for real activity, coefficients are closer to zero and imprecise. Interpreted as a

diagnostic of the forecasting system, the coefficients summarize whether revisions are followed by predictable errors, indicating systematic updating patterns. *Notes:*

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

The regression follows Coibion and Gorodnichenko (2015) but is applied to model-implied forecasts, so coefficients are interpreted as a diagnostic of internal error-revision consistency in a regularized forecasting system, not as structural evidence about belief formation. Sign patterns can be read as underreaction or overreaction patterns and can reflect regularization, misspecification, or instability, so the paper emphasizes qualitative shifts across information sets rather than a behavioral mechanism.

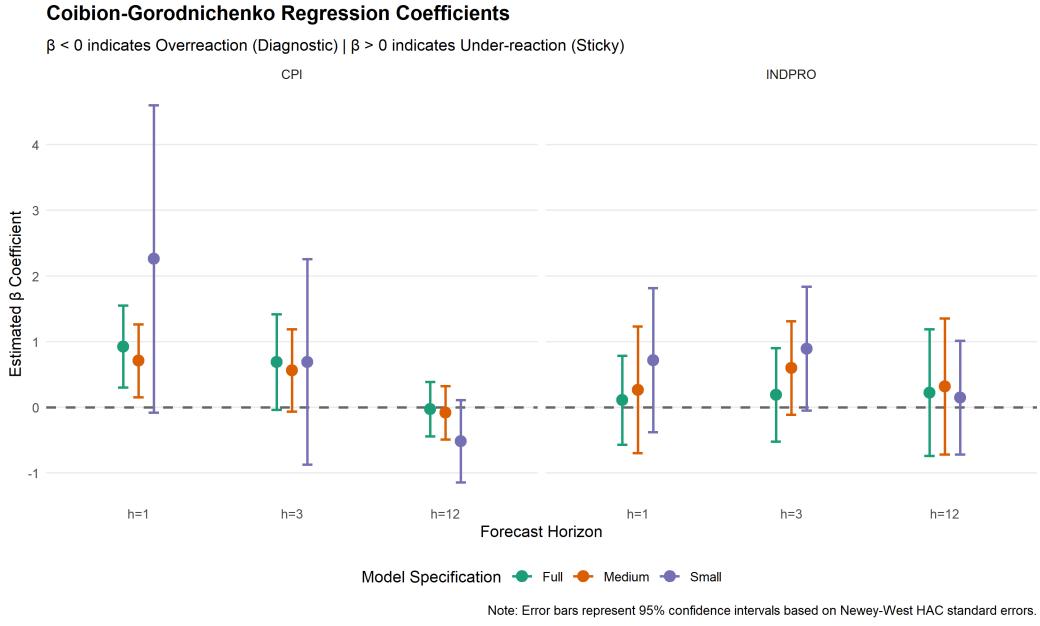


Figure 2: Revision diagnostic coefficients across information sets
Notes: Estimated coefficients from the revision diagnostic. Coefficients summarize internal error-revision consistency in the model-based forecasting system and are interpreted as updating patterns rather than behavioral claims. See INTERNAL_MAPPING.md for source files.

4.3 Regularization and model stability

Figure 3 reports the evolution of the learned shrinkage tightness parameter. The key message is methodological: hierarchical regularization adapts the forecasting system's effective complexity as information sets expand and as the data environment changes, which helps make horse-race comparisons less sensitive to ad hoc tuning choices.

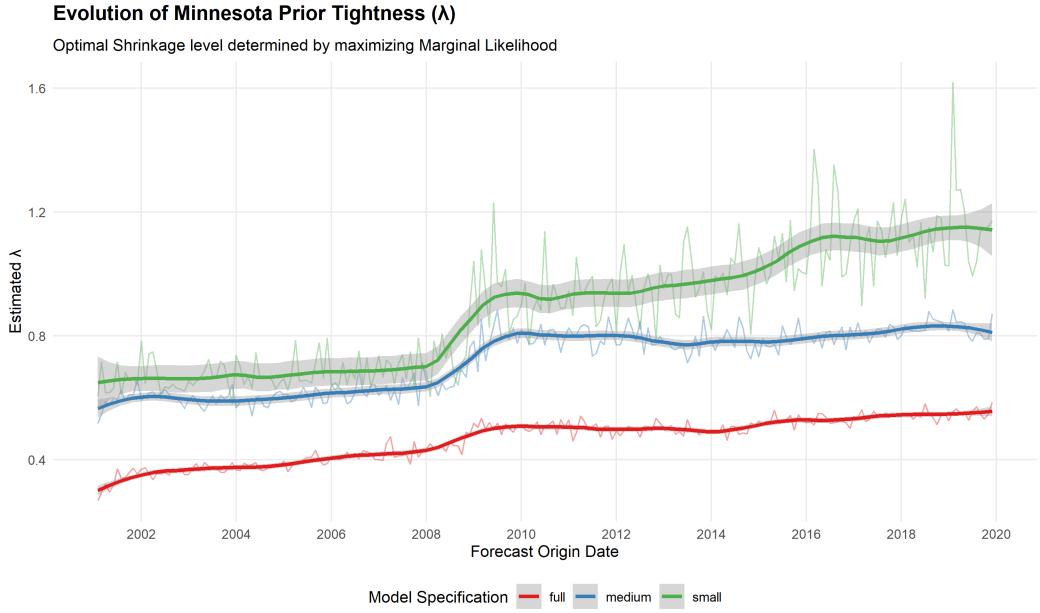


Figure 3: Evolution of learned shrinkage tightness
 Notes: The figure plots posterior means of λ at each recursive forecast origin for each model specification. See INTERNAL_MAPPING.md for source files.

4.4 Economic interpretation and mechanisms

A simple information-aggregation and signal-extraction view helps interpret why sentiment adds limited incremental accuracy once prices are included. Financial prices aggregate dispersed signals and react quickly to news about growth and inflation, so they already embed forward-looking content. Survey sentiment can add information about household plans and expectations, but it is also noisy, subject to framing, and slower-moving. In a linear forecasting system that already includes macro aggregates and prices, the marginal signal from sentiment is therefore small or state dependent, and its incremental contribution to point-forecast accuracy can be limited.

Long-horizon inflation forecasts are dominated by slow-moving trends and persistence, making parsimonious benchmarks competitive. Hierarchical shrinkage encourages persistence and can dampen short-run updates, so revision diagnostics may display underreaction or overreaction patterns that reflect regularization, misspecification, or structural change rather than beliefs. This mechanism view is interpretive rather than identified: the regression is a diagnostic of the forecasting system, and the evidence is consistent with sentiment affecting the pattern of revisions more than average point accuracy.

4.5 Limitations

The revision regression is a diagnostic tool, and its patterns can arise from shrinkage, misspecification, or structural change. The nested design limits power for incremental comparisons, and alternative sentiment measures, real-time data vintages, or density-forecast evaluation could alter the conclusions. Overall, the evidence supports a single narrative: sentiment adds limited incremental point-forecast gains beyond prices, while revision diagnostics reveal systematic updating patterns that vary with the information set.

5 Conclusion

This paper evaluates whether consumer sentiment adds incremental predictive content in a hierarchical BVAR with nested information sets and a revision-based diagnostic. The evidence indicates limited incremental point-forecast gains from sentiment once financial prices are included, while revision diagnostics show systematic updating patterns that vary with the information set (Table 1; Table 2). These findings are descriptive and consistent with information overlap, regularization effects, or changing environments rather than causal claims. Future work can assess state dependence, alternative sentiment measures, real-time data vintages, and density-forecast evaluation.

References

- Atkeson, A., & Ohanian, L. E. (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1), 2–11.
- Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector autoregressions. *Journal of Applied Econometrics*, 25(1), 71–92.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2018). Diagnostic expectations and credit cycles. *American Economic Review*.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2020). Diagnostic expectations and the macroeconomy. *Journal of Economic Perspectives*.
- Bram, J., & Ludvigson, S. (1998). Does consumer confidence forecast household expenditure? a sentiment index horse race. *Federal Reserve Bank of New York Economic Policy Review*, 4(2), 59–78.
- Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? if so, why? *American Economic Review*, 84(5), 1397–1408.
- 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
- Croushore, D. (2005). Do consumer confidence indexes help forecast consumer spending in real time? *Journal of Business & Economic Statistics*.
- 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
- Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic Perspectives*, 18(2), 29–50.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1), 1–48.
- Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to

forecast? *Journal of Money, Credit and Banking*, 39(s1), 3–33.

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). 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: The Clark–West test adjusts mean squared prediction error comparisons to account for nested model structures (Clark & West, 2007). Under the null hypothesis that the smaller (nested) model is adequate, the test statistic has a standard normal distribution. Because specifications are nested and forecasting relationships may be unstable, results are interpreted as suggestive rather than definitive. Source file: see INTERNAL_MAPPING.md.

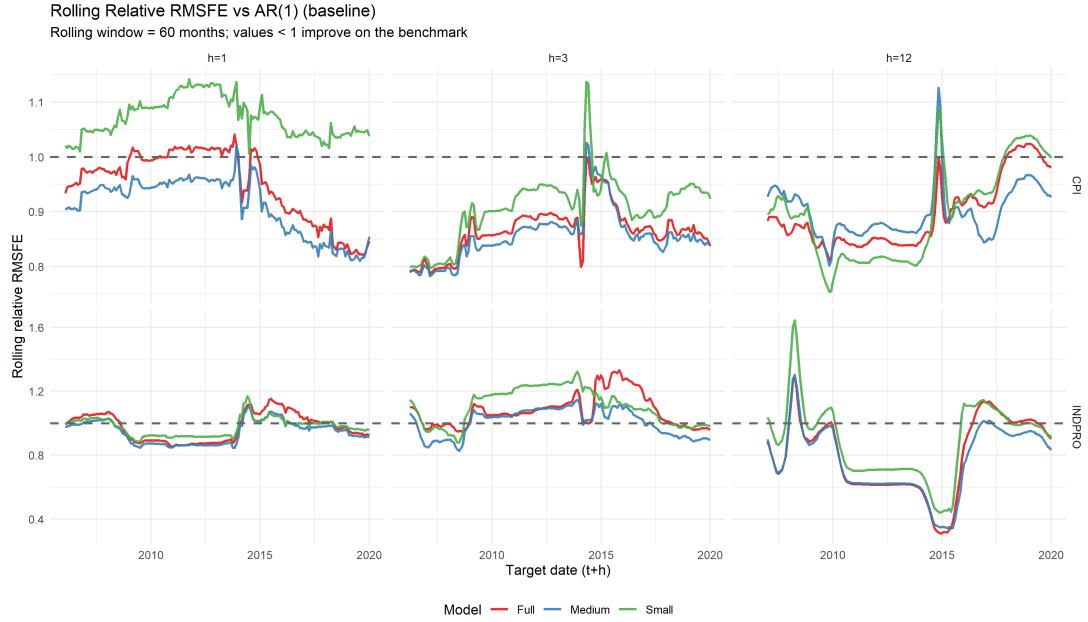


Figure 4: Rolling relative RMSFE versus an autoregressive benchmark.
 Notes: Rolling-window relative RMSFEs. Lower values indicate smaller forecast errors relative to the benchmark. See INTERNAL_MAPPING.md for source files.

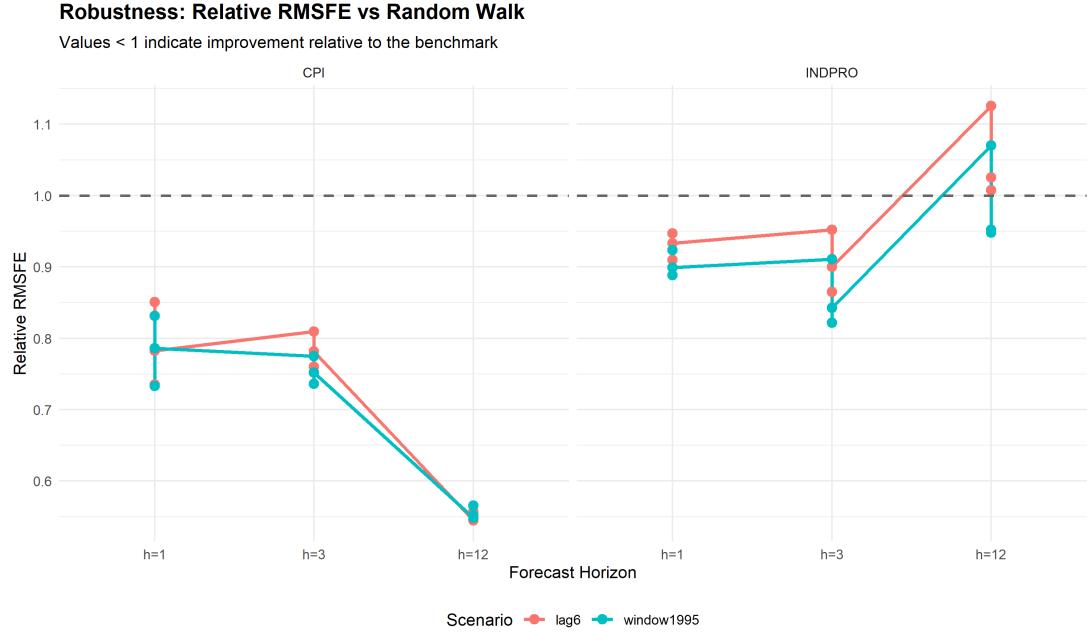


Figure 5: Robustness: relative RMSFE versus no-change benchmark
 Notes: Relative RMSFEs under alternative implementation choices, benchmarked to the same reference forecast. See INTERNAL_MAPPING.md for source files.

Robustness: Relative RMSFE vs AR(1)

Values < 1 indicate improvement relative to the benchmark

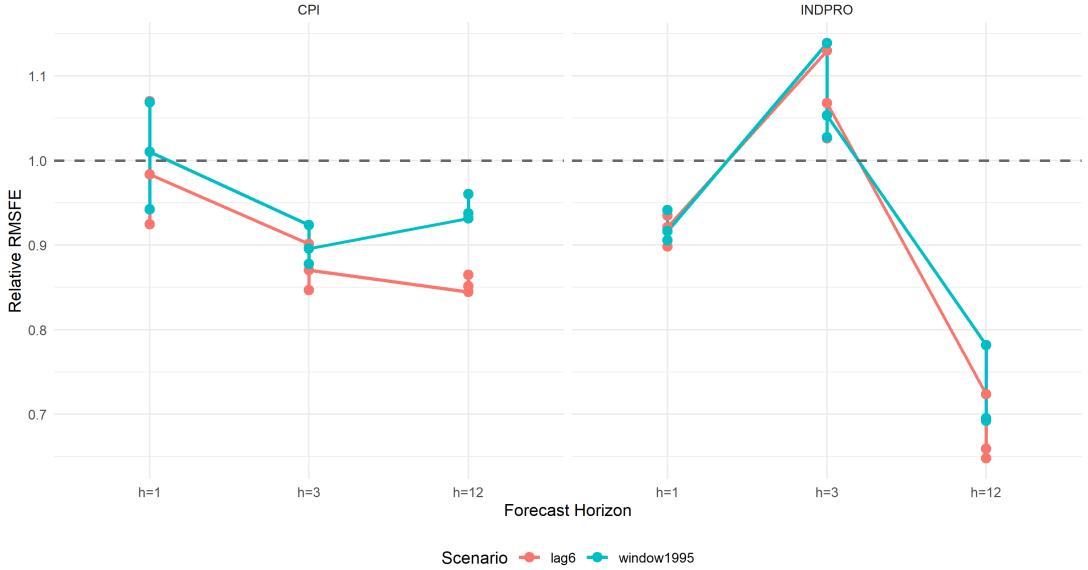


Figure 6: Robustness: relative RMSFE versus an autoregressive benchmark
Notes: Relative RMSFEs under an alternative implementation choice, benchmarked to the autoregressive reference forecast. See INTERNAL_MAPPING.md for source files.

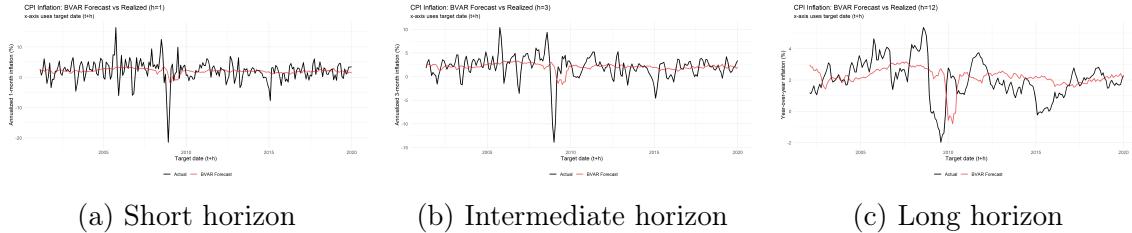


Figure 7: CPI inflation: forecast versus realized (multiple horizons)
Notes: Each panel plots the model-implied predictive mean and the realized target on the evaluation scale. See INTERNAL_MAPPING.md for source files.

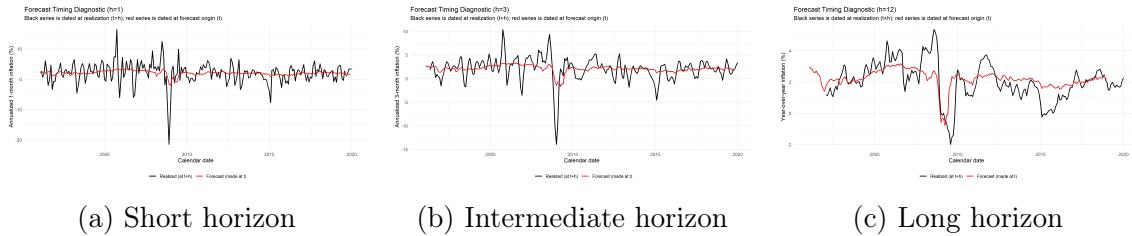


Figure 8: Forecast timing diagnostic (multiple horizons)
Notes: Realizations are dated at the target date; forecasts are dated at the origin date. See INTERNAL_MAPPING.md for source files.

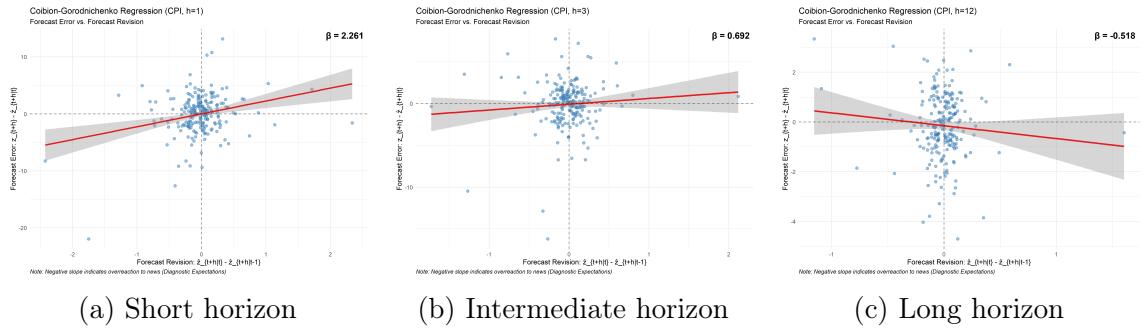


Figure 9: Revision diagnostic scatter (multiple horizons)
 Notes: Scatter of forecast errors against forecast revisions for CPI inflation in the baseline design. The fitted line corresponds to the Coibion and Gorodnichenko (2015) regression. See INTERNAL_MAPPING.md for source files.