Forecasting Horse Races and “Belief Distortions”: A Hierarchical Bayesian VAR Study with Sentiment Signals

Jingle Fu

January 22, 2026

1. Research Question and Motivation

Recent literature on Diagnostic Expectations (DE) suggests that economic agents overreact to news, whereas information rigidity models predict underreaction. This project investigates two precise questions:

1. Does expanding the information set of a hierarchical BVAR to include forward-looking financial prices and consumer sentiment reduce Root Mean Squared Forecast Error (RMSFE) relative to smaller baselines?

2. Whether adding sentiment changes the Coibion-Gorodnichenko(CG) error-revision coefficient in a direction consistent with diagnostic overreaction.

2. Data

I use monthly U.S. macroeconomic series from the FRED-MD database, covering the period **1985M1–2019M12**. The sample ends in 2019 to avoid COVID-19 outliers that would require complex volatility modeling beyond the scope of this term paper. To ensure consistent evaluation, variables are estimated in log-levels (to preserve cointegration) but evaluated in growth rates. The analysis compares three nested information sets:

- **Small Model (Baseline):** Industrial Production (INDPRO), Consumer Price Index (CPIAUCSL), Unemployment Rate (UNRATE), and Federal Funds Rate (FEDFUNDS).
- **Medium Model (Financial Extension):** Adds the 10-Year Treasury Yield (GS10) and S&P 500 Index (S&P500) to capture forward-looking financial cycles.
- **Full Model (Sentiment Extension):** Adds the University of Michigan Consumer Sentiment Index (UMCSENT) to test the marginal predictive power of “soft” data.

2.1 Data Transformation

In time-series analysis, (weak) stationarity is often crucial. The FRED-MD database provides, for each variable, a recommended *transformation code* ("tcode") intended to remove unit roots.[1] However, the BVAR literature (e.g., Sims; and Giannone, Lenza, and Primiceri, 2015) typically favors estimating the model in **levels** or **log-levels**. The key reason is that the Minnesota prior includes a "unit-root prior" (setting $\delta_i = 1$), which is designed explicitly for nonstationary data: by shrinking coefficients toward a random-walk specification, the prior allows the model—when supported by the data—to capture long-run comovement and potential cointegration relationships. If one differences the data mechanically, stationarity is ensured, but valuable long-run equilibrium information (e.g., the long-run link between interest rates and inflation) may be lost.

Accordingly, we adopt the following strategy:

- **Estimation stage.**
  - `INDPRO`, `CPIAUCSL`, `S&P 500`: use log-levels,
  $$x_t = \ln(X_t).$$
  - `UNRATE`, `FEDFUNDS`, `GS10`, `UMCSENT`: keep in levels (no logs), since these are already rates or index-type series.
- **Forecast evaluation stage.** The model produces forecasts in levels or log-levels. For forecast evaluation, we transform forecasts into annualized growth rates (for `CPIAUCSL` and `INDPRO`) or into changes in levels, so that our RMSFE comparisons align with standard practice in the literature.

For example, inflation forecast errors based on `CPIAUCSL` can be computed using the annualized monthly inflation rate

$$\pi_t^{\text{ann}} = 1200 \times \ln\left(\frac{P_t}{P_{t-1}}\right),$$

or alternatively using year-over-year inflation,

$$\pi_t^{\text{yoy}} = 100 \times \ln\left(\frac{P_t}{P_{t-12}}\right).$$

[1] See the FRED-MD documentation for the definition of tcodes and their recommended transformations.

Analogous transformations are applied to `INDPRO` to obtain annualized growth rates for forecast evaluation.

2.2 Implementation with R

We employ the R package `BVAR` developed by Kuschnig and Vashold (2021), which provides an authoritative and convenient implementation of the hierarchical prior framework in Giannone, Lenza, and Primiceri (2015). [2]

Step 1: Prior Setup We specify priors using the function `bv_priors()`.

- **Minnesota prior.** We place a distribution on the key shrinkage hyperparameter λ. Unlike the conventional approach that fixes λ (e.g., `lambda = 0.2`), we set a *mode* and *standard deviation*, and use the data to select the optimal λ via marginal likelihood maximization.

- **Dummy observations.** We include *sum-of-coefficients* and *dummy-initial-observation* priors to further help the model accommodate unit-root behavior and nonstationarity. These components are also treated hierarchically and are automatically tuned within the estimation.

- **Lag length.** We fix the lag order at $p = 12$ (12 monthly lags). Rather than truncating higher-order lags manually, we rely on the strong shrinkage implied by the hierarchical prior to control the parameters on distant lags.

Step 2: Recursive Pseudo-Out-of-Sample Forecasting To mimic a real-time forecasting environment, we conduct recursive forecasting with an expanding window:

- **Initial window:** 1985M1–2000M12 (approximately 15 years of data for initial training).

- **Recursive loop:** At each forecast origin T (starting at 2000M12),

  1. estimate the model using data from 1985M1 through T, jointly optimizing hyperparameters;
  2. produce $h = 1, 3, 12$ step-ahead forecasts;
  3. expand the sample by one month to $T + 1$, re-estimate the model, and re-optimize the hyperparameters.

[2]See Kuschnig and Vashold (2021) for details on the `BVAR` package and its implementation of hierarchical shrinkage priors.

- **Output:** This procedure yields three forecast time series (for $h = 1$, $h = 3$, and $h = 12$), covering roughly 230 forecast origins over 2001–2019.

3. Econometric Framework

The core methodology relies on a reduced-form VAR estimated with a Minnesota-style Normal-Inverse-Wishart prior.

1. **Hierarchical BVAR:** Let y_t be the vector of endogenous variables. We estimate three nested BVAR systems with $p = 12$ lags:

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

   across three distinct information sets. Shrinkage is selected endogenously by treating λ as a hyperparameter with a hyperprior and choosing it by marginal likelihood. We compute RMSFE for $h = 1, 3, 12$ and report RMSFE ratios relative to the benchmark, together with Diebold-Mariano tests for pairwise comparisons. Benchmarks are a random-walk-type forecast and an AR(1) forecast defined on the same evaluation transforms used for the BVAR outputs.

2. **Identification of "Behavioral" Bias:** Let z_{t+h} denote the realized growth rate of a target variable ($z \in \{\text{INDPRO}, \text{CPI}\}$) at horizon h. Let $\hat{z}^{(m)}_{t+h|t}$ denote the forecast generated by Model $m \in$ {Small, Med, Full} at time t. We estimate the following regression linking ex-post forecast errors to forecast revisions:

$$(z_{t+h} - \hat{z}^{(m)}_{t+h|t}) = \alpha_h + \beta_h(\hat{z}^{(m)}_{t+h|t} - \hat{z}^{(m)}_{t+h|t-1}) + \varepsilon_{t+h} \tag{2}$$

   where the term in the first parenthesis represents the forecast error and the term in the second parenthesis represents the forecast revision. Since the forecast horizon h creates overlapping observations, inference on β_h relies on Newey-West HAC standard errors.

3.1 Minnesota prior and hierarchical tightness

Stacking observations yields $Y = X\Phi + U$, where Φ collects $(c, B_1, \ldots, B_p)$. I impose a Minnesota-style Gaussian prior on Φ conditional on Σ:

$$\text{vec}(\Phi) \mid \Sigma, \lambda \sim \mathcal{N}(\text{vec}(\underline{\Phi}), \Sigma \otimes \underline{\Omega}(\lambda)), \qquad \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}),$$

where $\underline{\Phi}$ encodes the random-walk / near-random-walk belief on own first lags, and $\underline{\Omega}(\lambda)$ implements lag decay and cross-variable shrinkage. In particular, for coefficient $(B_\ell)_{ij}$,

$$\mathbb{V}\left[(B_\ell)_{ij} \mid \lambda\right] = \begin{cases} \lambda^2/\ell^2, & i = j, \\ (\lambda^2/\ell^2) \cdot (\sigma_i^2/\sigma_j^2), & i \neq j, \end{cases}$$

with σ_i^2 set from residual scales in univariate AR benchmarks.

The key departure from ad hoc calibration is that the overall tightness λ is *endogenized.* Following Giannone–Lenza–Primiceri, I treat λ as a hyperparameter selected by the data through the marginal data density (MDD),

$$p(Y \mid \lambda) = \int p(Y \mid \Phi, \Sigma)\, p(\Phi, \Sigma \mid \lambda)\, d\Phi\, d\Sigma,$$

and use the `BVAR` package implementation to obtain $\hat{\lambda}$ (and other prior hyperparameters) via hierarchical prior selection. Conjugacy implies that, conditional on λ, the posterior is Normal–Inverse-Wishart, which yields closed-form posterior moments and a tractable posterior predictive distribution for y_{t+h}.

We first give λ a Gamma hyperprior, $\lambda \sim \mathcal{G}(a, b)$, and based on data Y, we compute the best posterior mode $\hat{\lambda}$ by maximizing $p(Y \mid \lambda)p(\lambda)$ over a grid of λ values. Then, conditional on $\hat{\lambda}$, we obtain the posterior of (Φ, Σ) and the predictive distribution of y_{t+h}.

3.2 Pseudo out-of-sample forecasting and evaluation

I implement an expanding-window pseudo out-of-sample exercise. The initial estimation window is 1985M1–2000M12; I then recursively re-estimate and forecast through 2019M12, generating predictive means for $h \in \{1, 3, 12\}$.

Forecast accuracy. For target i and horizon h, compute RMSFE,

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

and report relative RMSFEs versus RW/AR(1). I will conduct Diebold–Mariano tests for differences in predictive loss between Small and Medium specifications.

3.3 Evaluation of Forecasting Competition

We evaluate forecasting performance using the following metrics:

- **Absolute accuracy:** For each model and forecast horizon, we compute the root mean squared forecast error (RMSFE).

- **Relative accuracy:** We normalize each model's RMSFE by the RMSFE of a random-walk (RW) benchmark to obtain the *relative RMSFE*:

$$\text{RelRMSFE}_{m,h} \;=\; \frac{\text{RMSFE}_{m,h}}{\text{RMSFE}_{\text{RW},h}}.$$

  Values below one indicate that model m outperforms the benchmark at horizon h.

- **Statistical significance:** We apply the Diebold–Mariano (DM) test to assess whether the medium-scale BVAR delivers significantly better predictive accuracy than (i) the small-scale BVAR and (ii) univariate benchmark models.

4. Interpretation and Expected Results

First, We verify whether the Hierarchical BVAR outperforms AR(1) benchmarks. The key test is whether the *Full Model* lowers RMSFE for real activity variables at short horizons ($h = 1, 3$) (and at long horizens $h = 12$).

Our key object is $\Delta\beta_h$. A shift in β_h *toward zero* when adding UMCSENT is interpreted as sentiment improving the informational content of revisions, making updates less systematically biased. By contrast, if this pushes β_h *further away from zero*—especially into negative—it suggests that sentiment generating an overreaction pattern consistent with the diagnostic-expectations sign prediction.