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

Jingle Fu

December 13, 2025

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.

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. To test the efficiency of the algorithmic forecasts, I 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.

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 $h = 12$). Then, by comparing β_h across the three nested models, we can isolate the marginal effect of sentiment. The key empirical object is the change in β_h when adding sentiment (Medium-Finance vs Medium-Finance+Sentiment), which might connect to diagnostic expectations.