

Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.

ISSN 1936-2854 (Print)

ISSN 2767-3898 (Online)

Bells and Whistles of Nowcasting Models: Which Matter and When?

Freddy García-Albán and Manuel González-Astudillo

2026-068

Please cite this paper as:

García-Albán, Freddy, and Manuel González-Astudillo (2026). “Bells and Whistles of Nowcasting Models: Which Matter and When?,” Finance and Economics Discussion Series 2026-068. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2026.068>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

Bells and Whistles of Nowcasting Models: Which Matter and When?*

Freddy García-Albán[†]

Manuel González-Astudillo[‡]

September 18, 2026

Abstract

This paper investigates which features of a Bayesian dynamic factor model improve U.S. GDP nowcasts and whether their contribution varies across historical episodes. Using pseudo-real-time information sets for 25 series, we estimate 16 combinations of dynamic heterogeneity, stochastic volatility, time-varying long-run growth, and a multiplicative outlier adjustment. We evaluate their marginal contributions to point, density, and quantile score forecasts using balanced paired factorial contrasts. Dynamic heterogeneity improves point, density, and tail forecasts, and it is the only feature that helps consistently across all three criteria. Stochastic volatility improves density and tail forecasts in some periods but not point forecasts consistently. Time-varying long-run growth lowers squared-error loss in some periods but worsens tail forecasts. The outlier adjustment does not improve point or density forecasts overall, has favorable but imprecisely estimated effects on upper-tail quantile scores during the pandemic, and worsens those scores afterward.

Keywords: Bayesian dynamic factor models; forecast evaluation; stochastic volatility; long-run growth; outliers.

JEL classification: C11; C32; C53; E37.

*We thank Gianni Amisano for helpful discussions and feedback. All remaining errors are our own.

[†]Universidad Ecotec, Km 13 1/2 Vía Samborondón, Samborondón, Guayas, Ecuador. Email: fgarcia@ecotec.edu.ec. Corresponding author.

[‡]Board of Governors of the Federal Reserve System, Washington, D.C., USA and Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador. Email: manuel.p.gonzalez-astudillo@frb.gov. The views expressed are those of the authors and do not necessarily reflect those of the Board of Governors or the Federal Reserve System.

1 Introduction

Which bells and whistles of a modern nowcasting model actually improve its performance, and during which historical periods? Bayesian dynamic factor models used for nowcasting have recently been enriched along several dimensions, each tied to a specific feature of the macroeconomy. Dynamic heterogeneity relates each high-frequency observable variable of the model to the common factor and its lag, accommodating differences in the timing with which leading and lagging variables respond to the business cycle. Stochastic volatility lets the variance of the model’s shocks change over time, a natural response to episodes in which uncertainty about the state of the economy rises and falls abruptly. Time-varying long-run growth lets trend GDP growth evolve gradually instead of staying fixed, capturing persistent slowdowns or accelerations. A multiplicative outlier adjustment accommodates unusually large observations without letting them distort the estimated volatility of the rest of the sample. Each of these features has a clear economic or statistical rationale, but each also adds complexity and computational cost, and it is not obvious that any of them, alone or in combination, improves nowcasts in practice.

Although some of these features such as dynamic heterogeneity and time-varying long-run growth had already been proposed prior to the COVID-19 pandemic period (see [D’Agostino et al., 2016](#); [Antolín-Díaz et al., 2017](#), for instance), the extreme movements in macroeconomic data during 2020 renewed interest in stochastic volatility and methods robust to outliers, capable of handling abrupt changes in volatility and fat tails (see [Antolín-Díaz et al., 2024](#); [Carriero et al., 2024](#), among others). This raises a natural question: are the gains from these features broad based, or concentrated in particular episodes such as the pandemic?

A large body of research uses mixed-frequency factor models for nowcasting. The frameworks of [Giannone et al. \(2008\)](#) and [Banbura et al. \(2013\)](#) provide the basis for many of these systems. Subsequent work has incorporated Bayesian estimation and stochastic volatility ([Carriero et al., 2015](#)), real-time data-flow considerations, heterogeneous factor dynamics, and explicit treatments of pandemic outliers and fat tails (see [Almuzara et al., 2023](#); [Antolín-Díaz et al., 2024](#); [Hong, 2025](#), for instance). These studies typically introduce and evaluate one such feature at a time, against a single benchmark specification. This makes it difficult to know whether a given feature would still help once combined with the others, or whether its apparent usefulness is tied to the particular sample period on which it was evaluated.

Most studies in this literature also compare a subset of complete model specifications over whole sample periods. [Carriero et al. \(2019\)](#), for instance, summarize a large number of point- and density-forecast comparisons across model classes, specifications, forecast horizons, and evaluation samples. These comparisons are useful for choosing a forecasting

model, but they do not reveal which specific element of a richer specification drives any improvement in performance. We contribute to this literature in two ways. First, rather than evaluating dynamic heterogeneity, stochastic volatility, time-varying long-run growth, and the multiplicative outlier adjustment one at a time, we consider all four jointly and use a full factorial design that isolates each feature’s marginal contribution independently of any single benchmark choice. Second, we assess these marginal contributions using point, density, and quantile-score criteria, separately for the periods before, during, and after the pandemic, tying each feature’s usefulness to a specific forecasting objective and historical episode rather than to average performance alone.

We estimate all $2^4 = 16$ combinations of these four features. This full factorial structure allows us to compare two models that differ only in one feature while keeping the remaining three features fixed. We then average those paired comparisons across the eight possible configurations of the other features to obtain a measure of the marginal contribution of each modeling choice that does not depend on the selection of a single benchmark model.

Our forecast-evaluation object therefore differs from the conventional pairwise loss differential used in the literature following [Diebold and Mariano \(1995\)](#) and [Giacomini and White \(2006\)](#). For every target quarter and forecast horizon, we compute a balanced average of paired differences associated with turning a particular feature on and off, using log-loss differences for squared forecast error and the continuous ranked probability score (CRPS), and quantile-score differences in levels. We then average those contrasts over the full evaluation sample and over three periods: before the pandemic, during the pandemic, and after the pandemic. To account for uncertainty, we bootstrap target quarters while keeping together all seven forecast horizons associated with a given quarter.

Our results from a pseudo-real-time forecasting exercise over target quarters from 2015Q1 to 2024Q4 indicate that dynamic heterogeneity delivers the clearest improvement in point, density, and quantile score forecasts: it lowers squared-error loss by 16.8 percent and CRPS by 7.6 percent on average, and it is the only feature with a precisely estimated improvement in an overall quantile score, cutting the 90th-percentile score by 6.2 percent. Stochastic volatility also tends to improve forecasting overall and particularly prior to the pandemic, when it reduces squared-error loss, CRPS, and both tail quantile scores, although the statistical evidence is less conclusive than for dynamic heterogeneity. Time-varying long-run growth lowers squared-error loss by 13.3 percent on average, but the effect is estimated with substantial uncertainty, does not extend to CRPS, and comes at a cost: it raises the lower-tail quantile scores by 3 to 4 percent. The multiplicative outlier adjustment improves neither average point nor density forecast performance; its apparent overall gains of more than 20 percent in the upper-tail quantile scores are imprecisely estimated. Looking across historical

periods, dynamic heterogeneity and stochastic volatility show their most noticeable gains before the pandemic, whereas time-varying long-run growth produces its largest improvement in point forecasts after the pandemic. Of note, the multiplicative outlier adjustment, despite being particularly motivated by unusually large observations, does not show a clear benefit during the pandemic in terms of point or density forecasts, but its point estimate for the upper-tail quantile scores does improve during that period; after the pandemic, its point estimates turn detrimental across squared-error loss, CRPS, and the upper-tail quantile scores alike.

The paper is organized as follows. Section 2 describes the data. Section 3 presents the nowcasting model and the four features under consideration. Section 4 discusses the pseudo-real-time forecasting exercise, whereas Section 5 describes the balanced paired factorial contrasts used to evaluate the marginal contribution of each feature. Section 6 presents the results, and Section 7 concludes.

2 Data

We use a data set with 25 U.S. series constructed as a subset of the New York Fed Staff Nowcast data environment described by [Almuzara et al. \(2023\)](#). The data include three quarterly variables and a set of monthly indicators that become available at different points within the quarter. Table 1 reports the complete list of variables, their transformations, and frequencies.¹

Following [Almuzara et al. \(2023\)](#), the estimation sample begins in 1985 and ends in December 2024, with series that have shorter histories entering the panel as they become available. For forecast evaluation, we use revised observations from the latest available vintage of each series, downloaded at the end of 2025. Monthly pseudo-real-time information sets are then constructed by masking observations according to their publication lags. Thus, the exercise reproduces the month-to-month ragged edge implied by these lags, but not real-time data revisions. We transform and standardize all series before estimation. Figure A.1 in the Appendix shows the evolution of each of the variables used in the estimation of the model.

¹We retain the publicly available indicators of real activity from the New York Fed Staff Nowcast data set. We exclude its monthly price indicators to keep mostly economic activity variables and omit the proprietary ISM series so that the panel can be reproduced from public sources.

Table 1: **Variables used in the nowcasting exercise**

Variable	Frequency	Units	Publication lag
Real gross domestic product	Q	QoQ % ch., ann.	1
Real gross domestic income	Q	QoQ % ch., ann.	2
Nonfarm business sector unit labor cost	Q	QoQ % ch., ann.	2
All employees: total nonfarm	M	Level ch., thous.	1
JOLTS: total job openings	M	Level ch., thous.	2
Civilian unemployment rate	M	Ppt. ch.	1
ADP nonfarm private payroll employment	M	Level ch., thous.	1
Industrial production index	M	MoM % ch.	1
Manufacturers' new orders: durable goods	M	MoM % ch.	1
Merchant wholesalers' inventories: total	M	MoM % ch.	2
Total business inventories	M	MoM % ch.	2
Manufacturers' shipments: durable goods	M	MoM % ch.	2
Manufacturers' unfilled orders: all industries	M	MoM % ch.	2
Manufacturers' inventories: durable goods	M	MoM % ch.	2
Retail sales and food services	M	MoM % ch.	1
Real personal consumption expenditures	M	MoM % ch.	1
New single-family houses sold	M	MoM % ch.	1
Housing starts	M	MoM % ch.	1
Value of construction put in place	M	MoM % ch.	2
Building permits	M	Level ch., thous.	1
Exports: goods and services	M	MoM % ch.	2
Imports: goods and services	M	MoM % ch.	2
Real disposable personal income	M	MoM % ch.	1
Philadelphia Fed Manufacturing Business	M	Index	0
Outlook: current activity			
Empire State Manufacturing Survey: general business conditions	M	Index	0

Notes: Q denotes quarterly and M denotes monthly. QoQ and MoM denote quarter-on-quarter and month-on-month changes, respectively; ann. denotes annualized, ch. change, thous. thousands, and ppt. percentage points. Publication lags are measured in months from the end of the reference period. A lag of 0 indicates availability within the reference month, while lags of 1 and 2 indicate availability one and two months later, respectively.

3 Nowcasting Model

Let y_t be an $n \times 1$ vector containing the monthly variables and the (latent) monthly representation of the quarterly variables. We specify the measurement equation as

$$y_t = c_t + \Lambda(L)f_t + u_t, \quad (1)$$

where c_t contains deterministic or slowly moving long-run growth components, f_t is a common factor, $\Lambda(L)$ is a lag polynomial of factor loadings, and u_t contains the variable-specific idiosyncratic components.

The common factor follows a stationary AR(2) process, as follows:

$$(1 - \phi(L))f_t = \varepsilon_t, \quad \varepsilon_t \sim N(0, \exp(h_{\varepsilon,t})), \quad (2)$$

while the idiosyncratic component of each variable i also follows stationary AR(2) dynamics and satisfies

$$(1 - \rho_i(L))u_{i,t} = \sqrt{o_{i,t}}\eta_{i,t}, \quad \eta_{i,t} \sim N(0, \exp(h_{\eta_{i,t}})), \quad i = 1, \dots, n, \quad (3)$$

where $o_{i,t}$ is a strictly positive multiplicative outlier.²

The first feature we consider in our evaluation of model features is dynamic heterogeneity. Building on [D’Agostino et al. \(2016\)](#) and adopting the parsimonious one-lag specification of [Antolín-Díaz et al. \(2024\)](#), we allow the common factor to enter the measurement equations contemporaneously and with one lag, as indicated below:

$$\Lambda(L) = \Lambda_0 + \Lambda_1 L.$$

In the restricted setup, $\Lambda(L) = \Lambda_0$. We normalize the contemporaneous loading of the factor in the real GDP equation to one.

The second feature is stochastic volatility. When it is included, the log volatilities of the common-factor and idiosyncratic shocks evolve as driftless random walks, as in [Primiceri](#)

²We use the multiplicative specification of [Hong \(2025\)](#) rather than the additive outlier component in [Antolín-Díaz et al. \(2024\)](#). Because the idiosyncratic components are autocorrelated, quasi-differencing an additive outlier makes its effective variance depend mechanically on the idiosyncratic autoregressive coefficients; the multiplicative specification applies the adjustment directly to the innovation variance and avoids this unintended dependence.

(2005):

$$h_{\varepsilon,t} = h_{\varepsilon,t-1} + v_{\varepsilon,t}, \quad v_{\varepsilon,t} \sim \text{i.i.d. } N(0, \omega_{v_\varepsilon}^2), \quad h_{\varepsilon,0} \sim N(\mu_{h_{\varepsilon,0}}, \sigma_{h_{\varepsilon,0}}^2) \quad (4)$$

$$h_{\eta_i,t} = h_{\eta_i,t-1} + v_{\eta_i,t}, \quad v_{\eta_i,t} \sim \text{i.i.d. } N(0, \omega_{v_{\eta_i}}^2), \quad h_{\eta_i,0} \sim N(\mu_{h_{\eta_i,0}}, \sigma_{h_{\eta_i,0}}^2). \quad (5)$$

In the restricted specification, the volatilities of ε_t and $\eta_{i,t}$ are constant over time.

The third feature allows long-run growth to vary over time. Following [Antolín-Díaz et al. \(2017\)](#), we specify

$$c_t = c + bg_t,$$

where g_t is a common long-run growth component that enters the real GDP and real GDI equations. The remaining variables do not load on g_t .³ Time-varying long-run growth has also been incorporated into other nowcasting models to allow for persistent changes in trend growth and has been shown to improve their performance (see [González-Astudillo and Baquero, 2019](#); [Almuzara et al., 2023](#); [Antolín-Díaz et al., 2024](#), for example). In our model,

$$g_t = g_{t-1} + \zeta_t, \quad \zeta_t \sim \text{i.i.d. } N(0, \omega_g^2). \quad (6)$$

Finally, we consider a multiplicative adjustment for unusual observations. When this feature is inactive, $o_{i,t} = 1 \forall i, t$. When it is active, we follow [Hong \(2025\)](#) and assume that the standardized idiosyncratic innovation, $\sqrt{o_{i,t}}\eta_{i,t}/\sqrt{\exp(h_{\eta_i,t})}$, follows a Student- t distribution with ν_i degrees of freedom.⁴ We use the equivalent normal-inverse-gamma representation for estimation:

$$o_{i,t} \mid \nu_i \sim \text{IG}\left(\frac{\nu_i}{2}, \frac{\nu_i}{2}\right), \quad i = 1, \dots, n. \quad (7)$$

Equivalently, $\nu_i/o_{i,t} \sim \chi_{\nu_i}^2$.

We estimate the model with Bayesian methods. For every specification and monthly pseudo-real-time information set, we discard the first 10,000 Gibbs-sampler iterations as burn-in and retain every 50th subsequent draw until 1,000 posterior draws are obtained. We use Gaussian shrinkage priors for the factor-loading and autoregressive coefficients and inverse-gamma priors for the variance parameters. The Gibbs sampler uses the simulation smoother of [Durbin and Koopman \(2002\)](#) and alternates among the factors, loadings, au-

³Here, c is the vector of time-invariant intercepts and b is the selection vector with $b_i = 1$ for real GDP and real GDI and zero otherwise; in the standardized empirical implementation, $c = 0$.

⁴An alternative formulation of multiplicative outliers is the uniform mixture proposed by [Stock and Watson \(2016\)](#). [Hong \(2025\)](#) shows that the Student- t approach tends to identify outliers too frequently, leading to an underestimation of the stochastic volatility processes, whereas the uniform mixture misses moderate outliers and understates their frequency. Hong proposes a hybrid model that combines these two approaches. Such a methodology is not the object of study in this paper.

toregressive parameters, stochastic volatilities (using the [Kim et al. \(1998\)](#) sampler), outliers and degrees of freedom (following [Jacquier et al. \(2004\)](#)), as applicable to each model. We handle quarterly variables at monthly frequency using the aggregation identity of [Mariano and Murasawa \(2003\)](#). Following [Hong \(2025\)](#), we impose the aggregation restriction on the latent quarterly idiosyncratic components instead of directly on the observed quarterly variables. This specification allows for a faster Gibbs sampler. Sections [A.2–A.6](#) of the Appendix describe the mixed-frequency state-space representation, the priors, and the estimation algorithm in more detail.

4 Forecasting Design

We evaluate the four binary features described in the previous section using all 16 possible model specifications. Hence, unlike an exercise that adds each feature sequentially to a particular benchmark, our design evaluates each feature across every possible configuration of the remaining ones. Let s index each model specification. Figure [1](#) lists the compact specification labels used throughout the paper and how the features vary from one specification to another.

We conduct a pseudo-real-time forecasting exercise. The evaluation unit is a forecast task $\tau = (q, h)$, where q is the target quarter and h is the monthly horizon relative to the end of that quarter, with the sign convention

$$h \in \{-5, -4, -3, -2, -1, 0, +1\},$$

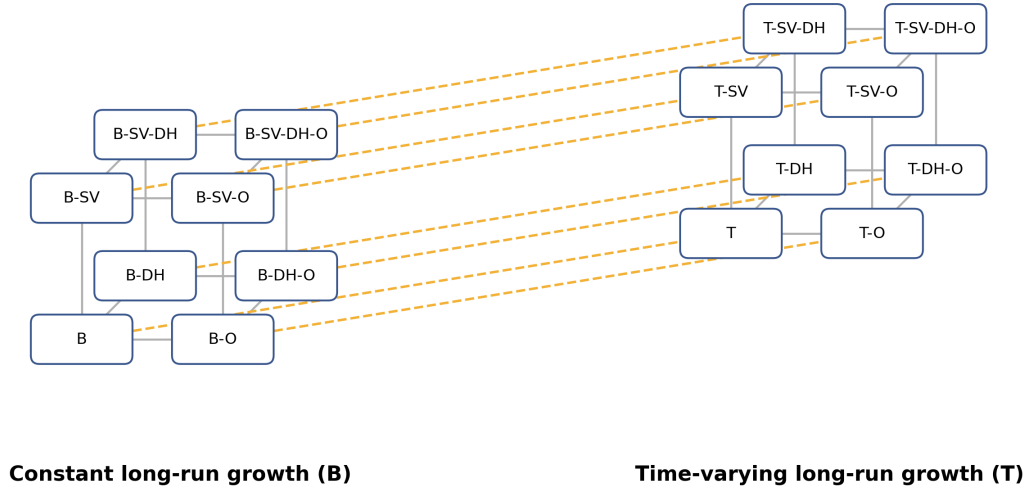
where $h = 0$ denotes the final month of the target quarter, negative values denote months before the quarter ends, and $h = +1$ denotes the first month after the quarter ends. Each target quarter therefore contributes seven forecast tasks: $h = -5, \dots, -1$ are forecast horizons, $h = 0$ is the nowcast horizon, and $h = +1$ is a backcast horizon.

For each monthly vintage, the data set is truncated according to the publication calendar and the corresponding ragged edge, and each of the 16 specifications in Figure [1](#) is re-estimated on the truncated data. Each re-estimated specification then forecasts real GDP for the target quarters implied by the vintage date and the GDP publication lag.⁵ Forecasts

⁵At the ragged edge, unreleased observations are treated as missing and simulated within the state-space model, and the log-volatility paths are updated conditional on the resulting completed paths. Beyond the vintage endpoint, the common-factor and idiosyncratic log volatilities are propagated recursively from their terminal posterior values using their transition equations. When the outlier adjustment is active, the multiplicative scales associated with unreleased observations and future periods are drawn independently, conditional on each posterior draw of ν_i , from the distribution in equation [\(7\)](#) and multiply the idiosyncratic innovation variance. Without stochastic volatility, the innovation variances are held fixed at their posterior

Figure 1: **Model specification labels**

16 model specifications as a 4-dimensional hypercube
(edges connect specifications differing in exactly one feature)



Notes: Each node is one of the 16 model specifications. B denotes a specification without time-varying long-run growth, whereas T denotes a specification with time-varying long-run growth; the suffixes SV , DH , and O are appended when stochastic volatility, dynamic heterogeneity, and the outlier adjustment are active, respectively. The 16 specifications form a 4-dimensional hypercube, drawn here as two 3-dimensional cubes (left: constant long-run growth; right: time-varying long-run growth) over the remaining three features, with the outlier adjustment on the horizontal axis, stochastic volatility on the vertical axis, and dynamic heterogeneity on the diagonal axis within each cube. An edge connects two specifications whenever they differ in exactly one feature: solid gray edges toggle stochastic volatility, dynamic heterogeneity, or the outlier adjustment within a cube, and dashed orange edges toggle long-run growth between the two cubes.

are evaluated against revised GDP values from the latest available vintage used to construct the data set. Our evaluation sample covers target quarters from 2015Q1 to 2024Q4. We divide it into three periods: 2015Q1–2019Q4, 2020Q1–2021Q4, and 2022Q1–2024Q4, which we refer to as the pre-pandemic, pandemic, and post-pandemic periods, respectively. The classification is based on the target quarter. This partition allows us to investigate whether a feature that is useful on average has a different contribution depending on the macroeconomic environment.

5 Forecast Evaluation

5.1 Forecast construction and loss functions

For task τ and specification s , let $\{y_{\tau s}^{(r)}\}_{r=1}^{R_{\tau s}}$ denote posterior predictive draws of real GDP growth and let y_τ denote the realized value. The point forecast is the posterior predictive mean,

$$\hat{y}_{\tau s} = \frac{1}{R_{\tau s}} \sum_{r=1}^{R_{\tau s}} y_{\tau s}^{(r)}. \quad (8)$$

We score each specification with three loss functions. The squared forecast error loss measures point-forecast accuracy,

$$L_{\tau s}^{sfe} = (y_\tau - \hat{y}_{\tau s})^2, \quad (9)$$

and the CRPS measures density-forecast accuracy,

$$L_{\tau s}^{crps} = \frac{1}{R_{\tau s}} \sum_{r=1}^{R_{\tau s}} |y_{\tau s}^{(r)} - y_\tau| - \frac{1}{2R_{\tau s}^2} \sum_{r=1}^{R_{\tau s}} \sum_{\ell=1}^{R_{\tau s}} |y_{\tau s}^{(r)} - y_{\tau s}^{(\ell)}|. \quad (10)$$

Both $L_{\tau s}^{sfe}$ and $L_{\tau s}^{crps}$ are nonnegative by construction and strictly positive in our evaluation sample. Neither, however, reveals how a feature affects the extremes of the predictive distribution, so we also score the tails of the predictive distribution directly. Let $q_{\alpha, \tau s}$ denote the predictive α -quantile for task τ and specification s . The quantile score is

$$QS_{\alpha, \tau s} = (\alpha - \mathbf{1}\{y_\tau < q_{\alpha, \tau s}\}) (y_\tau - q_{\alpha, \tau s}), \quad (11)$$

which we evaluate at $\alpha \in \{0.05, 0.10, 0.90, 0.95\}$; $QS_{\alpha, \tau s} \geq 0$ by construction, and lower values indicate better tail-quantile forecasts.

draws; without the outlier adjustment, $o_{i,t} = 1$.

5.2 Balanced paired factorial contrasts

We isolate the marginal contribution of each feature with a single construction applied in turn to each of the three losses above, namely $L_{\tau s}^{sfe}$, $L_{\tau s}^{crps}$, and $QS_{\alpha, \tau s}$. For the squared forecast error and CRPS losses, we work with the log loss; for the quantile score, we work with the score directly:

$$z_{\tau s} = \begin{cases} \log L_{\tau s}, & L \in \{sfe, crps\}, \\ QS_{\alpha, \tau s}, & \text{quantile score, evaluated separately at each } \alpha. \end{cases} \quad (12)$$

Let $\mathcal{F}$ denote the set of four features described in Section 3 and define $x_{js} = 1$ if feature $j \in \mathcal{F}$ is included in specification s , with $x_{js} = 0$ otherwise. Because a specification is completely characterized by the four feature indicators, we write $z_{\tau}(x_j, x_{-j})$ for its value at task τ , where x_{-j} collects the indicators for the other three features.

For each task and feature, we compare the specifications that differ only in feature j . We average this comparison over all eight configurations of the other three features rather than fixing them at a single benchmark combination, because the estimated effect of feature j would otherwise depend on which companion features happen to be active; averaging over the full 2^3 set of configurations removes that dependence. We define the balanced paired factorial contrast as

$$d_{\tau j} = \frac{1}{8} \sum_{x_{-j} \in \{0,1\}^3} [z_{\tau}(x_j = 1, x_{-j}) - z_{\tau}(x_j = 0, x_{-j})], \quad (13)$$

where there are eight paired comparisons because the remaining three features generate 2^3 possible configurations. A negative value of $d_{\tau j}$ indicates that, for a given target quarter and forecast horizon, including feature j reduces z on average across those configurations: it lowers average log loss when $z = \log L_{\tau s}^{sfe}$ or $z = \log L_{\tau s}^{crps}$, and it lowers the average quantile score when $z = QS_{\alpha, \tau s}$. In this last case we write the contrast as $d_{\tau j \alpha}^{QS}$ to make the dependence on the quantile level explicit.

We obtain the overall contribution of feature j by averaging the task-level contrasts:

$$\bar{d}_j = \frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} d_{\tau j}, \quad (14)$$

where $\mathcal{T}$ denotes the set of all target-quarter-by-horizon tasks. Similarly, for period p , we calculate

$$\bar{d}_{jp} = \frac{1}{|\mathcal{T}_p|} \sum_{\tau \in \mathcal{T}_p} d_{\tau j}, \quad (15)$$

where $\mathcal{T}_p$ contains the tasks associated with target quarters in period $p \in \{\text{pre-pandemic, pandemic, post-pandemic}\}$. Hence overall and period-specific effects are obtained the same way for all three losses.

We report these effects in percent. For the squared forecast error and CRPS losses, $\bar{d}_j$ and $\bar{d}_{jp}$ are average log-loss differentials, which we convert into a percent change in loss with

$$100 (\exp(\bar{d}_j) - 1) \quad (16)$$

(and analogously with $\bar{d}_{jp}$ in place of $\bar{d}_j$). This exponential transformation is not meaningful for the quantile score, since $z = QS_{\alpha, \tau s}$ is not itself a log loss. Instead, we normalize $\bar{d}_{j\alpha}^{QS}$ by the average quantile score of the matched specifications in which feature j is inactive, $\overline{QS}_{\alpha}^{(-j)}$ (or its period-specific counterpart), and report

$$100 \frac{\bar{d}_{j\alpha}^{QS}}{\overline{QS}_{\alpha}^{(-j)}}. \quad (17)$$

Both transformations preserve the sign of $\bar{d}_j$ and $\bar{d}_{jp}$, so a negative percent change continues to indicate an improvement from including feature j .

We assess uncertainty with a bootstrap at the target quarter level, applied in the same way to all three losses and computed both for the full sample and within each period. For the full-sample contrasts, we resample target quarters with replacement from the complete evaluation sample; for the period-specific results, we resample quarters within each period. In both cases, all seven forecast horizons corresponding to the same quarter are kept together when bootstrapping each feature contrast.⁶ We use 5,000 bootstrap replications and report 90 percent confidence intervals.

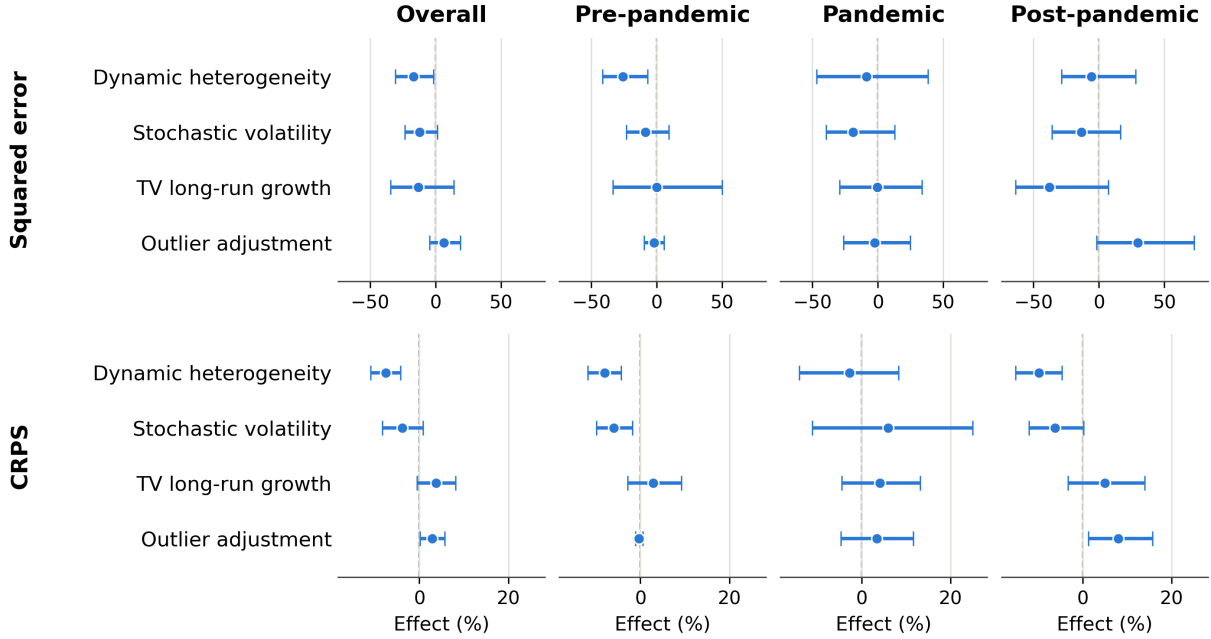
6 Results

Figure 2 presents the overall and regime-specific effects of the four model features in terms of percent differences in squared forecast error and CRPS losses, so negative numbers indicate an improvement in forecast performance when the corresponding feature is included.⁷

⁶This mirrors the broader concern that forecast performance across horizons for a common target quarter should not be treated as a collection of unrelated single-horizon comparisons (see [Quaedvlieg, 2021](#)). We do not implement his test, but our bootstrap design reflects the same dependence.

⁷Appendix B provides figures comparing the point nowcasts at different horizons with realized GDP growth across specifications and sample periods, as well as heatmaps of the root-mean-squared errors (RMSEs) and CRPS. It also compares the estimated latent factor, which we denominate “economic activity,” from each specification with the annualized growth rate of real private domestic final purchases (PDFP) and the GDPplus indicator of [Aruoba et al. \(2016\)](#). In addition, it reports estimates of long-run growth,

Figure 2: **Balanced paired factorial contrasts by loss function and period.**



Notes: Points are percent effects, $100(\exp(\bar{d}_j) - 1)$ (or $100(\exp(\bar{d}_{jp}) - 1)$ for regime panels); horizontal bars are 90 percent bootstrap percentile intervals from 5,000 target-quarter bootstrap replications. The vertical dashed line marks zero effect. TV denotes time-varying.

The most noticeable result in the Overall column of Figure 2 is the performance of dynamic heterogeneity. Its inclusion reduces squared-error loss by close to 17 percent and CRPS loss by about 8 percent. The 90 percent bootstrap intervals are below zero in both cases, in contrast with the other three features. Hence, among all the features we study, dynamic heterogeneity is the only one that likely produces an improvement for both point and density forecasts in the overall sample.⁸

This result is consistent with mechanisms emphasized by D’Agostino et al. (2016) and Antolín-Díaz et al. (2024). D’Agostino et al. (2016) show that, under dynamic homogeneity, persistent, low-volatility indicators can disproportionately shape the common factor, whereas allowing for heterogeneous lead–lag patterns produces a more timely indicator of the business cycle. Similarly, Antolín-Díaz et al. (2024) argue that, without heterogeneous loadings, the common factor is pulled toward survey indicators that move slowly and lags faster-moving indicators such as durable spending and investment, particularly around turning points in economic activity. Allowing each series its own lag structure corrects this mismatch.

stochastic volatilities, and outliers from the richest model.

⁸All the results in this section are reported in Appendix C in tabular form, which also allows us to report $P_{\text{boot}}(\Delta < 0)$, the proportion of bootstrap replications in which the corresponding feature lowers the loss (or, for the quantile score, lowers $QS_{\alpha, \tau s}$).

Although [D’Agostino et al. \(2016\)](#) do not find a statistically significant forecasting gain over the homogeneous model in their shorter evaluation sample, our factorial comparison identifies improvements in both squared-error loss and CRPS.

Moreover, comparing each specification’s full-sample factor to two external benchmarks of economic activity, such as growth in PDFP and GDPplus ([Aruoba et al., 2016](#)), shows that specifications with dynamic heterogeneity track both benchmarks far more closely than those without it: the correlation with PDFP growth rises on average from -0.10 to 0.50 , and the correlation with GDPplus rises from 0.17 to 0.69 , a pattern that holds regardless of whether the model also includes stochastic volatility or the shared trend.⁹

The point estimates for stochastic volatility imply reductions of 12.2 percent in squared forecast error loss and 3.8 percent in CRPS overall, but both 90 percent bootstrap confidence intervals narrowly include zero. Its overall effect on nowcasting performance is therefore less conclusive than that of dynamic heterogeneity. The favorable CRPS point estimate is consistent with evidence from the Bayesian VAR literature that stochastic volatility can improve density and interval forecasts (see [Clark, 2011](#); [Clark and Ravazzolo, 2015](#); [D’Agostino et al., 2013](#)). Regarding long-run growth, its point estimate implies a reduction in squared forecast error loss of about 13 percent, but its 90 percent confidence interval includes zero, while its effect on CRPS is likely detrimental. Finally, the multiplicative outlier adjustment raises both losses; the 90 percent interval includes zero for squared forecast error, but not for CRPS. Among these three remaining features, stochastic volatility is therefore the most consistently beneficial in the overall sample, time-varying long-run growth a distant second on the strength of its point-forecast effect alone, and the outlier adjustment the only one of the four features with no overall benefit under either loss function.¹⁰

The gains from dynamic heterogeneity are concentrated outside the pandemic, as the three remaining columns of [Figure 2](#) show for the pre-pandemic, pandemic, and post-pandemic periods separately. Before the pandemic, the feature improves squared-error performance, while CRPS improvements occur both before and after the pandemic. During 2020–21, the point estimates of the losses are close to zero and the intervals are wide enough that no clear effect can be established.

Stochastic volatility’s gains are likewise concentrated before the pandemic rather than during it. It improves CRPS before the pandemic, whereas the pandemic estimate leans to-

⁹These correlations are computed from the full-sample factor estimates shown in [Figure B.10](#) of [Appendix B.4](#), and are not part of the vintage-by-vintage forecast-evaluation exercise reported elsewhere in this section.

¹⁰We also investigate whether these findings depend on the bootstrap assumption that target quarters are sampled independently. [Appendix D](#) reports results from a four-quarter moving-block bootstrap that allows for local serial dependence in the loss differentials. The main conclusion is unchanged: dynamic heterogeneity remains the only feature with gains under both loss functions.

ward worsening the density forecast, although the estimate is imprecise. Thus, even though changes in volatility were especially pronounced in 2020, the results do not indicate that stochastic volatility produces a clear forecasting improvement during the pandemic period. This finding is consistent with related evidence that gains in density forecasts from stochastic volatility in Bayesian VARs and dynamic factor models are typically documented over tranquil or moderately volatile samples (see [Clark, 2011](#); [Clark and Ravazzolo, 2015](#); [Carriero et al., 2024](#)), rather than during episodes as extreme as the pandemic, when volatility itself shifts too abruptly for the assumed random-walk process to track it within the nowcasting horizon.

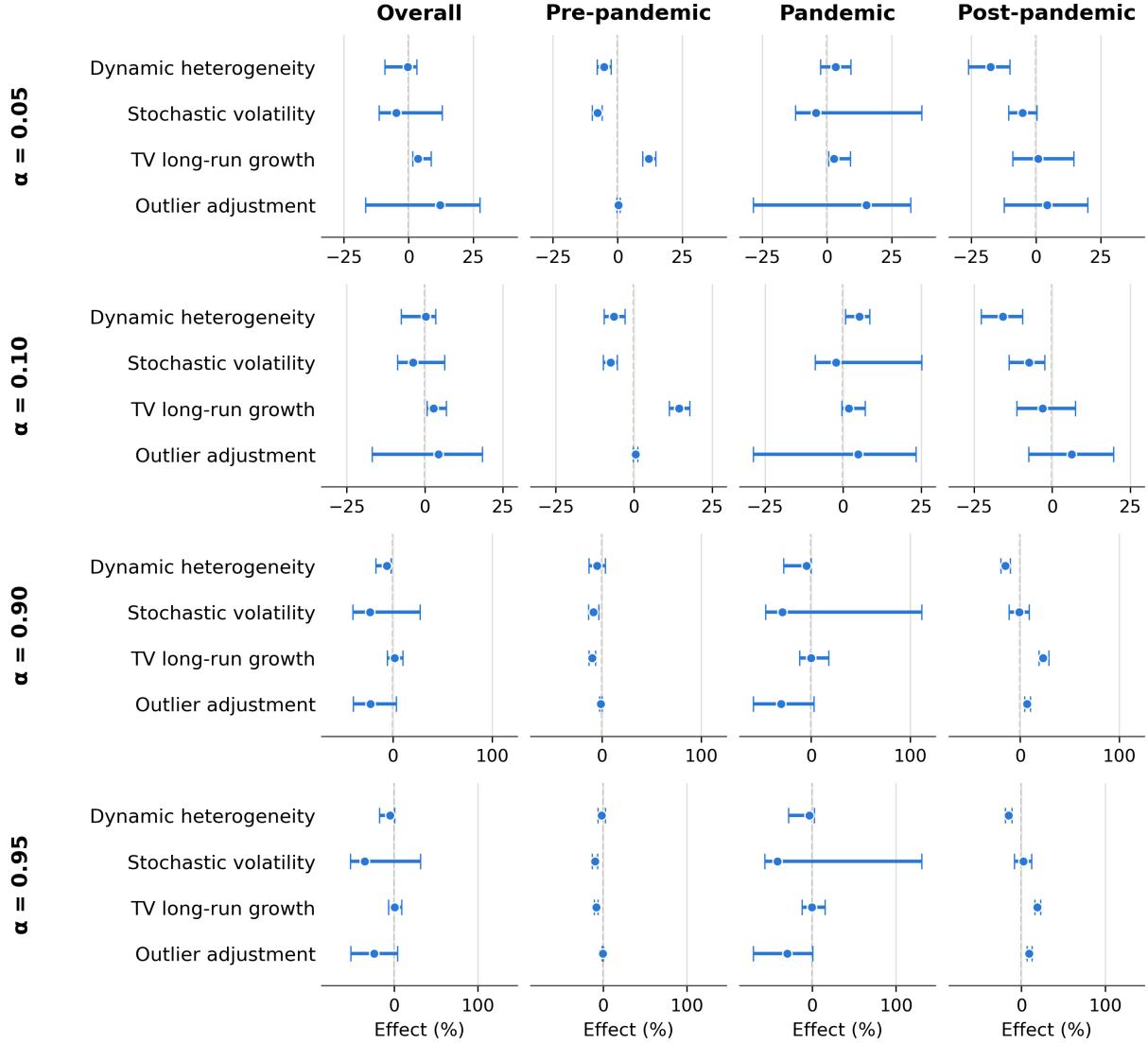
Including time-varying long-run growth does not provide a robust CRPS enhancement in any of the three periods. However, including this feature substantially lowers squared-error loss after the pandemic, with uncertainty around that estimate remaining relatively large. A slowly evolving trend can absorb the persistent shift in average growth that materialized after the pandemic and sharpen the point nowcast, but the added parameter uncertainty from estimating that trend widens the predictive distribution enough to offset the gain in density terms. Lastly, the multiplicative outlier adjustment does not show a clear benefit during the pandemic, the episode whose unusually large observations it was designed in part to accommodate. Its effects are imprecisely estimated before and during 2020–21 under both loss functions. After the pandemic, however, the adjustment raises CRPS (a worsening of density forecasting ability), with its 90 percent interval entirely above zero, and the squared-error point estimate also indicates an adverse effect, although its 90 percent interval narrowly includes zero. This post-pandemic cost likely reflects the same scale-mixing parameters that let the adjustment absorb the extreme observations of 2020 continuing to inflate the predictive variance around the more ordinary fluctuations that followed, at a point when that insurance is no longer needed.

These patterns, visible across the panels of [Figure 2](#), underscore that practitioners building nowcasting models should prioritize dynamic heterogeneity over the other three features when only one addition can be afforded. The outlier adjustment does, however, contain the size of individual forecast misses during the pandemic: as [Appendix B](#) shows, several specifications without the adjustment produce nowcasts that substantially miss realized GDP growth at the height of the pandemic, whereas specifications with the outlier adjustment do not.

Dynamic heterogeneity remains the strongest feature under the quantile-score contrasts defined in [Section 5.2](#): in the overall results reported in [Figure 3](#), it is the only feature with a 90 percent confidence interval entirely below zero, corresponding to the 90th percentile. Stochastic volatility and the outlier adjustment show large improvements at the 90th and

95th percentiles, but their intervals are wide and include zero. Time-varying long-run growth raises the 5th- and 10th-percentile scores, with intervals excluding zero, indicating worse lower-tail performance on average.

Figure 3: **Balanced paired factorial contrasts for quantile scores, by quantile level and period.**



Notes: Points are percent changes in mean quantile score relative to the matched feature-off specifications; horizontal bars are 90 percent bootstrap percentile intervals from 5,000 target-quarter bootstrap replications. The vertical dashed line marks zero effect. TV denotes time-varying.

The remaining columns of Figure 3 show the same contrasts by period. As with squared-error and CRPS losses, the clearest tail-forecasting gains tend to appear before or after the pandemic, although some favorable estimates also emerge during it. Starting with dynamic heterogeneity, it improves lower-tail scores before and after the pandemic and produces its

clearest upper-tail gains after it. Stochastic volatility also generates several gains at both the lower and upper tails, most of them before the pandemic. The outlier adjustment improves the 90th- and 95th-percentile scores during the pandemic, which is precisely when its effect should be most noticeable, although the estimates are imprecise; this adjustment turns detrimental for these same percentile scores after the pandemic. Finally, time-varying long-run growth shows no clear pattern: it improves upper-tail performance somewhat before the pandemic but worsens the lower tail over that same period, and after the pandemic it moves the other way, hurting the upper tail. Appendix C.2 reports these regime-specific effects in tabular form, split into lower- and upper-tail quantiles, in addition to the bootstrap probabilities that a given feature lowers the quantile score loss.

7 Conclusion

In this paper, we evaluate the contribution of four features commonly used to enrich Bayesian dynamic factor models for nowcasting: dynamic heterogeneity, stochastic volatility, time-varying long-run growth, and a multiplicative adjustment for unusual observations. Rather than comparing each feature with a single benchmark specification, we estimate all 16 combinations of these features and use balanced paired factorial contrasts to obtain their marginal contributions to forecast performance.

Our results indicate that dynamic heterogeneity is the most useful feature in the exercise. It improves average performance under both squared-error loss and CRPS, with reductions of 16.8 percent and 7.6 percent, respectively, and it is the only feature with a precisely estimated improvement in an overall quantile score, cutting the 90th-percentile score by 6.2 percent. These gains are more evident outside the pandemic, particularly before the pandemic period for squared-error loss and before and after the pandemic for CRPS. Moreover, this feature also improves lower- and upper-tail forecasting, especially outside the pandemic period.

The evidence for the remaining features is less conclusive. Stochastic volatility tends to reduce both losses overall and lowers CRPS and both tail quantile scores before the pandemic, but its effects are estimated with greater uncertainty and it does not produce a clear gain during the pandemic. Time-varying long-run growth does not materially improve density forecasts and worsens the lower-tail quantile scores before the pandemic; it produces a sizeable reduction in squared-error loss after the pandemic, but that same period is when it does the most damage to the upper-tail quantile scores, raising them by close to 20 percent. The multiplicative outlier adjustment does not improve average point or density forecast performance and is particularly detrimental after the pandemic, despite being a feature motivated in part by the unusually large observations associated with that episode. During

the pandemic itself, its point estimate for the upper-tail quantile scores drops by about 30 percent, though this is not precisely estimated.

All told, our results suggest that greater model flexibility does not necessarily translate into better nowcasts. The usefulness of a particular feature depends on the forecasting objective and on the period being evaluated. In our exercise, allowing for heterogeneous factor dynamics is the only one of the four bells and whistles that produces statistically supported improvements in point forecasts, density forecasts, and at least one quantile-score criterion over the full sample.

Data Availability

Data will be made available on request.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work the authors used large-language-model assistants, specifically Claude (Anthropic) and Codex (OpenAI), in order to streamline code, process results, and polish writeup. After using these tools, the authors reviewed and verified all outputs, edited the content as needed, and take full responsibility for the content of the published article.

CRedit author statement

Freddy García-Albán: Methodology, Software, Validation, Formal analysis, Data Curation, Visualization, Writing - Original Draft.

Manuel González-Astudillo: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - Review & Editing.

A Data, model, and estimation

This section provides additional detail on the data, the full model formulation, and the estimation procedure summarized in Sections 2 and 3.

A.1 Data

Table 1 reports the variables, release frequencies, and transformations used in the now-casting exercise. Figure A.1 plots each series in the units in which it enters the model – i.e. after the transformation reported in Table 1 (annualized quarter-over-quarter percent change for the three quarterly variables; month-over-month percent change, level change, or survey index for the monthly indicators) but before the standardization applied prior to estimation. All panels span the estimation sample through December 2024, while the plotted observations for each series begin at that series’ own availability date. These dates vary because the underlying source series differ in length (e.g. ADP nonfarm private payroll and JOLTS total job openings are available only from 2010 and 2001, respectively).

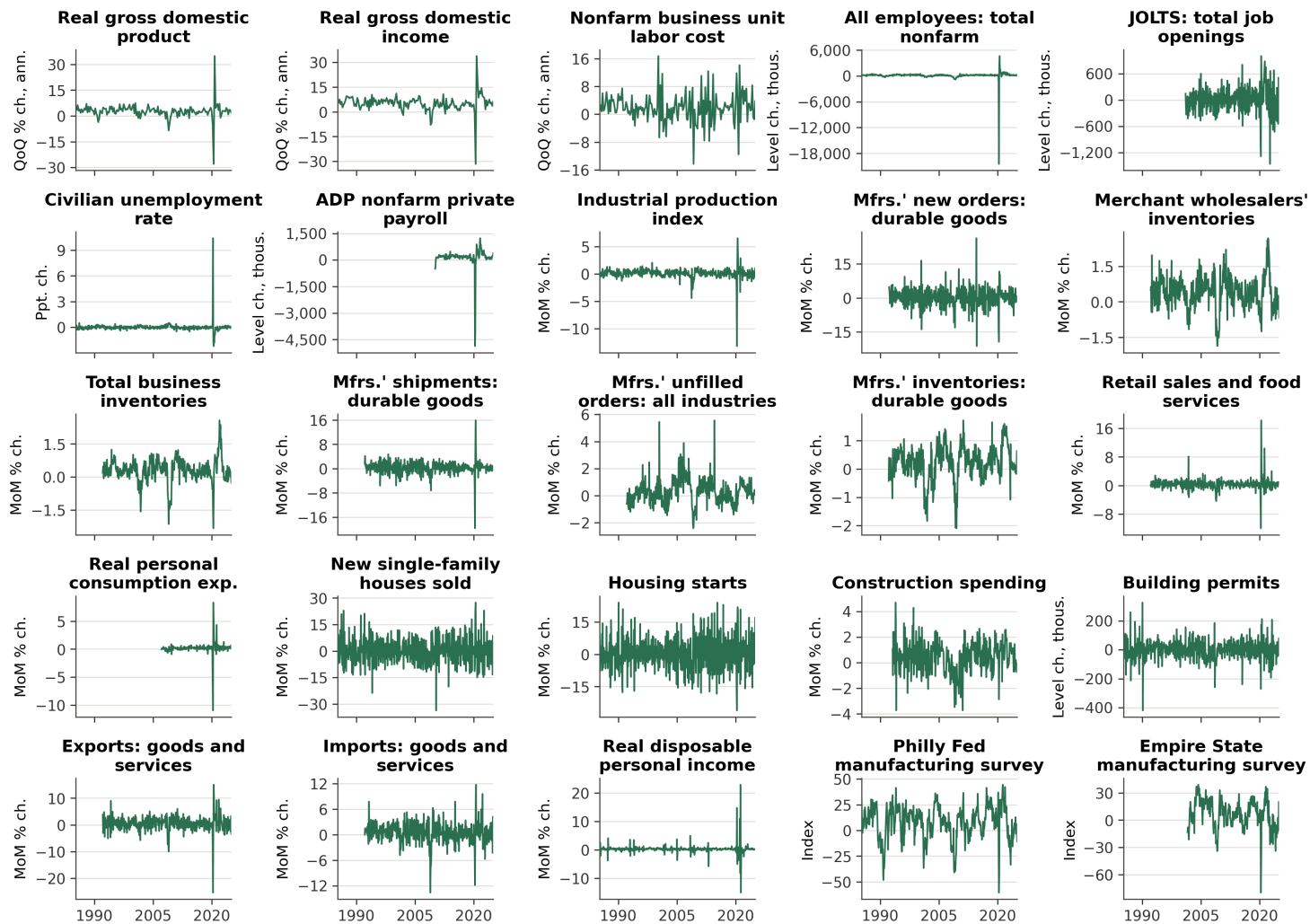


Figure A.1: Each of the 25 series used in the nowcasting exercise, in the units in which it enters the model (Table 1), before standardization, over the estimation sample through December 2024.

A.2 The model

Let $\mathbf{y}_t$ collect the $n = n_m + n_q$ monthly variables and latent monthly observations associated with quarterly variables, where n_m and n_q denote the numbers of monthly and quarterly variables. The model is written on a monthly grid as

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{\Lambda}(L)f_t + \mathbf{u}_t, \quad t = 1, \dots, T, \quad (\text{A.1})$$

where $\mathbf{c}_t$ contains the long-run growth terms, f_t is the single common factor, and $\mathbf{u}_t$ contains idiosyncratic components.

The loading polynomial is

$$\mathbf{\Lambda}(L) = \mathbf{\Lambda}_0 + \mathbf{\Lambda}_1 L + \dots + \mathbf{\Lambda}_m L^m \quad (\text{A.2})$$

The order m controls dynamic heterogeneity in the measurement equation. In principle, m can be any nonnegative integer. In the factorial design below we restrict attention to $m \in \{0, 1\}$. The case $m = 0$ is the model without dynamic heterogeneity: each observable loads only on the contemporaneous common factor. The case $m = 1$ lets observables load on the contemporaneous factor and one monthly lag. This is the dynamic-heterogeneity specification used by [Antolín-Díaz et al. \(2024\)](#), who report that $m = 1$ is sufficient to allow for rich lead-lag heterogeneity, while their deviance information criterion (DIC) comparison supports $m > 0$ but not higher lag orders. The common factor is normalized by setting the contemporaneous loading in the GDP equation equal to one.

Let

$$\phi(L) = \phi_1 L + \dots + \phi_p L^p, \quad \rho_i(L) = \rho_{i,1} L + \dots + \rho_{i,q} L^q.$$

In all specifications, we set $p = q = 2$.

The factor and idiosyncratic components are

$$(1 - \phi(L))f_t = \sigma_{\varepsilon,t}\varepsilon_t, \quad \varepsilon_t \stackrel{\text{iid}}{\sim} N(0, 1), \quad (\text{A.3})$$

$$(1 - \rho_i(L))u_{i,t} = \sigma_{\eta_i,t}\sqrt{O_{i,t}}\eta_{i,t}, \quad \eta_{i,t} \stackrel{\text{iid}}{\sim} N(0, 1), \quad i = 1, \dots, n. \quad (\text{A.4})$$

In stochastic-volatility specifications, define $h_{\varepsilon,t} \equiv \log \sigma_{\varepsilon,t}^2$ and $h_{\eta_i,t} \equiv \log \sigma_{\eta_i,t}^2$. Then

$$h_{\varepsilon,t} = h_{\varepsilon,t-1} + v_{\varepsilon,t}, \quad v_{\varepsilon,t} \stackrel{\text{iid}}{\sim} N(0, \omega_{v_{\varepsilon}}^2), \quad h_{\varepsilon,0} \sim N(\mu_{h_{\varepsilon,0}}, \sigma_{h_{\varepsilon,0}}^2), \quad (\text{A.5})$$

$$h_{\eta_i,t} = h_{\eta_i,t-1} + v_{\eta_i,t}, \quad v_{\eta_i,t} \stackrel{\text{iid}}{\sim} N(0, \omega_{v_{\eta_i}}^2), \quad h_{\eta_i,0} \sim N(\mu_{h_{\eta_i,0}}, \sigma_{h_{\eta_i,0}}^2). \quad (\text{A.6})$$

In constant-volatility specifications, $\sigma_{\varepsilon,t}$ and $\sigma_{\eta_i,t}$ are time invariant.

The time-varying long-run growth specification imposes a common low-frequency component on the expenditure and income measures of aggregate output:

$$\mathbf{c}_t = \mathbf{c} + \mathbf{b}g_t, \quad g_t = g_{t-1} + \zeta_t, \quad \zeta_t \stackrel{\text{iid}}{\sim} N(0, \omega_g^2). \quad (\text{A.7})$$

Let $\mathcal{G} = \{\text{GDP}, \text{GDI}\}$. The selection vector is

$$b_i = \begin{cases} 1, & i \in \mathcal{G}, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.8})$$

Without time-varying long-run growth, $\mathbf{c}_t = \mathbf{c}$. In the standardized estimation system used in the empirical implementation, $\mathbf{c} = \mathbf{0}$. The transformed series are standardized before estimation, so no intercept vector is separately estimated.

When the outlier adjustment is inactive, $o_{i,t} = 1$. When it is active, following [Hong \(2025\)](#), the standardized idiosyncratic innovation $\sqrt{o_{i,t}}\eta_{i,t}$ follows a Student- t distribution with ν_i degrees of freedom. For estimation, we use the equivalent normal-inverse-gamma scale-mixture representation:

$$o_{i,t} \mid \nu_i \sim \mathcal{IG}\left(\frac{\nu_i}{2}, \frac{\nu_i}{2}\right), \quad i = 1, \dots, n. \quad (\text{A.9})$$

Equivalently, $\nu_i/o_{i,t} \sim \chi_{\nu_i}^2$.

Following [Mariano and Murasawa \(2003\)](#), for a quarterly growth rate x_t^q and its latent monthly growth rate x_t^m ,

$$x_t^q = \frac{1}{3}x_t^m + \frac{2}{3}x_{t-1}^m + x_{t-2}^m + \frac{2}{3}x_{t-3}^m + \frac{1}{3}x_{t-4}^m. \quad (\text{A.10})$$

Following [Hong \(2025\)](#), the sampler imposes (A.10) on the unobserved quarterly idiosyncratic components rather than on the quarterly variables directly.

A.3 Model specifications

We evaluate four common features of dynamic factor models: stochastic volatility, dynamic heterogeneity in $\mathbf{\Lambda}(L)$, time-varying long-run growth, and multiplicative outlier adjustment. Each feature can be switched on or off, yielding $2^4 = 16$ possible specifications. For ease of notation, the figures and tables below use the compact specification labels defined in Figure 1: B denotes a specification without time-varying long-run growth, whereas T denotes a specification with time-varying long-run growth. The suffixes SV , DH , and O are included when stochastic volatility, dynamic heterogeneity, and outlier adjustment are

active, respectively.

A.4 The state space representation

The sampler uses two state-space systems. The first is a univariate series-by-series system that recovers monthly idiosyncratic components for quarterly variables. The second is the multivariate dynamic-factor system for the panel that contains observed monthly variables and monthly-interpolated quarterly variables.

Monthly interpolation of quarterly idiosyncratic components. For each quarterly variable $i \in \{1, \dots, n_q\}$, define

$$x_{i,t} = [u_{i,t}^M \ u_{i,t-1}^M \ u_{i,t-2}^M \ u_{i,t-3}^M \ u_{i,t-4}^M]'. \quad (\text{A.11})$$

The monthly interpolation of the quarterly idiosyncratic component is obtained from

$$u_{i,t}^Q = Hx_{i,t} + \xi_{i,t}, \quad \xi_{i,t} \sim N(0, R), \quad (\text{A.12})$$

$$x_{i,t} = A_i x_{i,t-1} + e_{i,t}, \quad e_{i,t} \sim N(0, Q_{i,t}), \quad (\text{A.13})$$

with

$$H = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}, \quad A_i = \begin{bmatrix} \rho_{i,1} & \cdots & \rho_{i,q} & 0_{1 \times (5-q)} \\ I_4 & 0_{4 \times 1} \end{bmatrix}, \quad (\text{A.14})$$

$$Q_{i,t} = \text{diag}(\sigma_{\eta_{i,t}}^2 o_{i,t}, 0, 0, 0, 0),$$

and $R = 0$ in the implementation, corresponding to exact aggregation. Given $u_{i,t}^M$, the monthly-interpolated value of a quarterly series is reconstructed as

$$y_{i,t}^{Q,M} = c_{i,t} + \sum_{\ell=0}^m \lambda_{i,\ell} f_{t-\ell} + u_{i,t}^M, \quad i \leq n_q, \quad (\text{A.15})$$

with absent terms omitted when the corresponding feature is not included.

Dynamic-factor state space. Let

$$\tilde{y}_{i,t} = y_{i,t} - \rho_{i,1} y_{i,t-1} - \cdots - \rho_{i,q} y_{i,t-q} \quad (\text{A.16})$$

denote the quasi-differenced series. Ordering quarterly variables first,

$$\tilde{\mathbf{y}}_t = \begin{bmatrix} \tilde{y}_{1,t}^{Q,M} \\ \vdots \\ \tilde{y}_{n_q,t}^{Q,M} \\ \tilde{y}_{n_q+1,t}^M \\ \vdots \\ \tilde{y}_{n,t}^M \end{bmatrix}. \quad (\text{A.17})$$

Let $d = \max\{p, q\}$. The state-space system for the quasi-differenced data is

$$\tilde{\mathbf{y}}_t = \mathbf{C}_t \boldsymbol{\xi}_t + \mathbf{D}_t \tilde{\mathbf{u}}_t, \quad (\text{A.18})$$

$$\boldsymbol{\xi}_t = \mathbf{A}_t \boldsymbol{\xi}_{t-1} + \mathbf{B}_t \tilde{\mathbf{e}}_t. \quad (\text{A.19})$$

Here, $\tilde{\mathbf{u}}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{I}_n)$ and $\tilde{\mathbf{e}}_t \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{I}_{s_\xi})$ are mutually independent, where $s_\xi = d + m + 1$ without time-varying long-run growth and $s_\xi = d + m + q + 2$ with it. For the time-varying long-run growth specification,

$$\boldsymbol{\xi}_t = [g_t, \dots, g_{t-q}, f_t, \dots, f_{t-d-m}]'. \quad (\text{A.20})$$

Without time-varying long-run growth, the g_t block is dropped. When $m = 0$,

$$\mathbf{C}_t = \begin{bmatrix} \mathbf{C}^g & \mathbf{C}^\lambda \end{bmatrix}, \quad (\text{A.21})$$

where

$$\mathbf{C}^g = \text{diag}(\mathbf{b}) \begin{bmatrix} 1 & -\rho_{1,1} & \cdots & -\rho_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & -\rho_{n,1} & \cdots & -\rho_{n,q} \end{bmatrix}, \quad (\text{A.22})$$

and

$$\mathbf{C}^\lambda = \text{diag}(\boldsymbol{\Lambda}_0) \begin{bmatrix} 1 & -\rho_{1,1} & \cdots & -\rho_{1,q} \\ 1 & -\rho_{2,1} & \cdots & -\rho_{2,q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & -\rho_{n,1} & \cdots & -\rho_{n,q} \end{bmatrix}. \quad (\text{A.23})$$

When $m = 1$, $\mathbf{C}^\lambda$ is expanded to include the corresponding quasi-differenced loading on f_{t-1} .

The transition matrix is

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_2 \end{bmatrix}, \quad \mathbf{A}_1 = \begin{bmatrix} 1 & 0_{1 \times q} \\ I_q & 0_{q \times 1} \end{bmatrix}, \quad (\text{A.24})$$

and

$$\mathbf{A}_2 = \begin{bmatrix} \phi_1 & \cdots & \phi_p & 0_{1 \times (d+1+m-p)} \\ I_{d+m} & 0_{(d+m) \times 1} & & \end{bmatrix}. \quad (\text{A.25})$$

The shock-loading matrices are

$$\mathbf{D}_t = \text{diag}(\sigma_{\eta_1,t}\sqrt{o_{1,t}}, \dots, \sigma_{\eta_n,t}\sqrt{o_{n,t}}), \quad (\text{A.26})$$

$$\mathbf{B}_t = \text{diag}(\omega_g, 0_{1 \times q}, \sigma_{\varepsilon,t}, 0_{1 \times (d+m)}). \quad (\text{A.27})$$

A.5 Priors

For the loading blocks,

$$\lambda_{i,0} \sim N(\bar{\lambda}_{i,0}, V_\lambda), \quad \bar{\lambda}_{i,0} = 1, \quad V_\lambda = 0.2, \quad (\text{A.28})$$

$$\lambda_{i,\ell} \sim N(0, V_\lambda), \quad \ell = 1, \dots, m. \quad (\text{A.29})$$

The GDP contemporaneous factor loading is fixed at one. In the time-varying long-run growth specification, the trend loadings of GDP and GDI are fixed at one; the GDI factor loading is estimated.

For autoregressive parameters,

$$\phi_k \sim N(\bar{\phi}_k, V_{\phi,k}), \quad \bar{\phi}_1 = 0.9, \quad \bar{\phi}_k = 0 \ (k > 1), \quad V_{\phi,k} = \frac{0.2}{k^2}, \quad (\text{A.30})$$

$$\rho_{i,k} \sim N(0, V_{\rho,k}), \quad V_{\rho,k} = \frac{0.2}{k^2}. \quad (\text{A.31})$$

Draws implying nonstationary $\phi(L)$ or $\rho_i(L)$ are rejected.

In constant-volatility specifications,

$$\sigma_\varepsilon^2 \sim \mathcal{IG}(3, 1), \quad \sigma_{\eta_i}^2 \sim \mathcal{IG}(3, 1). \quad (\text{A.32})$$

In stochastic-volatility specifications, the innovation variances of the log-volatility random walks have priors

$$\omega_{v_\varepsilon}^2 \sim \mathcal{IG}(1, 10^{-4}), \quad \omega_{v_{\eta_i}}^2 \sim \mathcal{IG}(1, 10^{-4}). \quad (\text{A.33})$$

The initial log volatilities satisfy

$$h_{\varepsilon,0} \sim N(0, 1), \quad h_{\eta_i,0} \sim N(0, 1). \quad (\text{A.34})$$

For the long-run growth innovation variance,

$$\omega_g^2 \sim \mathcal{IG}(1, 10^{-3}). \quad (\text{A.35})$$

The priors for the loadings, autoregressive coefficients, stochastic volatility, and long-run growth follow the shrinkage logic of [Antolín-Díaz et al. \(2024\)](#). The contemporaneous factor loadings are centered on the scale of the standardized variables, lagged loadings and idiosyncratic autoregressive coefficients are centered at zero, and prior variances decline with lag length. The priors on the innovation variances of the stochastic-volatility and long-run-growth processes shrink toward constant-volatility and constant-long-run-growth specifications.

For the multiplicative outlier adjustment, the scale distribution is given by [\(A.9\)](#). For each monthly indicator $i > n_q$, the degrees of freedom are series specific and have the discrete prior

$$\mathbb{P}(\nu_i = k) \propto k \exp(-k/10), \quad k \in \{3, \dots, 40\}. \quad (\text{A.36})$$

For each quarterly variable $i \leq n_q$, ν_i is fixed at 30. Thus all series receive multiplicative scale draws, while the quarterly innovations are restricted to have substantially thinner tails.

A.6 The Gibbs sampler algorithm

The sampler has a block structure. Superscript j denotes the current iteration. Blocks associated with stochastic volatility, time-varying long-run growth, dynamic heterogeneity, or outlier adjustment are omitted when the corresponding feature is inactive.

0. *Initialization.* Initialize the model parameters, stochastic volatility paths, multiplicative scales, degrees of freedom, and a monthly interpolation of the quarterly variables. Set $j = 1$.
1. *Quasi-difference the current monthly panel.* Given $\boldsymbol{\rho}^{j-1}$ and the current monthly-interpolated quarterly variables, construct

$$\tilde{\mathbf{y}}_t^j = \mathcal{Q}(\mathbf{y}_t^{M,j-1}; \boldsymbol{\rho}^{j-1}),$$

where $\mathcal{Q}(\cdot; \boldsymbol{\rho})$ denotes the series-by-series quasi-difference operator. The idiosyncratic

innovation covariance used in the simulation smoother has diagonal elements $\sigma_{\eta_i,t}^{2,j-1} \sigma_{i,t}^{j-1}$.

2. *Draw the common states.* Draw the state sequence $\{\xi_t^j\}_{t=1}^T$, and hence $\{f_t^j\}_{t=1}^T$ and, when present, $\{g_t^j\}_{t=1}^T$, conditional on the current parameters, volatility paths, and multiplicative scales, using the simulation smoother applied to (A.18)–(A.19).
3. *Draw the loading blocks.* Conditional on $\{\xi_t^j\}_{t=1}^T$, ρ^{j-1} , the idiosyncratic volatility paths, and the multiplicative scales, the measurement equations reduce to independent Gaussian regressions in quasi-differences. Draw $\Lambda_0^j, \dots, \Lambda_m^j$ equation by equation. Impose $\lambda_{\text{GDP},0} = 1$; in the time-varying long-run growth specification, impose unit trend loadings for GDP and GDI and estimate the GDI factor loading.
4. *Draw idiosyncratic autoregressive coefficients and volatilities.* For each series, draw ρ_i^j from the idiosyncratic autoregression conditional on the current common states and loadings. Reject nonstationary draws. Conditional on ρ_i^j , standardize the idiosyncratic innovations by $\sqrt{\sigma_{i,t}^{j-1}}$ and update the idiosyncratic volatility path: use the random-walk log-volatility block when stochastic volatility is active and an inverse-gamma update otherwise.
5. *Draw the long-run growth innovation variance.* If the time-varying long-run growth block is active, draw $\omega_g^{2,j}$ conditional on the current path $\{g_t^j\}_{t=1}^T$. Given the inverse-gamma prior in Section A.5 and the random-walk law of motion for g_t , the conditional posterior is inverse-gamma.
6. *Draw factor autoregressive coefficients and volatility.* Draw ϕ^j from the factor autoregression conditional on $\{f_t^j\}_{t=1}^T$ and the factor volatility path. Reject nonstationary draws. Conditional on ϕ^j , update the factor volatility path using the stochastic-volatility block when active and an inverse-gamma update otherwise.
7. *Draw outliers and degrees of freedom.* If the outlier adjustment is active, draw $\{\sigma_{i,t}^j\}$ conditional on the standardized idiosyncratic innovations. Then update each monthly ν_i^j on the grid $\{3, \dots, 40\}$. Quarterly degrees of freedom remain fixed at 30. If the block is inactive, set $\sigma_{i,t}^j = 1$.
8. *Update monthly-interpolated quarterly variables.* For each quarterly variable, compute the quarterly idiosyncratic component net of common factors. Draw $\{u_{i,t}^{M,j}\}_{t=1}^T$ from (A.12)–(A.13) using innovation variance $\sigma_{\eta_i,t}^2 \sigma_{i,t}^j$, and reconstruct the monthly-interpolated quarterly series $y_{i,t}^{Q,M,j}$.
9. *Iterate.* Set $j = j + 1$ and repeat until convergence.

For each specification and monthly information set, we discard the first 10,000 Gibbs-sampler iterations as burn-in and retain every 50th subsequent draw until 1,000 posterior draws are obtained.

B Additional nowcasting results

B.1 GDP forecasts

Figures B.1–B.3 plot realized real GDP growth and horizon-specific nowcasts by target quarter for the 16 specifications. The evaluation windows are 2015Q1–2019Q4, 2020Q1–2021Q4, and 2022Q1–2024Q4. Realized GDP growth is shown with one year of pre-window history. Panel labels encode the specification: *B* excludes time-varying long-run growth, *T* includes it, *SV* denotes stochastic volatility, *DH* denotes dynamic heterogeneity, and *O* denotes outlier adjustment. Each panel uses a fixed vertical range chosen for legibility of the bulk of the forecasts within that period; a nowcast that falls outside this range is instead plotted as a triangle at the panel’s upper or lower edge, annotated with its numerical value. In the pandemic figure (Figure B.2) these markers appear only in the panels without the outlier adjustment (*B*, *B-DH*, *B-SV*, *B-SV-DH*, *T*, *T-DH*, *T-SV*, *T-SV-DH*). Several of these nowcasts miss realized GDP growth by more than 100 percentage points; the annotated forecast values include approximately -193 and -228 percent. No such markers appear in the panels with the outlier adjustment. This is consistent with the discussion in Section 6 that the outlier adjustment, while not improving average point or density forecast accuracy, contains the size of the most extreme individual pandemic-period misses.

GDP forecasts and actuals by specification: Pre-pandemic

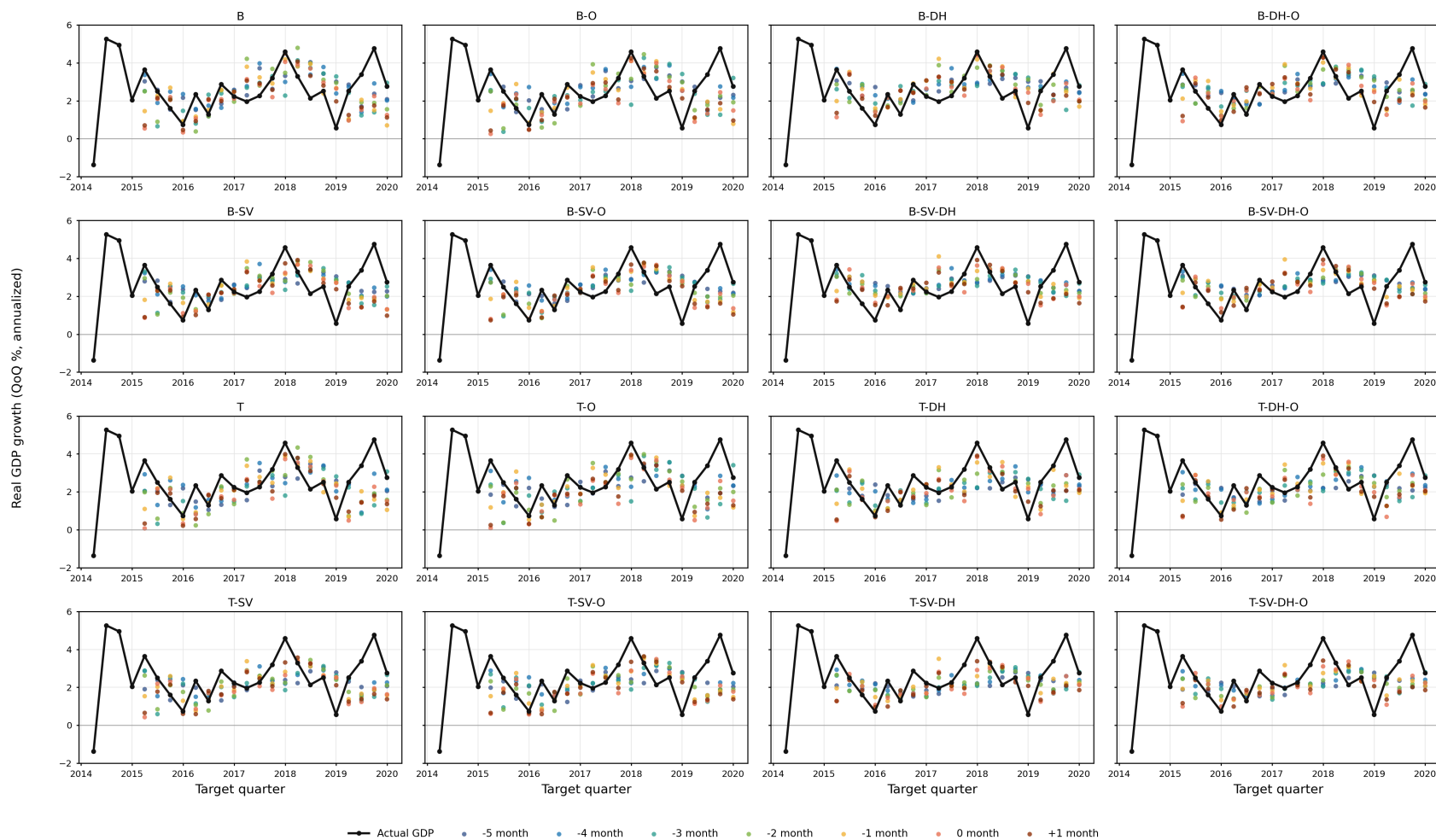


Figure B.1: Real GDP growth forecasts and realizations by specification: pre-pandemic period. Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

GDP forecasts and actuals by specification: Pandemic

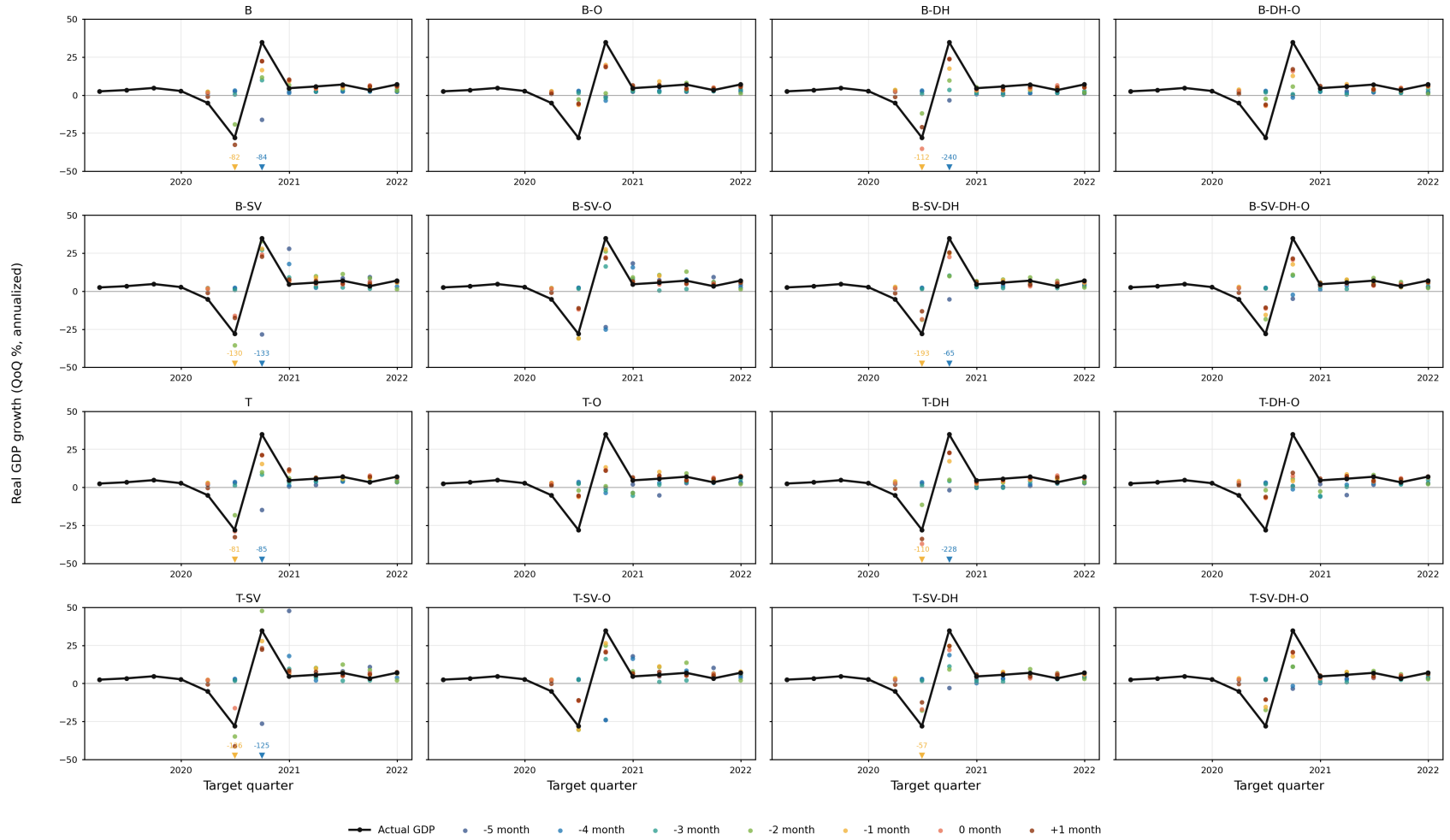


Figure B.2: Real GDP growth forecasts and realizations by specification: pandemic period. Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

GDP forecasts and actuals by specification: Post-pandemic

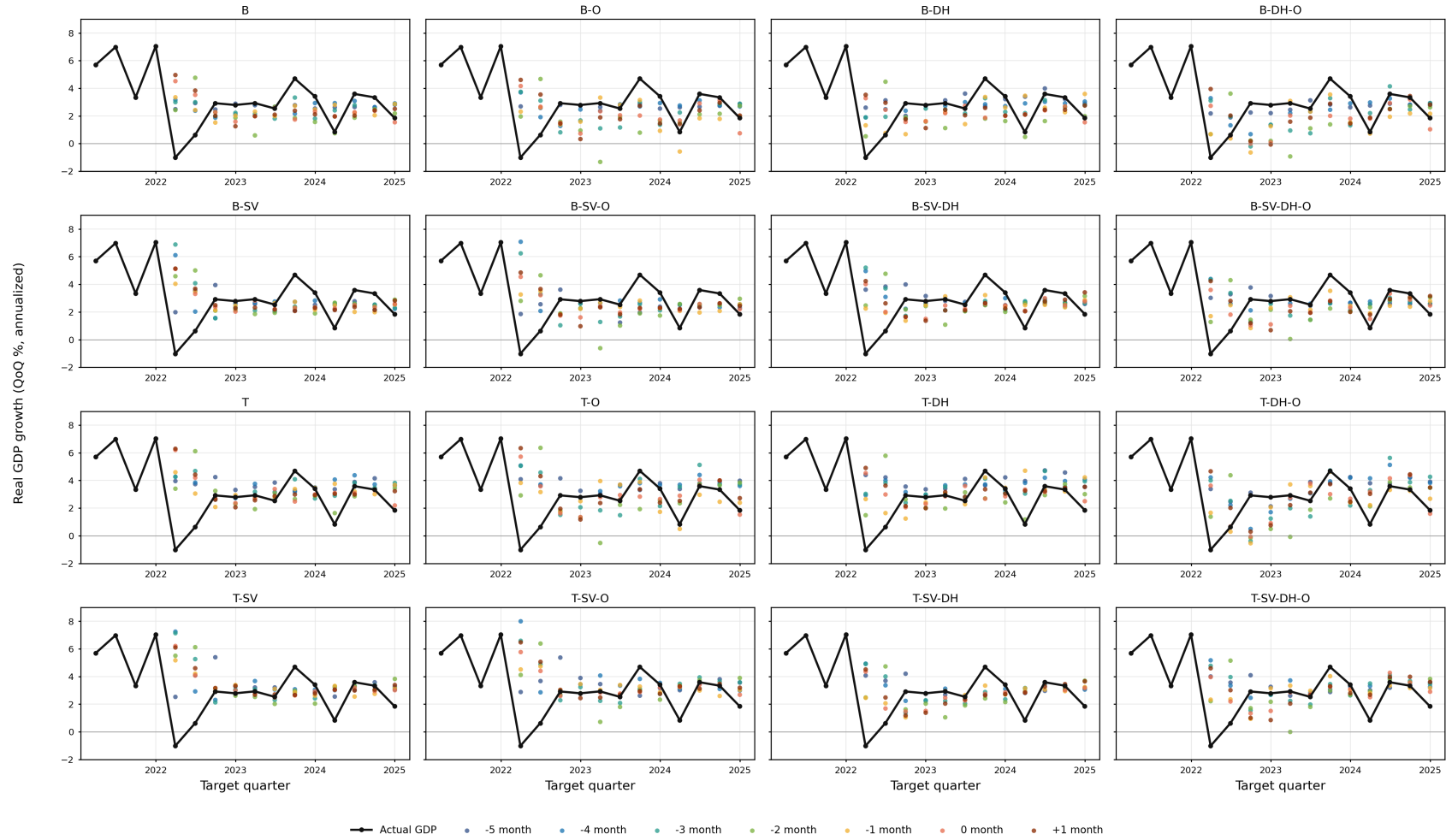


Figure B.3: Real GDP growth forecasts and realizations by specification: post-pandemic period. Specification labels: B excludes and T includes time-varying long-run growth; SV , DH , and O denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

B.2 RMSE by horizon and specification

Figures B.4–B.6 report RMSE by forecast horizon and model specification for each evaluation period. Rows are horizons and columns are specifications. Cell entries are RMSE; colors are normalized within each horizon, so the color scale compares specifications within a forecast horizon.

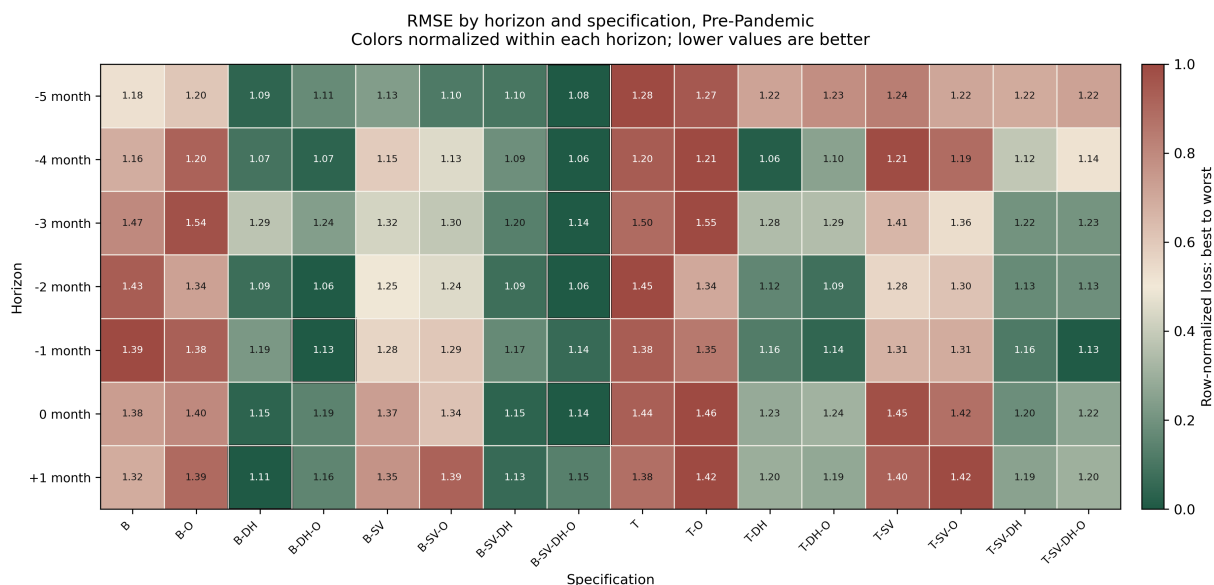


Figure B.4: RMSE by horizon and specification: pre-pandemic period. Cell entries are in percentage points of annualized quarter-on-quarter real GDP growth. Colors are normalized within each horizon; a black outline marks the lowest loss within that horizon. Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

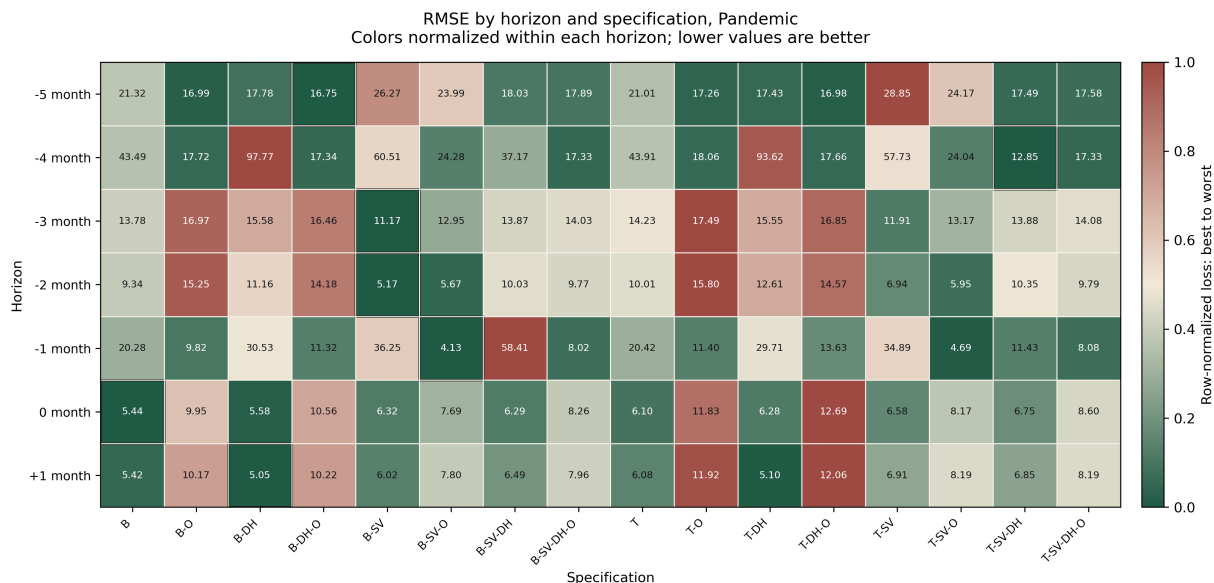


Figure B.5: RMSE by horizon and specification: pandemic period. Cell entries are in percentage points of annualized quarter-on-quarter real GDP growth. Colors are normalized within each horizon; a black outline marks the lowest loss within that horizon. Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

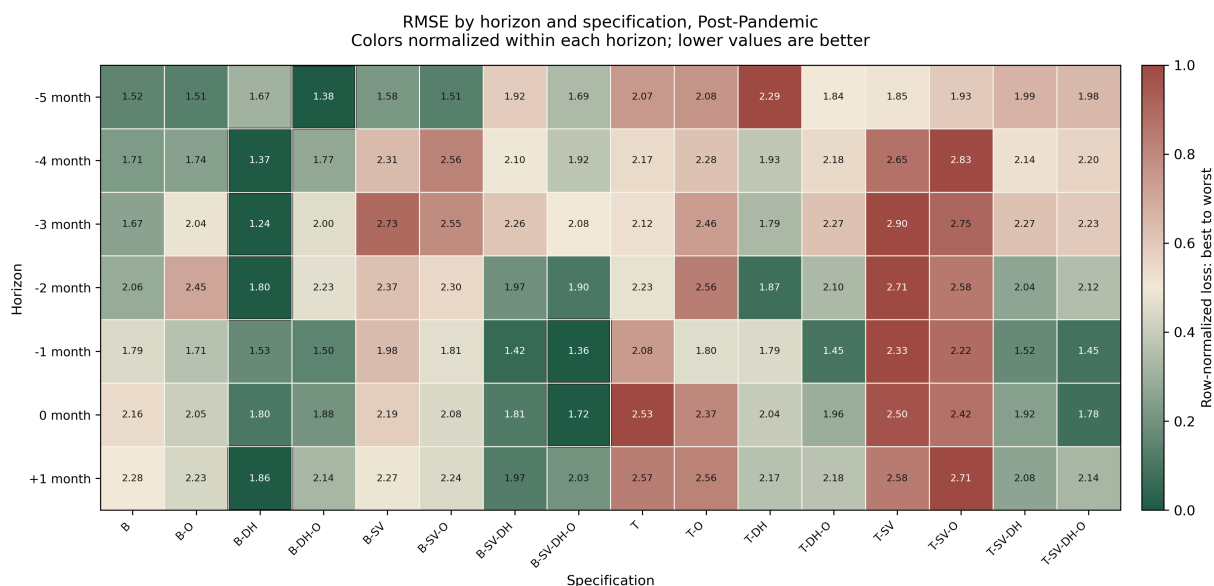


Figure B.6: RMSE by horizon and specification: post-pandemic period. Cell entries are in percentage points of annualized quarter-on-quarter real GDP growth. Colors are normalized within each horizon; a black outline marks the lowest loss within that horizon. Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

B.3 CRPS by horizon and specification

Figures B.7–B.9 report average CRPS by forecast horizon and model specification for each evaluation period. Rows are horizons and columns are specifications. Cell entries are average CRPS; colors are normalized within each horizon, so the color scale compares specifications within a forecast horizon.

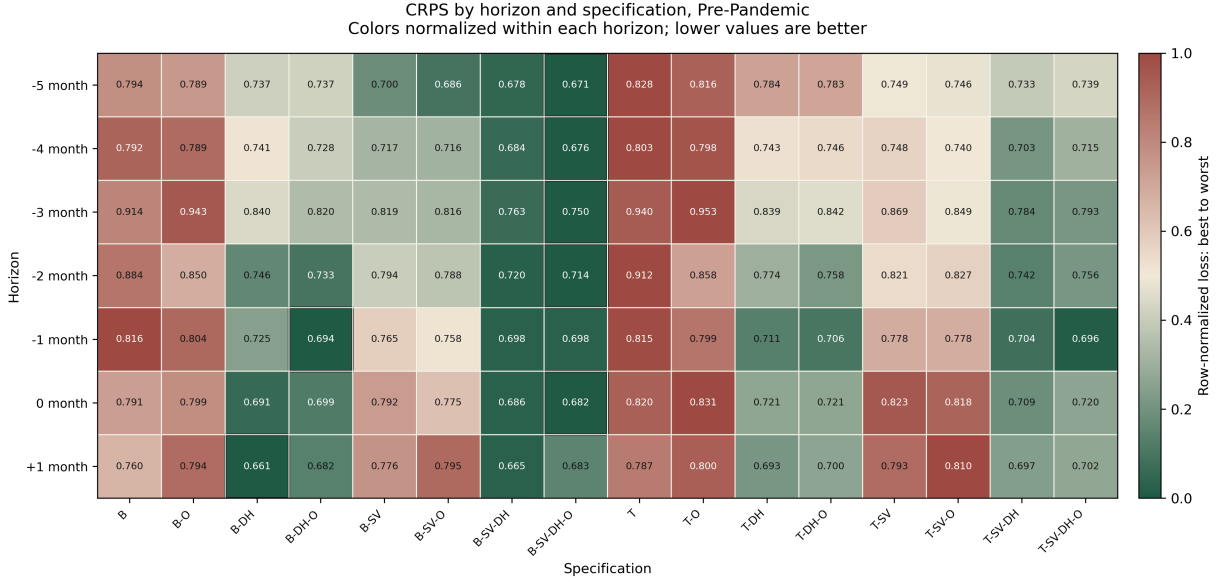


Figure B.7: CRPS by horizon and specification: pre-pandemic period. Cell entries are in percentage points of annualized quarter-on-quarter real GDP growth. Colors are normalized within each horizon; a black outline marks the lowest loss within that horizon. Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

B.4 Economic activity factor

Figures B.1–B.9 evaluate forecast accuracy using the pseudo-real-time exercise described in Section 4, in which every specification is re-estimated at each monthly forecast origin. This subsection instead reports a single, separate full-sample estimation of each of the 16 specifications in Figure 1: each specification is estimated once on the complete data set through the latest available vintage, rather than re-estimated month by month. This exercise is therefore distinct from, and not directly comparable to, the forecast-evaluation results reported elsewhere in the paper; it is intended only to visualize the common factor extracted by each specification. The underlying smoothed estimates condition on data through November 2025; Figures B.10–B.13 display observations only through December 2024 to match the

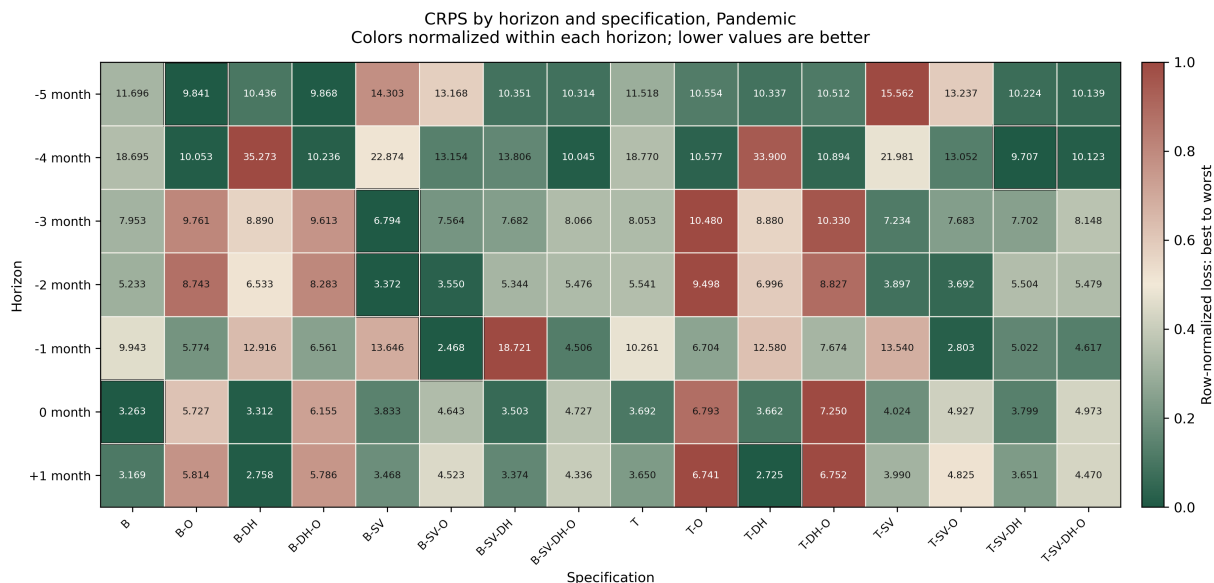


Figure B.8: CRPS by horizon and specification: pandemic period. Cell entries are in percentage points of annualized quarter-on-quarter real GDP growth. Colors are normalized within each horizon; a black outline marks the lowest loss within that horizon. Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

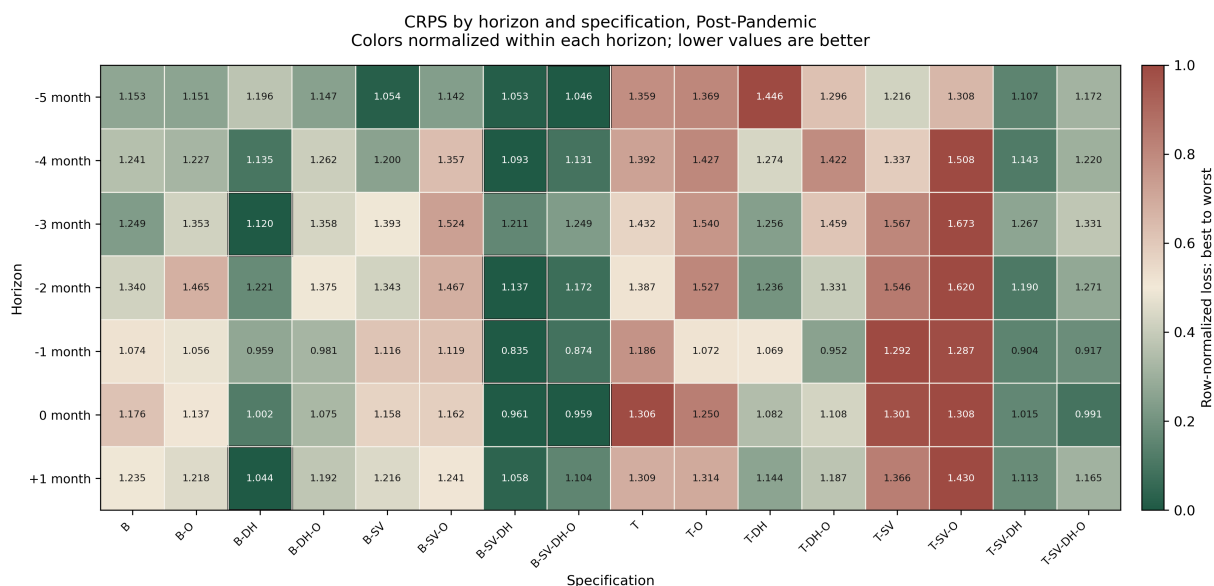


Figure B.9: CRPS by horizon and specification: post-pandemic period. Cell entries are in percentage points of annualized quarter-on-quarter real GDP growth. Colors are normalized within each horizon; a black outline marks the lowest loss within that horizon. Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

empirical period considered in the paper.

For each specification, Figure B.10 plots the posterior mean of the estimated common “economic activity” factor together with a 68 percent credible band (the 16th and 84th posterior percentiles across posterior draws) against two external benchmarks of real growth. PDFP growth is the compounded annual rate of change of real final sales to private domestic purchasers – GDP excluding government spending, net exports, and inventories – reported by the Federal Reserve Bank of St. Louis’s [FRED database](#). At a July 2025 press conference, Federal Reserve Chair Jerome Powell described private domestic final purchases as “a narrower but better signal for future, for where the economy is going” ([Powell, 2025](#)). GDPplus (also known as GDP+) is an alternative measure of real economic activity growth, produced by the Federal Reserve Bank of Philadelphia, that combines information from both the expenditure- and income-side estimates of GDP ([Aruoba et al., 2016](#)). Comparing the factor with both benchmarks, rather than with headline real GDP alone, provides an additional descriptive check on whether the latent factor co-moves with alternative measures of economic activity. Because the factor and the two growth series are on different scales, the comparison is informal: the factor is plotted on the left-hand axis and the two growth benchmarks on the right-hand axis. The two scales are common across panels, and the right-hand scale is positioned so that the pooled time average of PDFP and GDPplus coincides with zero on the factor scale. The series are therefore shown together to assess co-movement in timing and turning points rather than exact numerical correspondence.

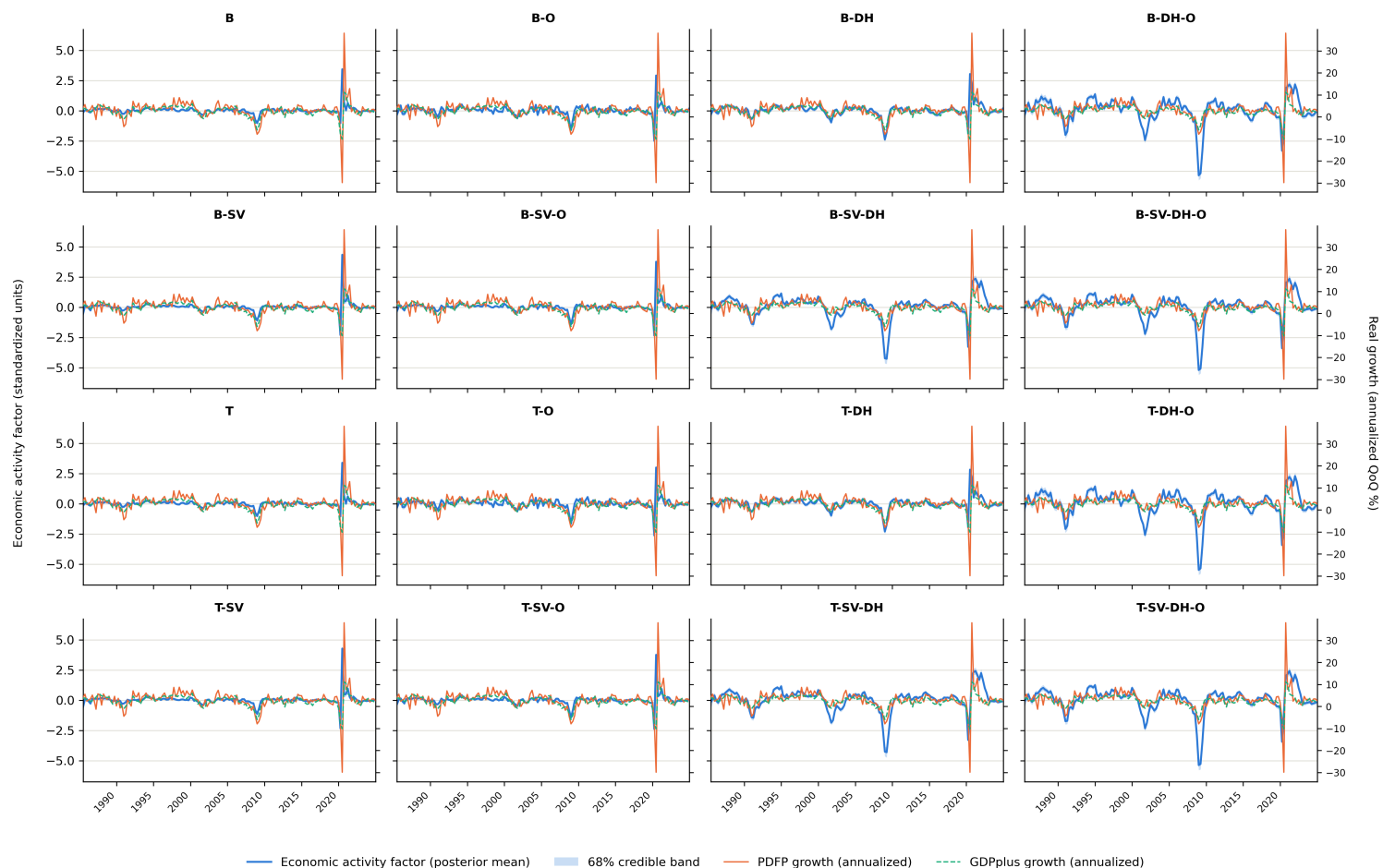


Figure B.10: Estimated economic activity factor and external growth benchmarks by specification. The posterior mean and 68 percent credible band of the factor are shown at calendar-quarter ends through 2024Q4. The factor is measured in standardized model units on the left-hand axis; PDFP and GDPplus are annualized quarter-over-quarter growth rates, measured in percent on the right-hand axis. Both axes are common across panels. PDFP denotes real final sales to private domestic purchasers; GDPplus combines expenditure- and income-side estimates of real GDP growth (Aruoba et al., 2016). Specification labels: *B* excludes and *T* includes time-varying long-run growth; *SV*, *DH*, and *O* denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

B.5 Additional model estimates

For the richest specification, $T\text{-}SV\text{-}DH\text{-}O$, the same full-sample estimation also delivers the shared long-run growth component g_t defined in equation (A.7) above. Because g_t enters the latent monthly equations for real GDP and GDI, we first map each posterior draw to quarterly frequency using the mixed-frequency weights in equation (A.10). We then undo GDP’s standardization with the mean and standard deviation used in estimation and convert the resulting annualized quarter-over-quarter path to a four-quarter growth rate.

Figure B.11 plots this quarterly long-run-growth component against PDFP and GDPplus four-quarter growth, putting the three series on a common, visually comparable scale. Its posterior mean is much smoother than the two benchmarks and is intended to capture low-frequency shifts in growth rather than business-cycle fluctuations.

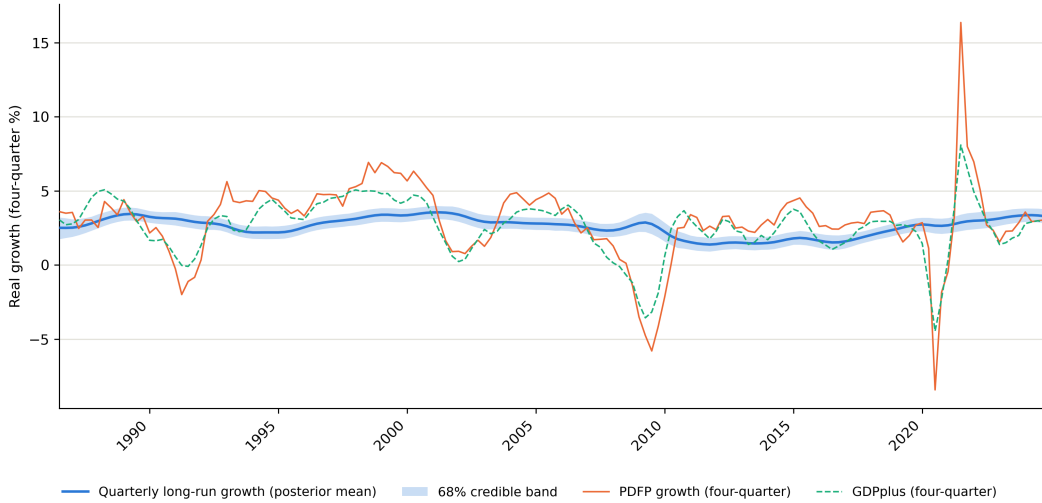


Figure B.11: Quarterly long-run-growth component implied by g_t for $T\text{-}SV\text{-}DH\text{-}O$: posterior mean and 68 percent credible band from the same full-sample estimation as Figure B.10, compared with PDFP and GDPplus four-quarter growth. Only observations through December 2024 are shown. Specification labels: B excludes and T includes time-varying long-run growth; SV , DH , and O denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

The same full-sample estimation of $T\text{-}SV\text{-}DH\text{-}O$ delivers the baseline idiosyncratic innovation variance $\exp(h_{\eta_{i,t}})$ for every series. Figure B.12 plots calendar-quarter-end readings of the posterior mean and 68 percent credible band of its standard deviation, $\sqrt{\exp(h_{\eta_{i,t}})}$, rescaled to the units reported in Table 1 and restricted to dates on which the series is observed. This baseline (“clean”) standard deviation excludes the multiplicative outlier scale and is intended to capture gradual, persistent changes in idiosyncratic volatility.

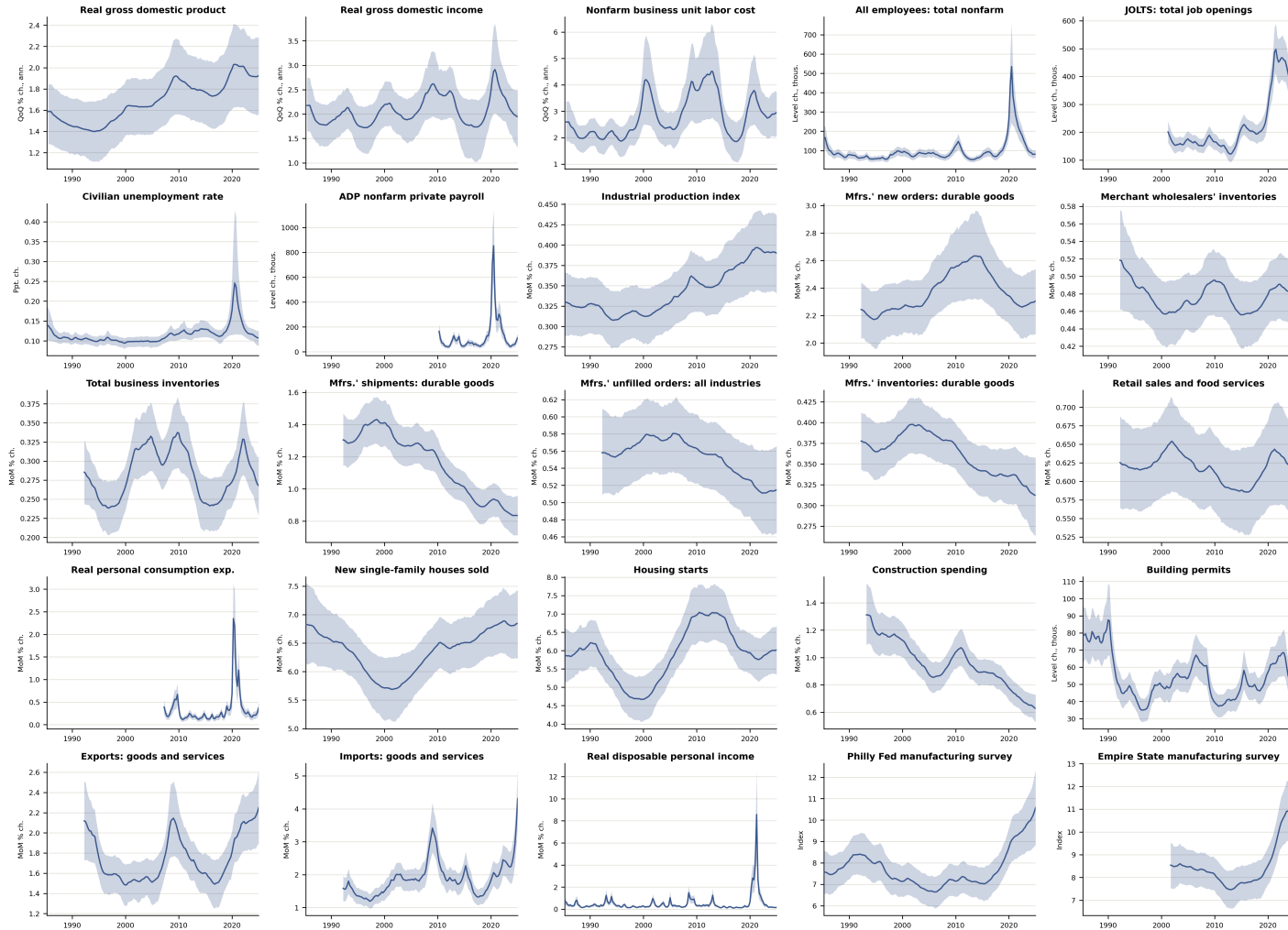


Figure B.12: Estimated idiosyncratic innovation standard deviation by series for T - SV - DH - O : posterior mean and 68 percent credible band of $\sqrt{\exp(h_{\eta_i,t})}$, rescaled to the units reported in Table 1. Calendar-quarter-end readings are shown through December 2024 and only over each series' own availability window, from the same full-sample estimation as Figure B.10. Specification labels: B excludes and T includes time-varying long-run growth; SV , DH , and O denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

Figure B.13 isolates how the multiplicative outlier scale changes the baseline idiosyncratic innovation volatility in the richest specification. Because $o_{i,t}$ is a strictly positive variance scale, the figure describes changes in dispersion. Positive (negative) values mean that the scale raises (lowers) the conditional standard deviation relative to the stochastic-volatility baseline, and zero corresponds to $o_{i,t} = 1$.

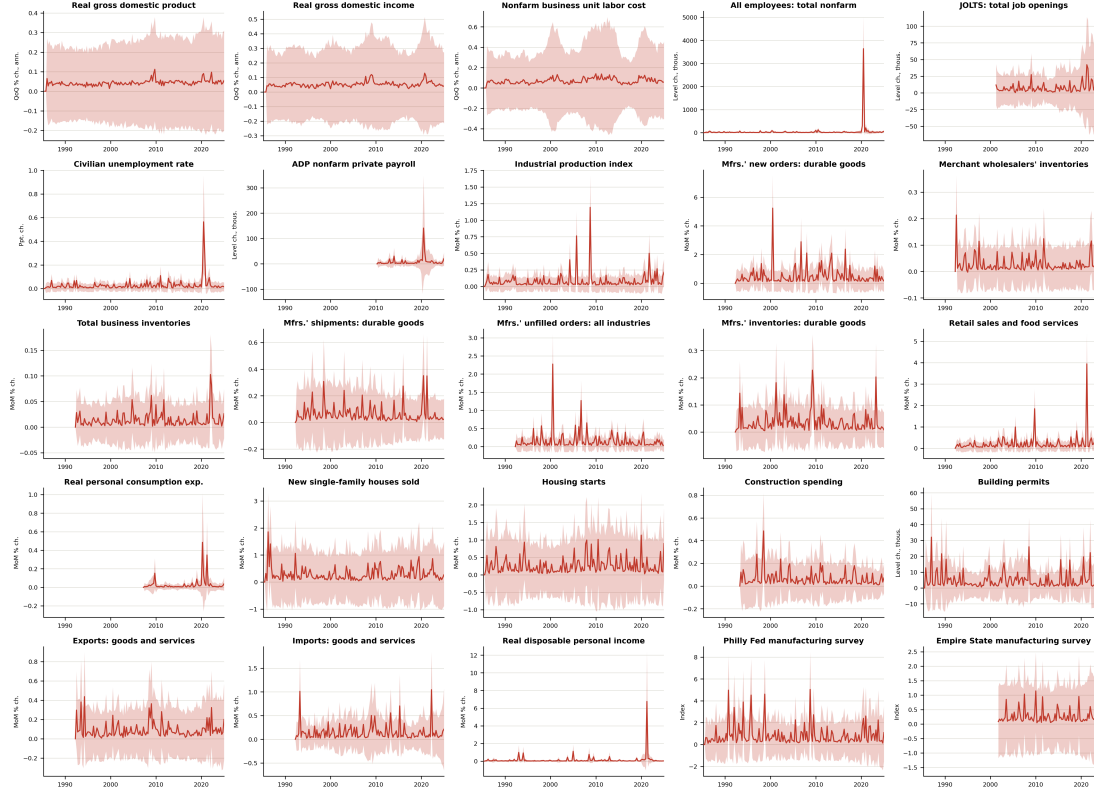


Figure B.13: Change in the idiosyncratic innovation standard deviation attributable to the multiplicative outlier adjustment, by series, for T - SV - DH - O : posterior mean and 68 percent credible band of $\sqrt{\exp(h_{\eta_i,t}) o_{i,t}} - \sqrt{\exp(h_{\eta_i,t})}$ in the units reported in Table 1. Calendar-quarter-end readings through December 2024 are shown over each series' availability window, using the same full-sample estimation as Figure B.10. Specification labels: B excludes and T includes time-varying long-run growth; SV , DH , and O denote stochastic volatility, dynamic heterogeneity, and outlier adjustment, respectively.

C Balanced paired factorial contrasts

C.1 Squared forecast error and CRPS losses

The squared forecast error and CRPS losses, and the corresponding balanced paired factorial contrast, are defined in Section 5 (equations 9–16). The overall and regime-specific effects for both losses are reported graphically in Figure 2; because that figure does not display the underlying bootstrap probabilities, Tables C.1 and C.2 below report the same effects and confidence intervals in numeric form, together with $P_{\text{boot}}(\Delta < 0)$.

Table C.1: Overall squared forecast error and CRPS effects

Feature	Effect %	90% CI	$P_{\text{boot}}(\Delta < 0)$
<i>Squared forecast error</i>			
Stochastic volatility	-12.2 [-23.5, 1.4]		0.93
Dynamic heterogeneity	-16.8 [-30.8, -1.4]		0.96
TV long-run growth	-13.3 [-34.3, 14.0]		0.80
Outlier adjustment	6.4 [-4.4, 18.9]		0.18
<i>CRPS</i>			
Stochastic volatility	-3.8 [-8.3, 0.9]		0.91
Dynamic heterogeneity	-7.6 [-10.9, -4.2]		1.00
TV long-run growth	3.8 [-0.5, 8.2]		0.07
Outlier adjustment	2.9 [0.2, 5.7]		0.04

Notes: Entries are percent loss changes, $100(\exp(\bar{d}_j) - 1)$. Intervals are 90 percent percentile intervals from 5,000 target-quarter bootstrap replications. $P_{\text{boot}}(\Delta < 0)$ is the share of bootstrap replications in which the feature lowers log loss. Negative effects indicate lower forecast loss. TV denotes time-varying.

Table C.2: Regime-specific squared forecast error and CRPS effects

Feature	Regime	Effect %	90% CI	$P_{\text{boot}}(\Delta < 0)$
<i>Squared forecast error</i>				
Stochastic volatility	Pre-pandemic	-8.5 [-23.2, 9.4]		0.80
Stochastic volatility	Pandemic	-18.8 [-39.3, 12.9]		0.85
Stochastic volatility	Post-pandemic	-13.5 [-35.8, 16.3]		0.78
Dynamic heterogeneity	Pre-pandemic	-25.8 [-41.3, -7.0]		0.99
Dynamic heterogeneity	Pandemic	-8.6 [-46.9, 38.3]		0.59
Dynamic heterogeneity	Post-pandemic	-5.6 [-28.6, 28.1]		0.64
TV long-run growth	Pre-pandemic	-0.0 [-33.6, 49.8]		0.49
TV long-run growth	Pandemic	-0.3 [-29.1, 33.8]		0.50
TV long-run growth	Post-pandemic	-37.7 [-63.5, 7.2]		0.92
Outlier adjustment	Pre-pandemic	-2.1 [-9.6, 5.7]		0.65
Outlier adjustment	Pandemic	-2.4 [-26.4, 24.8]		0.56
Outlier adjustment	Post-pandemic	29.7 [-1.8, 72.6]		0.06
<i>CRPS</i>				
Stochastic volatility	Pre-pandemic	-6.0 [-9.9, -1.8]		0.99
Stochastic volatility	Pandemic	5.9 [-11.1, 25.0]		0.29
Stochastic volatility	Post-pandemic	-6.3 [-12.1, 0.3]		0.94
Dynamic heterogeneity	Pre-pandemic	-8.1 [-11.9, -4.3]		1.00
Dynamic heterogeneity	Pandemic	-2.7 [-14.0, 8.3]		0.63
Dynamic heterogeneity	Post-pandemic	-9.9 [-15.0, -4.7]		1.00
TV long-run growth	Pre-pandemic	2.9 [-2.9, 9.2]		0.21
TV long-run growth	Pandemic	4.2 [-4.5, 13.1]		0.22
TV long-run growth	Post-pandemic	5.0 [-3.3, 14.0]		0.18
Outlier adjustment	Pre-pandemic	-0.3 [-1.2, 0.5]		0.72
Outlier adjustment	Pandemic	3.5 [-4.6, 11.6]		0.25
Outlier adjustment	Post-pandemic	8.0 [1.3, 15.7]		0.02

Notes: Entries are percent loss changes, $100(\exp(\bar{d}_{jp}) - 1)$. Intervals are 90 percent percentile intervals from 5,000 target-quarter bootstrap replications conducted within each regime. $P_{\text{boot}}(\Delta < 0)$ is the share of bootstrap replications in which the feature lowers log loss. Negative effects indicate lower forecast loss. TV denotes time-varying.

The regime-specific squared forecast error and CRPS effects in Table C.2 mirror the pattern discussed in Section 6: dynamic heterogeneity’s gains are concentrated before and after the pandemic rather than during it, stochastic volatility’s CRPS improvement is clearest before the pandemic, time-varying long-run growth’s squared-error gain after the pandemic is imprecisely estimated, and the outlier adjustment raises CRPS after the pandemic while showing no clear effect in the other two regimes.

C.2 Quantile scores

The quantile score and the corresponding balanced paired factorial contrast are defined in Sections 5.1 and 5.2, respectively (equations 11–17). The overall quantile-score effects are reported graphically in Figure 3; Table C.3 below reports the same effects in numeric form, together with $P_{\text{boot}}(\bar{d}_{j\alpha}^{QS} < 0)$, and Tables C.4 and C.5 report the regime-specific effects, split into lower- and upper-tail quantiles.

Table C.3: Overall quantile-score effects

Feature	Effect %	90% CI	$P_{\text{boot}}(\bar{d}_{j\alpha}^{QS} < 0)$
$\alpha = 0.05$			
Stochastic volatility	-4.8 [-11.4, 12.9]		0.64
Dynamic heterogeneity	-0.3 [-9.2, 3.1]		0.57
TV long-run growth	3.6 [1.5, 8.6]		0.00
Outlier adjustment	12.1 [-16.6, 27.5]		0.29
$\alpha = 0.10$			
Stochastic volatility	-3.8 [-8.8, 6.3]		0.70
Dynamic heterogeneity	0.3 [-7.6, 3.4]		0.47
TV long-run growth	2.8 [0.7, 6.8]		0.01
Outlier adjustment	4.4 [-17.0, 18.4]		0.37
$\alpha = 0.90$			
Stochastic volatility	-23.2 [-40.6, 27.4]		0.65
Dynamic heterogeneity	-6.2 [-17.5, -1.9]		1.00
TV long-run growth	1.5 [-5.6, 10.0]		0.31
Outlier adjustment	-22.8 [-40.2, 3.3]		0.88
$\alpha = 0.95$			
Stochastic volatility	-35.0 [-52.2, 31.4]		0.64
Dynamic heterogeneity	-4.8 [-17.7, 0.7]		0.91
TV long-run growth	0.5 [-6.9, 8.8]		0.40
Outlier adjustment	-23.9 [-51.6, 3.9]		0.89

Notes: Entries are percent changes in mean quantile score relative to the matched feature-off specifications. Bootstrap intervals are 90 percent percentile intervals from 5,000 target-quarter bootstrap replications. $P_{\text{boot}}(\cdot < 0)$ is the share of bootstrap replications in which the feature lowers the quantile score. Negative effects indicate lower quantile score. TV denotes time-varying.

Table C.4: Regime-specific quantile-score effects: lower tail

Feature	Regime	Effect %	90% CI	$P_{\text{boot}}(\bar{d}_{jp\alpha}^{QS} < 0)$
$\alpha = 0.05$				
Stochastic volatility	Pre-pandemic	-7.9 [-10.0, -6.0]		1.00
Stochastic volatility	Pandemic	-4.3 [-12.2, 36.7]		0.66
Stochastic volatility	Post-pandemic	-5.1 [-10.6, 0.4]		0.94
Dynamic heterogeneity	Pre-pandemic	-5.4 [-8.0, -2.5]		1.00
Dynamic heterogeneity	Pandemic	3.4 [-2.5, 9.2]		0.10
Dynamic heterogeneity	Post-pandemic	-17.5 [-26.1, -10.2]		1.00
TV long-run growth	Pre-pandemic	12.0 [9.6, 14.6]		0.00
TV long-run growth	Pandemic	2.8 [0.7, 9.1]		0.02
TV long-run growth	Post-pandemic	0.9 [-9.0, 14.5]		0.43
Outlier adjustment	Pre-pandemic	0.3 [-0.3, 0.9]		0.21
Outlier adjustment	Pandemic	15.2 [-28.4, 32.3]		0.29
Outlier adjustment	Post-pandemic	4.3 [-12.4, 20.0]		0.28
$\alpha = 0.10$				
Stochastic volatility	Pre-pandemic	-7.5 [-9.8, -5.4]		1.00
Stochastic volatility	Pandemic	-2.3 [-9.0, 25.2]		0.61
Stochastic volatility	Post-pandemic	-7.4 [-13.8, -2.4]		1.00
Dynamic heterogeneity	Pre-pandemic	-6.4 [-9.6, -2.8]		1.00
Dynamic heterogeneity	Pandemic	5.2 [0.8, 8.5]		0.04
Dynamic heterogeneity	Post-pandemic	-15.8 [-22.8, -9.5]		1.00
TV long-run growth	Pre-pandemic	14.4 [11.3, 17.9]		0.00
TV long-run growth	Pandemic	1.8 [-0.4, 7.0]		0.09
TV long-run growth	Post-pandemic	-3.1 [-11.4, 7.4]		0.73
Outlier adjustment	Pre-pandemic	0.5 [-0.2, 1.2]		0.14
Outlier adjustment	Pandemic	4.8 [-28.7, 23.3]		0.40
Outlier adjustment	Post-pandemic	6.3 [-7.5, 19.7]		0.23

Notes: Entries are percent changes in mean quantile score relative to the matched feature-off specifications. Bootstrap intervals are 90 percent percentile intervals from 5,000 target-quarter bootstrap replications conducted within each regime. $P_{\text{boot}}(\cdot < 0)$ is the share of bootstrap replications in which the feature lowers the quantile score. Negative effects indicate lower quantile score. TV denotes time-varying.

Table C.5: Regime-specific quantile-score effects: upper tail

Feature	Regime	Effect %	90% CI	$P_{\text{boot}}(\bar{d}_{jp\alpha}^{QS} < 0)$
$\alpha = 0.90$				
Stochastic volatility	Pre-pandemic	-8.7 [-13.7, -3.4]		0.99
Stochastic volatility	Pandemic	-28.9 [-45.9, 111.8]		0.66
Stochastic volatility	Post-pandemic	-0.9 [-11.4, 8.9]		0.57
Dynamic heterogeneity	Pre-pandemic	-4.9 [-13.4, 3.1]		0.84
Dynamic heterogeneity	Pandemic	-4.7 [-27.7, -0.3]		0.97
Dynamic heterogeneity	Post-pandemic	-15.1 [-19.5, -10.2]		1.00
TV long-run growth	Pre-pandemic	-9.8 [-13.2, -6.6]		1.00
TV long-run growth	Pandemic	0.0 [-11.7, 17.6]		0.46
TV long-run growth	Post-pandemic	22.9 [18.7, 28.7]		0.00
Outlier adjustment	Pre-pandemic	-1.2 [-2.5, -0.2]		0.98
Outlier adjustment	Pandemic	-30.4 [-58.4, 2.8]		0.93
Outlier adjustment	Post-pandemic	6.8 [4.4, 10.3]		0.00
$\alpha = 0.95$				
Stochastic volatility	Pre-pandemic	-9.9 [-13.4, -6.9]		1.00
Stochastic volatility	Pandemic	-41.7 [-56.9, 130.8]		0.67
Stochastic volatility	Post-pandemic	2.3 [-8.4, 12.2]		0.37
Dynamic heterogeneity	Pre-pandemic	-1.9 [-6.4, 2.6]		0.77
Dynamic heterogeneity	Pandemic	-3.6 [-28.3, 2.2]		0.80
Dynamic heterogeneity	Post-pandemic	-15.5 [-19.2, -11.6]		1.00
TV long-run growth	Pre-pandemic	-8.4 [-10.6, -6.2]		1.00
TV long-run growth	Pandemic	-0.7 [-12.5, 15.3]		0.54
TV long-run growth	Post-pandemic	19.1 [16.2, 22.8]		0.00
Outlier adjustment	Pre-pandemic	-0.6 [-1.5, 0.2]		0.90
Outlier adjustment	Pandemic	-30.1 [-70.4, 0.4]		0.95
Outlier adjustment	Post-pandemic	9.2 [6.8, 12.5]		0.00

Notes: Entries are percent changes in mean quantile score relative to the matched feature-off specifications. Bootstrap intervals are 90 percent percentile intervals from 5,000 target-quarter bootstrap replications conducted within each regime. $P_{\text{boot}}(\cdot < 0)$ is the share of bootstrap replications in which the feature lowers the quantile score. Negative effects indicate lower quantile score. TV denotes time-varying.

The quantile-score results do not overturn the density-forecast evidence. For dynamic heterogeneity, the 90th-percentile effect is the only overall quantile effect with an interval entirely below zero. The regime-specific estimates show lower-tail improvements before and after the pandemic and the clearest upper-tail improvements after the pandemic. The outlier adjustment has no clear overall effect at any of the four quantiles, but after the pandemic it raises the 90th- and 95th-percentile scores. Stochastic volatility and time-varying long-run growth also show effects that vary across quantiles and regimes.

D Moving-block bootstrap robustness

The target-quarter bootstrap described in Section 5.2 is matched to the multihorizon structure of the forecasting exercise: when a quarter is resampled for a given feature contrast, all seven horizons associated with that quarter are kept together. This preserves the dependence induced by evaluating several forecast horizons for the same realized GDP outcome. It does, however, treat the resulting quarter-level objects as independent. This may be restrictive because forecasts for neighboring quarters are produced from heavily overlapping information sets.

As a robustness check, Table D.1 reports a moving-block bootstrap over target quarters. The procedure forms overlapping four-quarter blocks, resamples these blocks with replacement, concatenates them, and truncates the resulting sequence to the length of the evaluation sample. All seven horizons associated with each selected quarter are retained before recomputing the average balanced paired factorial contrast. We report this exercise only for overall effects. The regime-specific samples are too short, especially during the pandemic, for block resampling to be informative.

Table D.1: Overall effects with a four-quarter moving-block bootstrap

Feature	Effect %	90% CI	$P_{\text{MBB}}(\Delta < 0)$
<i>Squared forecast error</i>			
Stochastic volatility	-12.2	[-22.1, 0.1]	0.95
Dynamic heterogeneity	-16.8	[-25.7, -5.5]	0.99
TV long-run growth	-13.3	[-30.2, 4.6]	0.90
Outlier adjustment	6.4	[-5.0, 23.0]	0.23
<i>CRPS</i>			
Stochastic volatility	-3.8	[-7.4, 3.0]	0.81
Dynamic heterogeneity	-7.6	[-10.4, -4.8]	1.00
TV long-run growth	3.8	[0.1, 6.3]	0.04
Outlier adjustment	2.9	[0.1, 6.5]	0.04

Notes: Entries are percent loss changes, $100(\exp(\bar{d}) - 1)$. Intervals are 90 percent percentile intervals from 5,000 moving-block bootstrap replications using overlapping four-quarter blocks of target quarters. $P_{\text{MBB}}(\Delta < 0)$ is the share of block bootstrap replications in which the feature lowers log loss. Negative effects indicate lower forecast loss. TV denotes time-varying.

The moving-block bootstrap leaves the main qualitative ranking unchanged. Dynamic heterogeneity remains the only feature with intervals entirely below zero under both loss functions. The squared-error effect of the outlier adjustment remains imprecise, while its CRPS effect is 2.9 percent with a 90 percent interval of 0.1 to 6.5 percent. Stochastic volatility remains imprecise. Time-varying long-run growth has an imprecise squared-error effect and raises CRPS by 3.8 percent, with a 90 percent interval of 0.1 to 6.3 percent.

References

- Almuzara, M., Baker, K., O’Keeffe, H., and Sbordone, A. (2023). The New York Fed Staff Nowcast 2.0. New York Fed Staff Nowcast Technical Paper, Federal Reserve Bank of New York. <https://www.newyorkfed.org/medialibrary/media/research/blog/2023/NYFed-Staff-Nowcast-technical-paper>.
- Antolín-Díaz, J., Drechsel, T., and Petrella, I. (2017). Tracking the slowdown in long-run GDP growth. *The Review of Economics and Statistics*, 99(2):343–356.
- Antolín-Díaz, J., Drechsel, T., and Petrella, I. (2024). Advances in nowcasting economic activity: The role of heterogeneous dynamics and fat tails. *Journal of Econometrics*, 238(2):105634.
- Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorfheide, F., and Song, D. (2016). Improving GDP measurement: A measurement-error perspective. *Journal of Econometrics*, 191(2):384–397.
- Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, pages 195–237. Elsevier.
- Carriero, A., Clark, T. E., and Marcellino, M. (2015). Realtime nowcasting with a bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A*, 178(4):837–862.
- Carriero, A., Clark, T. E., Marcellino, M., and Mertens, E. (2024). Addressing COVID-19 outliers in BVARs with stochastic volatility. *The Review of Economics and Statistics*, 106(5):1403–1417.
- Carriero, A., Galvão, A. B., and Kapetanios, G. (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting*, 35(4):1226–1239.

- Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3):327–341.
- Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575.
- D’Agostino, A., Gambetti, L., and Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1):82–101.
- D’Agostino, A., Giannone, D., Lenza, M., and Modugno, M. (2016). Nowcasting business cycles: A bayesian approach to dynamic heterogeneous factor models. In Hillebrand, E. and Koopman, S. J., editors, *Dynamic Factor Models*, volume 35 of *Advances in Econometrics*, pages 569–594. Emerald Group Publishing Limited.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.
- Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.
- Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.
- Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting gdp and inflation: The real-time informational content of macroeconomic data releases. *Journal of Monetary Economics*, 55(4):665–676.
- González-Astudillo, M. and Baquero, D. (2019). A nowcasting model for Ecuador: Implementing a time-varying mean output growth. *Economic Modelling*, 82:250–263.
- Hong, S. S. (2025). Nowcasting economic activity with fat tails and outliers. Working Papers halshs-05046472, HAL.
- Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185–212.
- Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3):361–393.
- Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443.

- Powell, J. H. (2025). Transcript of Chair Powell’s press conference, July 30, 2025. Press conference transcript, Board of Governors of the Federal Reserve System, Federal Open Market Committee.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852.
- Quaadvlieg, R. (2021). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, 39(1):40–53.
- Stock, J. H. and Watson, M. W. (2016). Core inflation and trend inflation. *The Review of Economics and Statistics*, 98(4):770–784.