Department of Economics

A GDP Nowcast for the UK

UCL Macro Monitor

May 2026

Corrado Calemma Tiziano M. Di Biase Thomas Lazarowicz

This document is a companion to the model description. It describes the real-time process used to update nowcasts as new data arrives, the backtesting used to evaluate model candidates, and the calibration process we follow to select our production model.

1 Real-time construction and evaluation

We use two types of model run. A *full* re-estimation calls the Gibbs sampler described in the model description on the full dataset, producing a joint posterior over parameters and latent states. A *partial* re-estimation uses the parameter posterior from the most recent full re-estimation and updates only the latent states for new data releases. Full re-estimation is triggered by each GDP outturn release; partial re-estimation runs on each intervening data release. The motivation for the split is twofold. First, partial re-estimation means that we can utilise the model's outlier detection algorithm to decide whether a release contains news or is likely a historical outlier (without running a full re-estimation). Secondly, following a quarterly re-estimation schedule means that parameters change at the pace of the slowest signal in our model (quarterly GDP), while the latent state changes at the pace of each release. Holding parameters fixed between GDP releases means that we can easily isolate the news content of each release and produce an updated density nowcast at each release without conflating parameter re-estimation and news.

Mini-Gibbs procedure. Partial re-estimation is non-trivial under the Bayesian framework adopted here. Two issues prevent a standard Kalman filter and simulation smoother combination from being used directly. First, the object held fixed between full re-estimations is not a single point estimate of the parameters but a posterior distribution. Second, the outlier component introduces a non-linearity that makes the joint sampling of factors, innovations, and outlier scales infeasible even when the parameters are known with point precision. We resolve both by replacing the analytical Kalman recursion with a nested Monte Carlo procedure, which we refer to as a "mini-Gibbs".

From the most recent full re-estimation, the parameter posterior is approximated by D independent draws. For each draw, parameters are held fixed at the drawn values and a reduced Gibbs sampler is run on the latent states alone - factors, trend, idiosyncratic components, log-volatilities, and outlier scales - via a recursive sequence of Kalman filtering and simulation smoothing steps. The output is D independent posterior trajectories of the latent states conditional on the new vintage. The full posterior of the latent states under the partially re-estimated model is obtained by aggregating across the D parameter draws.

This procedure is, in principle, no faster than a single full re-estimation, however unlike the full re-estimation this process is embarrassingly parallel. In production we use $D = 100$ parameter draws and $R = 1{,}000$ inner iterations..

2 Backtest design

To evaluate the model, we construct a real time historical vintage dataset. Each vintage is a snapshot of the macroeconomic information set as it appeared on any given day: each observation is the value that was available on a given data without the benefit of subsequent revisions (in this sense a revision can be thought of as a new release about historical information). The full sample of vintages used for backtesting and calibration covers 2018Q1 to 2026Q1. For evaluation purposes we produce these vintages on any given Wednesday rather than for every given release.

At each weekly vintage the model produces a density forecast of quarterly GDP growth at multiple horizons. For evaluation, we focus on horizon $h = 0$, the nowcast of the current target quarter, where the target quarter is the next quarter for which a GDP outturn has not yet been released, and summarise the posterior by its mean.

Two RMSE metrics are reported:

- *Last h=0 RMSE*: for each target quarter, the h=0 forecast from the last vintage before the GDP release for that quarter is extracted and squared error against the realised outturn is computed. This produces one observation per target quarter. This is the metric most directly comparable to externally published nowcasts, which report a single best-by-release forecast per quarter.
- *Average h=0 RMSE*: squared errors averaged across all weekly vintages whose target is a given quarter, weighting forecasts made when the information set is sparse equally with those made when it is rich.

Each metric is reported on three sub-samples: the full panel of target quarters in 2018Q1–2026Q1, a sub-sample excluding the COVID period (2020-2021), and the post-COVID sub-sample (2022Q1 onward). We exclude the COVID sub-sample from our headline assessment because the 2020Q2–Q3 outturns—a 20% contraction followed by a 17% rebound—are large enough to dominate full-sample RMSE on their own and are structurally distinct from the cyclical fluctuations the model is designed to track. The headline metric throughout this document is post-COVID Last h=0 RMSE.

3 Variable panel and factor count

The first calibration choice is over variable selection and factor counts. We consider five sets of variables and up to three factors for each for a combination of 15 cells. For the sets of variables we consider:

- 22 indicators (1 quarterly + 21 monthly) A parsimonious set of macroeconomic variables for the UK, supplemented by Euro Area survey and activity indicators. Full list in the model description.
- The same set of 22 indicators detailed above augmented with nine UK financial market series (FTSE All-Share, three sterling exchange rates, three gilt yields, Brent crude); 31 variables total.
- 24 indicators corresponding to the panel of Anesti et al. (2022), with the closest available substitutes for series we do not have in our pipeline (gilt yields proxy for the term spread).
- A broader panel of 39 indicators, all surveys and financial data detailed above, with the addition of labour market variables and house price series.
- An exhaustive pool of 65 indicators from the data pipeline; included as an upper bound on panel size.

Within each set of variables, factor counts of 1, 2, and 3 were tested. For each (preset, factor) combination, the hyperparameter grid described in Section 4 is run, and the best in cell post-COVID Last h=0 RMSE is recorded. The winning factor count for each is reported in Table 1.

The compact `v22_nc` panel at three factors dominates. The production specification is therefore 22 indicators with three common cyclical factors plus a stochastic GDP trend.

Table 1: Best-in-cell post-COVID Last $h=0$ RMSE across the variable-panel/factor-count grid.

Preset	Factors	Post-COVID RMSE
`22`	3	**0.870**
`anesti`	3	1.060
`39`	2	1.190
`22 + financial`	2	1.259

4 Hyperparameter calibration

Conditional on the panel and factor count, prior hyperparameters are calibrated via a 256-cell grid sweep over five dimensions:

- Trend innovation prior scale: $S_\tau \in \{0.005, 0.010, 0.020, 0.030\}$
- Trend innovation prior degrees of freedom: $\nu_\tau \in \{1, 5, 7, 10\}$
- Factor VAR Minnesota tightness: $\tau_{\text{VAR}} \in \{0.5, 0.7, 0.9, 1.1\}$
- Outlier Student-t degrees of freedom: $\nu_o \in \{4, 8\}$
- Loading Minnesota tightness: $\tau_\lambda \in \{0.5, 0.8\}$

Each cell of the grid is run as an independent backtest at a screening MCMC budget of 4,000 Gibbs draws (1,000 burn-in) per re-estimation; the winner is then re-run at production budget (8,000 / 2,000) for the figures reported elsewhere in this document. The screening budget is appropriate for ranking configurations on RMSE but not for producing final-quality posteriors.

Beyond the five canonical dimensions, a broader 836-configuration sweep additionally varies the Minnesota lag-decay parameters α_{VAR} and α_λ, the level of the factor and idiosyncratic stochastic-volatility priors, the volatility-of-volatility scales, and the idiosyncratic AR coefficient priors $\rho_{i,1}, \rho_{i,2}$. Within sensible neighbourhoods of the production values, perturbing these axes has small impact on post-COVID Last $h=0$ RMSE relative to the five canonical dimensions; they are therefore held fixed at the values listed in the model description's prior calibration table for the canonical grid above.

5 Computational resource

The calibration grid is run on UCL's Myriad HPC cluster. Each cell of the grid - one (preset, factor count, hyperparameter configuration, seed) tuple - is an independent job: a single backtest covering 2018Q1-2026Q1 on a single CPU core with single-threaded BLAS. To ensure reproducibility all jobs are run on a fixed node class with identical CPUs.

Per-job resources. An eight-year weekly backtest comprises 32 full re-estimations and roughly 420 partial-re-estimation (mini-Gibbs) updates, taking 20–24 wall-clock hours per cell.

6 Robustness

The chosen configuration is subjected to three robustness checks.

Seed sensitivity. The top five configurations from the 22 variable 3 factor sweep are re-run across a panel of RNG seeds, with full posterior parameter draws retained at each re-estimation, to verify that the cross-cell ranking is not an artefact of a single seed. Geweke convergence diagnostics and effective sample sizes for the leading factor loading are computed for each (configuration, seed) pair.

Local sensitivity around the winner. A fine-grained local sweep replicates the best performing configuration at neighbouring values of S_τ with all other hyperparameters held fixed, to confirm that 0.010 is not a knife-edge optimum and to characterise the local response of post-COVID RMSE to perturbations of the trend prior.

MCMC convergence at production settings. At the production MCMC budget (8,000 draws, 2,000 burn-in), Geweke p-values and effective sample sizes are computed for the leading factor loading at each full re-estimation across the 2018 - 2026 backtest.

7 Benchmark comparison

As a headline external benchmark for our UK GDP nowcasts we use a private sector nowcast provider, which publishes a single best-by-release forecast per target quarter using a related but independent dynamic factor model. Forecasts from both models are aligned to the same target quarters and the same release dates. On post-COVID Last h=0 RMSE - the metric on which benchmark's published performance are most directly comparable with our model - the production model achieves parity.

8 Production specification

Our backtesting and calibration procedure set out above chooses the variables, factor structure and hyperparameter set:

- *Variable panel*: 22 indicators (1 quarterly GDP target plus 21 monthly indicators). Full list in the model description.
- *Factor count*: 3 common cyclical factors plus a stochastic GDP trend.
- *Identification*: As documented in the model description (only GDP loads on the trend; three diagonal-unity normalisers fixed at GDP, GfK Consumer Confidence, and Industrial Production).
- *Hyperparameters*: $S_\tau = 0.010$, $\nu_\tau = 1$, $\tau_{\text{VAR}} = 0.9$, $\nu_o = 8$, $\tau_\lambda = 0.8$. All other priors as documented in the model description.
- *MCMC*: 8,000 Gibbs iterations with 2,000 burn-in (6,000 retained); no thinning. Mini-Gibbs at partial re-estimation: 100 parameter draws $\times$ 1,000 inner iterations.
- *Performance*: post-COVID Last h=0 RMSE of 0.870, at parity with the external benchmark on the same target quarters.

References

Antolin-Diaz, J., Drechsel, T., & Petrella, I. (2021). Advances in nowcasting economic activity: The role of heterogeneous dynamics and fat tails. *Journal of Econometrics*, 238(2), 105632.

Anesti, N., Galvão, A. B., & Miranda-Agrippino, S. (2022). Uncertain Kingdom: nowcasting GDP and its revisions. *Journal of Applied Econometrics*, 37(1), 42–62.