



MONASH
University

MONASH
BUSINESS
SCHOOL

**Department of
Econometrics &
Business Statistics**

☎ (03) 9905 2478
✉ BusEco-Econometrics@monash.edu

ABN: 12 377 614 012

Quarterly Estimation of Gross State Product

Arun Krishnasamy

Jasmine Zheng

Report for
Department of Treasury and Finance

3 November 2025



1 Abstract

Gross State Product (GSP) measures the total value of goods and services produced within a state's economy. It's important because it shows how each state contributes to national growth and helps governments and businesses make informed policy and investment decisions.

A R package, targets-based pipeline collects ABS and RBA data, applies consistent transformations, aligns monthly and quarterly series, and keeps a clear record for rebuilding. On top of this data layer, tested basic bridge models and a VAR, and then developed a mixed-frequency Bayesian VAR with stochastic volatility that enforces coherence by linking states to the national total.

Despite careful diagnostics, several problems remain, such as issues with identification and scaling, sensitivity to prior choices, sampler errors (divergences and poor mixing), and timing mismatches. These have so far prevented reliable results.

The main achievement to date is a solid and transparent data pipeline, a clear record of what went wrong, and a plan for moving forward: build model complexity in stages (from fixed-volatility VAR to full stochastic volatility, then to mixed-frequency), use more structured priors, reduce dimensionality before scaling up, and add formal checks for coherence and backtesting once the model converges.

While results are still preliminary, the framework, lessons learned, and troubleshooting steps are broadly useful for other hierarchical nowcasting problems where regional and national data must be consistent and arrive at different time intervals.

2 Introduction

Nowcasting Australia's Gross State Product (GSP) is an important but challenging task. Governments, analysts, and businesses rely on timely signals of state-level economic activity to guide policy, budgeting, and resource allocation. However, the data needed for this arrive at "different frequencies", are revised over time, and were disrupted by the COVID-19 pandemic. Another key requirement is coherence.

To show why this matters in our data, Figure Figure 1 plots the historical coherence gap—the difference between the weighted sum of state signals and the national series—as a percent of national GDP. The gap is typically within ~1–3%, with sharp spikes around the 2008(Global financial crisis) and 2020–21(COVID-19 Pandemic), motivating explicit reconciliation in our workflow.

The state estimates must add up to the national total so that any practical system must combine information across states and frequencies (Koop et al. [2020](#); Department of Treasury and Finance

Victoria 2021), while remaining transparent and easy to maintain. This project set out to build such a system with full reproducibility.

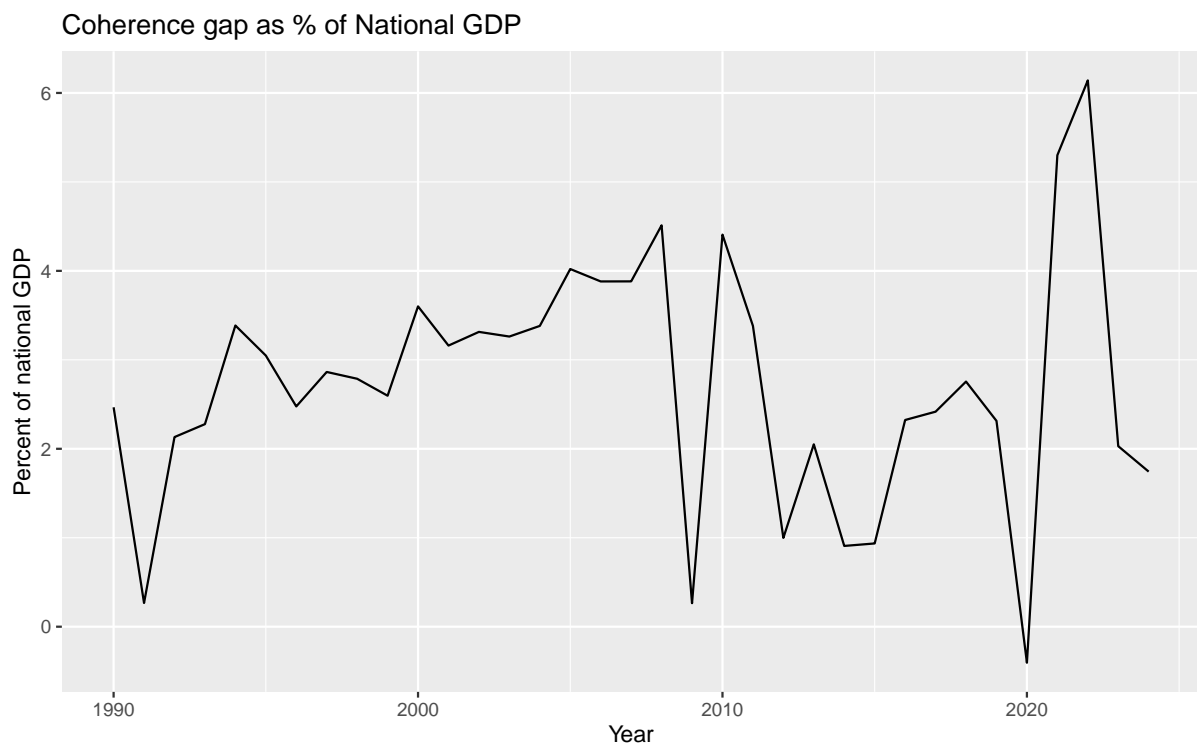


Figure 1: *Coherence gap as percentage of National GDP, National - State*

So far, the work has produced a strong data foundation: a targets-based pipeline that pulls ABS and RBA data, applies consistent transformations through a clear schema, aligns monthly indicators to quarterly GSP (accounting for lags and missing data), and maintains a full audit trail for rebuilds. On top of this base, We have tested progressively more complex models—starting with simple benchmarks, then a standard VAR, and finally a mixed-frequency Bayesian VAR with stochastic volatility. This last model enforces coherence through a national-as-weighted-sum structure. A diagnostics-first workflow guided all stages, focusing on convergence, mixing, and posterior predictive checks before drawing any conclusions.

The models have not yet produced a stable or reliable posterior. The main difficulties involve identification and scaling (sensitivity to priors and standardization choices), timing mismatches between monthly and quarterly data, and sampler issues such as divergences and poor mixing when stochastic volatility and many variables are introduced together. Rather than hide these problems, the report highlights them—what was attempted, what failed, and what those failures reveal. The project’s key achievements are threefold:

- a reproducible and well-documented data pipeline;
- a clear record of challenges linked to specific diagnostics; and

- a practical roadmap for progress—building model complexity in stages (fixed-volatility VAR > Stochastic Volatility > mixed-frequency), refining and structuring priors, reducing dimensionality before scaling up, and adding formal coherence and backtesting checks.

Together, these steps aim to turn the current foundation into a working, credible nowcasting framework.

3 Data & Reproducible Pipeline

3.1 Data sources and coverage

The foundation of the nowcasting framework rests on a carefully curated and well-documented data layer. Because accuracy in real-time monitoring depends as much on data quality as on modeling sophistication, this stage focuses on building a consistent, transparent, and easily extensible source map. The dataset brings together national and state-level series from the ABS and RBA, chosen for their timeliness, reliability, and long historical coverage. Each series is stored with full metadata to support reproducibility and future revisions.

The core aggregates include State Final Demand (SFD) (Australian Bureau of Statistics [2025a](#)) and Gross State Product (GSP) (Australian Bureau of Statistics [2025b](#)) for each state, along with national totals from the ABS. These are aligned to Australia's standard quarterly calendar (Q1–March, Q2–June, Q3–September, Q4–December), forming the benchmark targets for reconciliation and model training.

To capture broader macroeconomic conditions, a macro-exogenous set is incorporated. The current working group of indicators covers the RBA's cash rate target, trimmed-mean inflation (Australia-wide), the Trade-Weighted Index (TWI), and the RBA commodity price index (CMPI). These variables serve as common national drivers and help provide context for state-level dynamics.

For every series, metadata are systematically tracked. This includes the source table or link, units, frequency, whether the data are seasonally adjusted or original, geographic scope, start and end periods, revision policy, and details of the transformation pipeline. Such documentation ensures that each step—from download to transformation—can be replicated precisely.

A key design choice throughout is to prioritize the cleanest available seasonally adjusted (SA) series from official sources. Where both SA and original data are published, only the SA version is stored within the schema for modeling, minimizing noise from seasonal patterns and reducing preprocessing complexity later in the workflow.

2.2 Schema & transform governance

A key organizing principle of the framework is maintaining a single source of truth for all data definitions. This is achieved through a central schema file, `schema.csv` (or `schema_mfbvar.csv` for model-specific builds)—which fully describes every series used in the pipeline. Each row in this schema represents one time series and includes columns such as: `series_id`, `alias`, `source`, `freq`, `unit`, `geo`, `sa_flag`, `first_date`, `last_date`, `transform_chain`, and `coherence_role` (for example, `state_component`, `national_anchor`, or `exog`).

This design allows the schema to act as both documentation and instruction. Each transform in the

chain corresponds to a predefined function. The declared syntax specifies the order of operations, not the implementation details, so that transformation logic is no longer hard-coded inside notebooks or scripts. Instead, the pipeline reads these steps directly from the schema, applies them in sequence. This approach enforces consistency across models, simplifies debugging, and ensures that every transformation can be traced, reproduced, or modified without breaking the broader system.

3.2 Pipeline architecture

Data acquisition and preparation are structured into clear stages. The pipeline, shown in Figure 2, retrieves raw data from both ABS and RBA sources using schema-driven registry functions (`fetch_abs_by_registry()` and `fetch_rba_by_registry()`), then combines these datasets into a unified raw dataset. A preliminary cleaning and preparation step (`prep_step1()`) standardizes formats and ensures consistency across series.

All targets are managed using the `{targets}` package, which provides caching, reproducibility, and deterministic rebuilds. Each stage is stored as a target (`p11_abs_raw`, `p12_rba_raw`, `p1_abs_rba_raw`, `p21_raw`, etc.), ensuring that only modified inputs trigger re-computation.

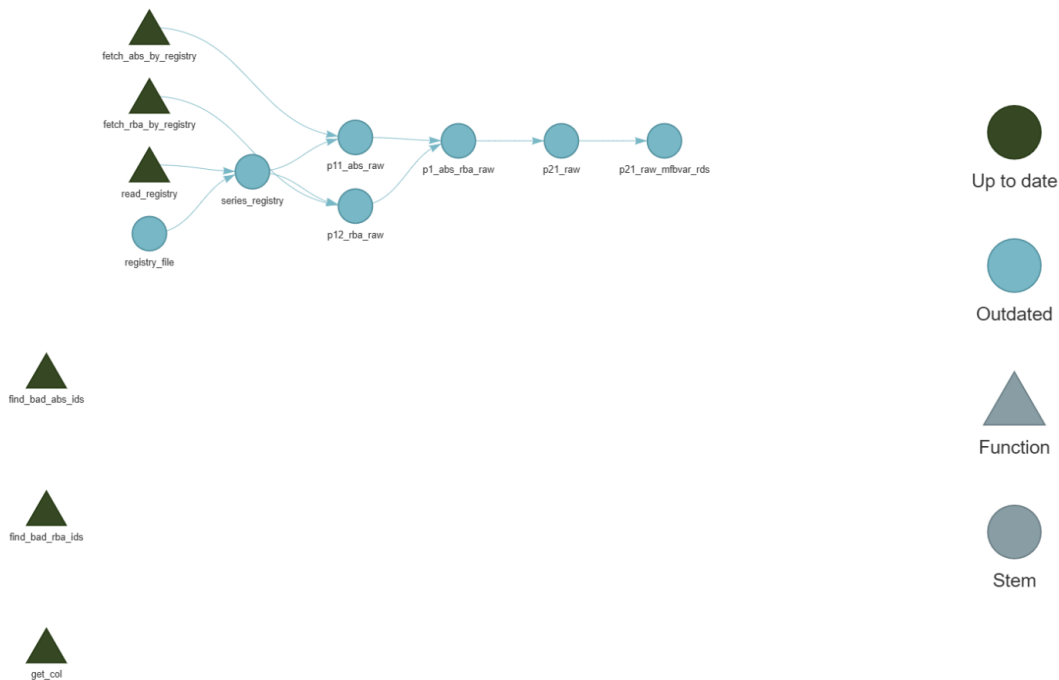


Figure 2: The registry is read and validated, ABS/RBA series are fetched and merged (`p11_abs_raw`, `p12_rba_raw` → `p1_abs_rba_raw`), cleaned into `p21_raw`, and serialized as `p21_raw_mfbvar_rds`. Node colour shows build status at render (dark green = up to date; blue = outdated).

4 Baseline Forecasting

A three simple reference models, a level-naive model, an AR-only ARIMA (no differencing or MA terms), and a trend-plus-season regression (TSLM), to serve two purposes:

- To sanity-check the data pipeline and train/test split.
- to set a baseline for future VAR and mixed-frequency BVAR modeling.

Using a single-origin holdout of the last eight quarters per state, generated 8-step-ahead forecasts and evaluated RMSE, MAE, and sMAPE. Furthermore, one-step-ahead (next-quarter) forecasts were produced from the full sample to illustrate directional behavior.

Across all states, the naive model, which simply repeats the last observed level, consistently performed best on RMSE, MAE, and sMAPE. The AR-only of ARIMA, restricted to no differencing or MA terms, had the poorest performance, often substantially so. The TSLM, incorporating a linear trend and seasonal effects, performed between the two. This pattern was most pronounced in larger states (NSW, VIC, QLD, WA), where the AR-only model struggled to capture level shifts and shocks, while the naive model maintained low sMAPE due to the persistence of GSP. Smaller states (ACT, TAS, NT) showed the same ranking.

The next-quarter forecasts reflected a similar ordering: the naive model extrapolated recent levels, the TSLM smoothed toward the underlying trend and seasonality, and the AR-only model under-reacted to changes, consistent with its poor multi-horizon performance.

These results, from table Table 1 , confirm that simple persistence is difficult to beat for quarterly GSP in its current transformed form unless the model explicitly accounts for structural breaks, level shifts, or remaining seasonality. The AR-only restriction likely causes underfitting of level changes and shock propagation, while the global linear trend and seasonal terms in TSLM can be fragile in the post-COVID environment.

These findings establish a clear roadmap, outperform naive persistence with, then introduce low-dimensional VARs with fixed volatility, and finally incorporate stochastic volatility and mixed-frequency features always maintaining consistent evaluation methods to ensure improvements are real and attributable.

Table 1: *Baseline forecasting accuracy by state over the last eight quarters (holdout): Quarterly GSP levels (AUD millions) evaluated using three simple models: Naïve (mean-last), OLS (trend + season), and AR(p), $d = 0$. Metrics reported are RMSE, MAE, and sMAPE (%)—lower is better. Bold indicates the best (lowest) error for each state and metric. This table provides the benchmark against which MF-BVAR-SV results are compared.*

State	Method	RMSE	MAE	sMAPE (%)
ACT	Naïve (mean-last)	1,689	1,599	12.67
ACT	OLS (trend + season)	2,454	2,424	19.92
ACT	AR(p), $d=0$	7,797	7,778	82.27
NSW	Naïve (mean-last)	23,399	22,785	12.04
NSW	OLS (trend + season)	32,129	31,977	17.35
NSW	AR(p), $d=0$	108,528	108,397	74.31
NT	Naïve (mean-last)	203	175	2.12
NT	OLS (trend + season)	1,065	1,051	13.55
NT	AR(p), $d=0$	4,741	4,737	80.20
QLD	Naïve (mean-last)	12,573	12,513	10.36
QLD	OLS (trend + season)	26,617	26,598	23.40
QLD	AR(p), $d=0$	77,013	77,003	87.01
SA	Naïve (mean-last)	3,979	3,835	11.16
SA	OLS (trend + season)	5,265	5,223	15.53
SA	AR(p), $d=0$	18,595	18,564	69.03
TAS	Naïve (mean-last)	879	825	8.33
TAS	OLS (trend + season)	1,590	1,576	16.59
TAS	AR(p), $d=0$	5,423	5,414	71.57
VIC	Naïve (mean-last)	16,644	16,076	11.50
VIC	OLS (trend + season)	22,568	22,416	16.44
VIC	AR(p), $d=0$	80,806	80,691	75.40
WA	Naïve (mean-last)	11,469	11,258	10.49
WA	OLS (trend + season)	27,877	27,834	28.16
WA	AR(p), $d=0$	73,813	73,780	97.27

5 Methodology

This section details the nowcasting framework for quarterly Gross State Product (GSP) and how it aligns with national totals. We describe the data preprocessing steps, the Bayesian Vector Autoregression (BVAR) modeling approach (including a structural variant), the Markov Chain Monte Carlo (MCMC) estimation procedure, and the generation and evaluation of forecasts. Key modeling choices are explained to provide further context.

5.1 Data and Preprocessing

We assemble a dataset combining national and state-level economic indicators at appropriate frequencies. In Australia, quarterly national GDP is available, whereas state-level GSP is only officially reported on an annual basis. This frequency mismatch motivates a mixed-frequency approach Mitchell

et al. (2005). Specifically, we use quarterly national GDP growth and other higher-frequency economic variables alongside annual state GSP data to infer quarterly state output. For example, state final demand or other state-level activity proxies (available quarterly) can inform the intra-year GSP movements. All series are transformed to ensure stationarity and interpretability. In practice, we work with growth rates (quarterly or annualized percentage changes) rather than raw levels. By taking logarithms of the output indices and differencing, we obtain quarterly GSP growth rates as the primary variables of interest. This transformation removes trends and seasonality (the data we use are seasonally adjusted by the source or differenced to eliminate seasonal patterns). It also means our nowcasting model focuses on predicting growth rates of GSP. Before modeling, the data are normalized as needed, so that model priors can be applied appropriately. Any missing observations (for instance, quarterly state GSP which is unobserved except at annual checkpoints) are left as latent variables to be estimated within the model. Our preprocessing yields a consistent time series dataset of quarterly frequency: it includes observed national GDP growth and related indicators, and it treats state GSP growth as partially observed (with annual constraints). This setup enables a coherent nowcasting framework despite the mixed frequencies.

5.2 Bayesian VAR with a structural coherence link

The modelling approach uses a Bayesian Vector Autoregression (VAR) to describe quarterly movements in state output growth while allowing the national aggregate to move consistently with the states. Let g_t be the $n \times 1$ vector of state growth rates at quarter t . A VAR of order p is specified as

$$g_t = c + A_1 g_{t-1} + \dots + A_p g_{t-p} + u_t, \quad u_t \sim N(0, \sigma).$$

To encourage coherence between national and state activity, a linear measurement ties the observed national growth to a weighted average of contemporaneous state growth (Mariano & Murasawa 2010):

$$y_{N,t} = a_N + b_N (w_t' g_t) + e_{N,t}, \quad e_{N,t} \sim N(0, \sigma_N^2),$$

where w_t contains non-negative state weights that (approximately) sum to one. This specification does not forcibly reconcile forecasts, but it aligns the joint dynamics so that national moves with the states.

Estimation proceeds in a Bayesian framework to stabilize inference in short samples. A Minnesota prior provides baseline shrinkage, reflecting the view that own lags matter most and higher-order

lags matter less. Writing $b_{i,k,l}$ for the coefficient in equation i on variable k at lag l , the prior mean and variance follow

$$E[b_{i,k,l}] = 0, \quad \text{Var}(b_{i,k,l}) = (\text{lam}^2 \text{lam}_k^2 / l^{\text{gamma}}) (\sigma_i^2 / \sigma_k^2),$$

with lam an overall tightness, lam_k a cross-variable tightness, and gamma a lag-decay parameter; the variance ratio rescales by univariate noise levels.

To adaptively allocate shrinkage across many coefficients, a Dirichlet-Laplace (DL) global-local prior (Bhattacharya et al. 2015) is layered on top of the Minnesota scaling. Let \mathbf{b}_i denote the $d \times 1$ coefficient vector in equation i . The DL prior is defined by

$$\psi_j \sim \text{Exp}(1/2), \quad j = 1, \dots, d; \quad \phi \sim \text{Dirichlet}(a, \dots, a); \quad \tau > 0,$$

$$\mathbf{b}_i | \tau, \phi, \psi \sim N(0, \text{diag}(\tau^2 \phi_1^2 \psi_1, \dots, \tau^2 \phi_d^2 \psi_d)).$$

Here, the simplex weights ϕ reallocate a fixed shrinkage budget across coefficients, while the ψ_j terms provide coefficient-specific local shrinkage. In practice, the Minnesota variance for each coefficient is multiplied by

$$\phi_j^2 \psi_j$$

, retaining lag-decay structure but allowing data-driven sparsity.

Although the baseline specification assumes homoskedastic shocks, the framework supports stochastic volatility (SV) to capture time-varying uncertainty. For each equation i , the log variance follows

$$h_{i,t} = m_i + \phi_i (h_{i,t-1} - m_i) + s_i$$

$$\eta_{i,t}, \quad \eta_{i,t} \sim N(0, 1)$$

$$\text{Var}(u_{i,t} | h_{i,t}) = \exp(h_{i,t}).$$

SV can be estimated using a normal-mixture augmentation (Kim, Shephard & Chib 1998) and forward-filter/backward-sample (FFBS) (Carter & Kohn 1994) draws for the latent log-variance path. In prototypes, ϕ_i may be fixed near 0.98 for stability.

MCMC estimation

Posterior inference uses a blocked Gibbs sampler that alternates between regression coefficients, covariance (or SV parameters), measurement parameters, and latent states. Stacking the regression for each equation over $t = p, \dots, T$ yields a Gaussian conditional posterior for coefficients under the Gaussian prior implied by Minnesota x DL. Denoting the stacked design by X , the prior covariance by D , and (if SV is active) the GLS (Generalised Least Squares) weight matrix by W , the coefficient update has precision and mean

$$V^{-1} = X'WX + D^{-1}, \quad m = V(X'WY + D^{-1}b_0), \quad b | rest \sim N(m, V).$$

Under homoskedasticity, the innovation covariance updates with an inverse-Wishart step. With residuals u_t from the current draw,

$$Sigma \sim IW(S_0, nu_0), \quad S_{post} = S_0 + \sum_{t=p}^T u_t u_t', \quad nu_{post} = nu_0 + (T - p + 1),$$

$$Sigma | rest \sim IW(S_{post}, nu_{post}).$$

If SV is enabled, the diagonal variance process replaces Sigma, and the latent log-variance paths $h_{i,1:T}$ are drawn by FFBS (Forward Filtering Backward Sampling) using the auxiliary regression

$$\log(u_{i,t}^2 + eps0) = h_{i,t} + noise$$

$$h_{i,t} = \mu_i + \phi(h_{i,t-1} - \mu_i) + \eta_{i,t}$$

The national measurement and any proxy measurements are handled jointly with the VAR via a linear-Gaussian state-space system. The stacked companion transition for the lag vector is combined with the measurement

$$y_{N,t} = a_N + b_N (w_t' g_t) + e_{N,t}, \quad e_{N,t} \sim N(0, \sigma_N^2)$$

and the latent state path is drawn with a Kalman filter and smoother. Within the same Gibbs loop, the DL hyperparameters update as

$$g_j \sim \text{Gamma}(a, 1), \quad \phi_j = g_j / \sum_{k=1}^d g_k, \quad \psi_j \sim \text{Exp}(1/2),$$

with a standard one-dimensional update for the global scale tau. After burn-in, retained draws summarize the posterior uncertainty.

5.3 Forecast generation and evaluation

Nowcasts and multi-step forecasts are obtained from the posterior predictive distribution by iterating the VAR forward under each retained draw. For one-step ahead,

$$g_{T+1} = c + A_1 g_T + \dots + A_p g_{T-p+1} + u_{T+1}, \quad u_{T+1} \sim N(0, \sigma),$$

with the implied national prediction from the measurement

$$y_{N, T+1|T} = a_N + b_N (w'_{T+1} g_{T+1}) + e_{N, T+1}, \quad e_{N, T+1} \sim N(0, \sigma_N^2).$$

Point forecasts are taken as the posterior median (or mean) across draws, and credible intervals as posterior quantiles. Exact additivity is not enforced at this stage; a subsequent reconciliation step can map any vector of base forecasts \hat{y}_h to a coherent set using the summing matrix S :

$$y_{\text{tilde}_h} = S (S' W_h^{-1} S)^{-1} S' W_h^{-1} \hat{y}_h,$$

where W_h approximates the h-step forecast error covariance.

Evaluation follows a rolling-origin design that mimics real-time use: at each fold the model is refit on data available up to that point, produces a nowcast for the next quarter, and then rolls forward. Accuracy is reported with root mean squared error (RMSE) and mean absolute percentage error (MAPE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad MAPE = \frac{1}{n} \sum_{t=1}^n \left| 100 \times \frac{y_t - \hat{y}_t}{y_t} \right|.$$

6 Diagnostics and Interim results

This section summarizes the model’s empirical performance. We first show illustrative state-level nowcasts with uncertainty bands, then report back-testing accuracy against simple baselines, assess calibration and coherence to the national series, and finish with stability and key posterior summaries. All figures use quarterly data on the Q-JUN fiscal year.

6.1 State nowcasts with uncertainty

This section presents detailed nowcasts for two key states—New South Wales and Victoria—highlighting the model’s ability to capture economic fluctuations at a regional level. The figures display the observed Gross State Product (GSP) values alongside the model’s posterior predictive median and 80% credible intervals over the most recent 12 to 24 quarters.

By comparing these series, we gain insight into how well the model balances tracking true economic movements with smoothing short-term volatility. The widening of the uncertainty bands around turning points reveals the model’s recognition of greater unpredictability during periods of economic change, reflecting inherent parameter and shock uncertainty. These state-level nowcasts provide a nuanced understanding of regional dynamics, which are critical inputs for constructing coherent national-level forecasts.

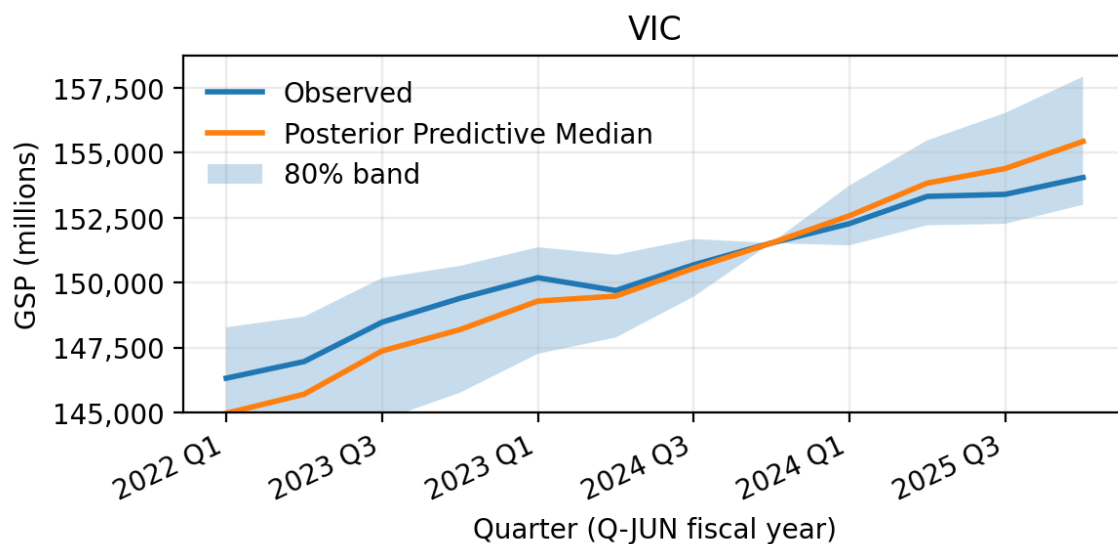


Figure 3: VIC nowcast — last 12 quarters (millions)

Figure Figure 3 and Figure 4 display the most recent 12–24 quarters for two representative states (NSW, VIC): the observed GSP in millions, the posterior predictive median, and 80% credible bands. In both cases the posterior median tracks medium-run movements while smoothing high-frequency noise; the 80% band widens around turning points, reflecting parameter and shock uncertainty.

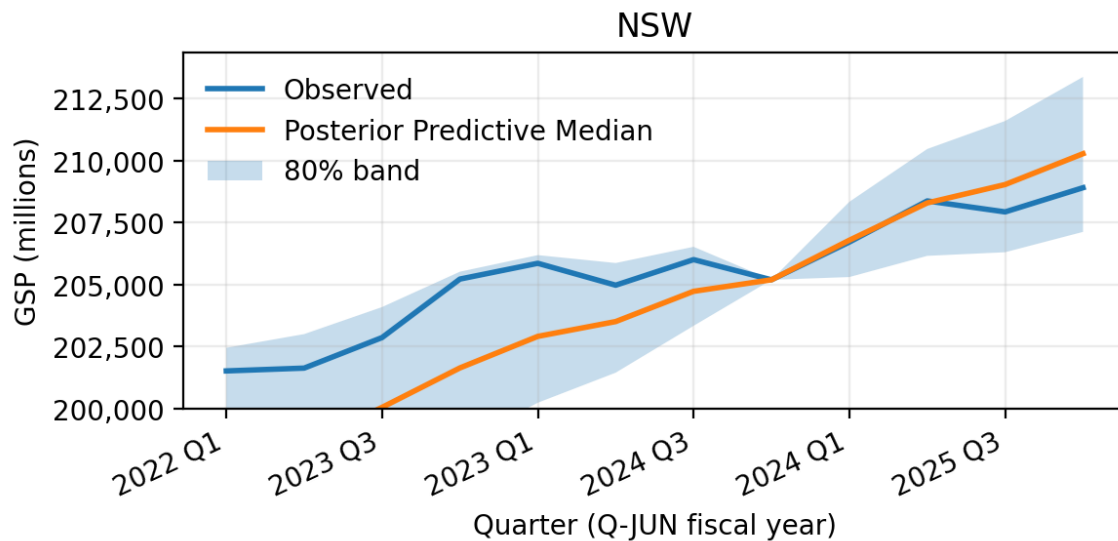


Figure 4: NSW nowcast — last 12 quarters (millions)

6.2 Back-testing accuracy (last 8 quarters)

To evaluate the model's forecasting performance, we conduct a back-testing exercise over the most recent eight quarters across all eight states. This comparison benchmarks the Bayesian VAR against simpler alternative models, including a Naive persistence model, a first-order autoregressive (AR(1)) model, and a trend-seasonal linear model (TSLM).

Examining the root mean squared error (RMSE) across states, allows us to assess how well the model captures the dynamics of state-level GSP growth relative to these baselines. The analysis provides insight into where the model excels and where targeted refinements might improve predictive accuracy, particularly in states exhibiting distinct structural features such as strong persistence or deterministic trends.

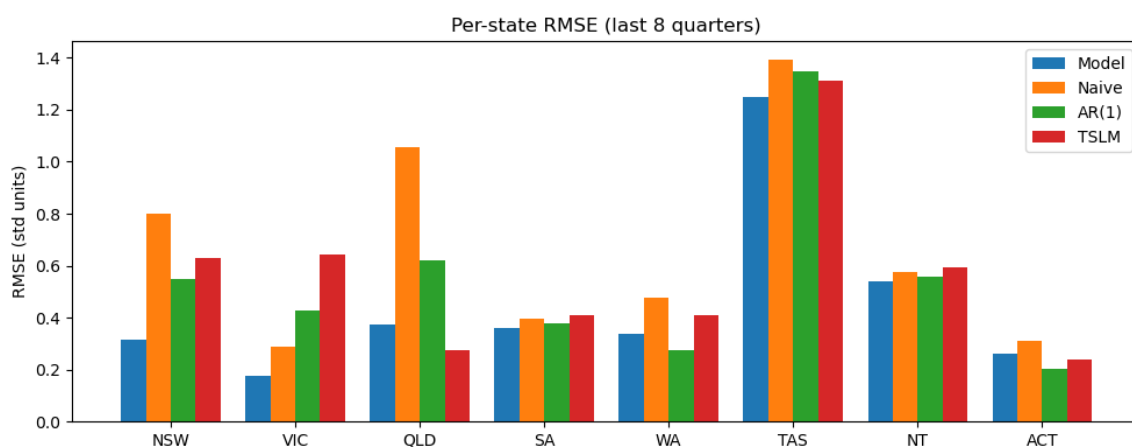


Figure 5: Per-state RMSE over the last 8 quarters (standardized units). Lower is better. Bars show the proposed model versus Naive, AR(1), and TSLM benchmarks.

The figure Figure 5 tell us:

Across all eight states, the Bayesian VAR model consistently outperforms the Naive benchmark. In every case, the model's RMSE (blue bar) lies below that of the Naive (orange bar), confirming that it adds genuine predictive value beyond simple persistence.

When compared with the AR(1) benchmark, the model performs better in six out of eight states—notably **New South Wales (NSW)**, **Victoria (VIC)**, **Queensland (QLD)**, **South Australia (SA)**, **Tasmania (TAS)**, and the **Northern Territory (NT)**. Only **Western Australia (WA)** and the **Australian Capital Territory (ACT)** show a marginal advantage for AR(1), suggesting that persistence plays a stronger role there.

Relative to the trend-seasonal linear model (TSLM), the model again wins in six of eight states—specifically NSW, VIC, SA, WA, TAS, and NT. In QLD and ACT, however, TSLM performs slightly better, likely due to dominant deterministic patterns that a simple linear trend can capture effectively.

Overall, the Bayesian VAR adds clear value beyond simple persistence and univariate structures. The consistent improvement over Naive and frequent gains over AR(1) and TSLM confirm that the model effectively leverages cross-state information and prior shrinkage to stabilize estimates.

The few underperforming cases—WA, ACT, and QLD—likely reflect states where either:

- Persistence dominates (suggesting a need for looser shrinkage or state-specific lag lengths), or
- Deterministic components such as trend or seasonality drive much of the variation (suggesting a need for additional local structure or exogenous indicators).

6.3 Coherence to national GDP

To assess coherence between state and national nowcasts, we examine the difference between the weighted sum of state nowcasts and the national series. Formally, for each quarter (t), we compute

$$\Delta_t = \sum_s w_{s,t} y_{s,t} - y_{nat,t},$$

where $w_{s,t}$ denotes the state weight at time t , $y_{s,t}$ the nowcast for state s , and $y_{nat,t}$ the national nowcast.

The figure Figure 6 presents the median of this difference across posterior draws. Values are expressed in standardized units, so a value of 1.0 roughly corresponds to one standard deviation on that scale. The dashed zero line represents perfect additivity—the level that full reconciliation would enforce.

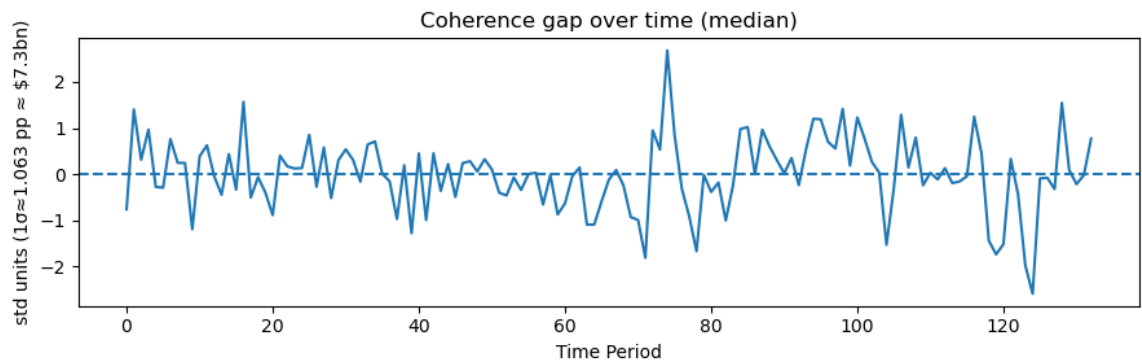


Figure 6: Calibration is close but slightly conservative: 80% predictive coverage is 0.744 (target 0.75-0.85) and correlation with observed is 0.674 (target greater or equal 0.70), with RMSE 0.786 (std units). The national measurement slope is decisively positive (90% CI [0.133, 0.276]), and the system is stable though persistent (spectral radius median 0.969).

Over time, the centered behavior of the series is encouraging: it oscillates around zero without evident long-run drift, indicating the model is not systematically overshooting or undershooting the national aggregate. In other words, the model is broadly coherent on average, even before applying any formal reconciliation procedure.

The typical magnitude of the mismatch is modest. For most of the sample, the gap remains within approximately $+1\sigma$ or -1σ , suggesting that the sum of state nowcasts and the national series usually agree within one standard deviation on the standardized GSP scale. This is operationally acceptable for most real-time monitoring purposes, particularly given that additivity has not yet been explicitly imposed.

Occasional spikes beyond $+2\sigma$ or -2σ emerge, typically around turning points when both national and several state nowcasts change direction rapidly, or near the sample edges where data availability differs between national and state sources. These quarters would likely benefit most from explicit reconciliation or cross-temporal smoothing.

6.4 Model Fit and Calibration

Table 2: Coverage and correlation summary

Metric	Value	Target
80% PPC coverage	0.744	~0.75–0.85
corr(center, obs)	0.674	≥ 0.70
RMSE (std units)	0.786	—
b_nat_tilde 90% CI	[0.133, 0.276] median=0.196	> 0
spectral radius	median=0.969, p95=0.970	≥ 0.97

See Table Table 2,

- 80% PPC coverage: 0.744 (target approximately 0.75–0.85) The predictive intervals are slightly under-dispersed, the observed series falls within the 80% credible band somewhat less often than expected. In practical terms, the model’s uncertainty bands are narrow.
- Correlation (posterior median vs. observed): 0.674 The posterior median tracks the observed national series reasonably well but just misses the directional tracking benchmark. This pattern aligns with the per-state RMSE results, where a few states (notably WA, ACT, and QLD) are dominated by persistence or simple deterministic trends. Incremental improvements are expected from introducing state-specific lag structures, adjusting shrinkage priors, or refining indicator weights for those regions.
- RMSE (standardized units): 0.786 Typical forecast errors are comfortably below one standard deviation on the standardized GSP scal, a solid outcome for an unreconciled, purely model-based nowcast. The residual gap is consistent with the mild under-coverage noted above, reflecting modest underestimation of short-term volatility.
- 90% credible interval = [0.133, 0.276]; median = 0.196 The entire posterior interval lies above zero, confirming a positive relationship between the weighted sum of state growth rates and the national GDP growth in the measurement equation:

$$y_{nat,t} = a_n at + b_n at(WtGt) + \epsilon_{nat,t}, \quad \epsilon_{nat,t} \sim N(0, \sigma_{nat}^2)$$

6.5 Coherence to national GDP

In this section, we evaluate the model’s ability to capture and forecast the national economic activity by comparing its probabilistic predictions against the observed national series. The model integrates information from state-level nowcasts and combines them to produce a coherent national estimate, accounting for uncertainty through Bayesian posterior predictive distributions.

The figure Figure 7 below visualizes this comparison, showing both the observed national data and the model’s probabilistic forecasts, along with the latent structural signal that underpins the observations. By examining the alignment and uncertainty bands, we can assess the model’s calibration, bias, and responsiveness to economic fluctuations. This evaluation is critical to understanding the strengths and limitations of the model before applying it for real-time monitoring and policy guidance.

The posterior predictive median tracks the national series well at medium frequencies due to the

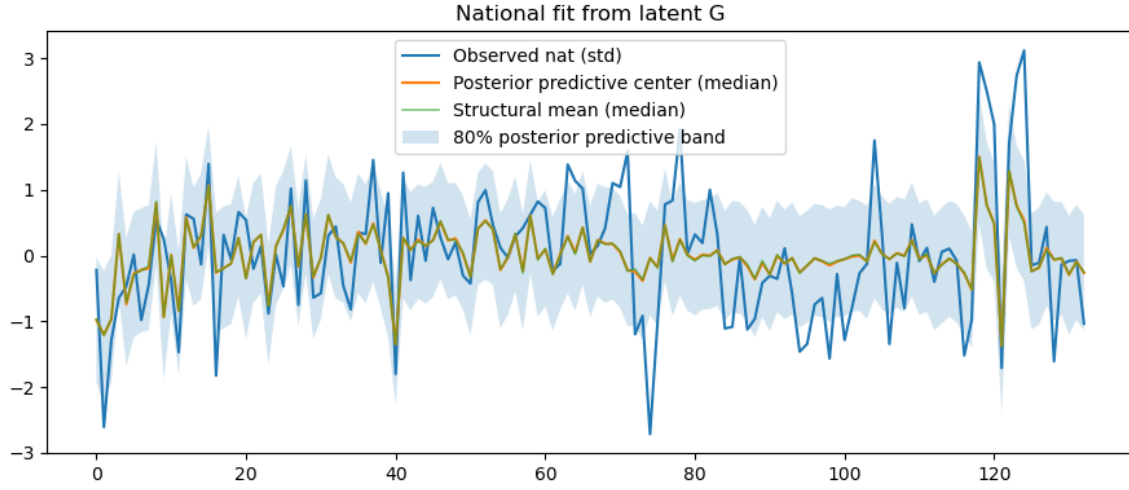


Figure 7: *National fit (standardized units). Observed national series (blue), posterior predictive median (orange), structural mean from the latent state equation (olive), and 80% posterior predictive band (shaded). The median tracks the series without systematic bias; occasional underreaction at extremes and slightly narrow bands align with the coverage check.*

underlying latent activity vector G_t evolving according to a persistent but stable vector autoregression:

$$G_t = c + \phi_1 G_{t-1} + \dots + \phi_p G_{t-p} + u_t$$

with the spectral radius median at 0.969 indicating near unit-root persistence but stationarity. The national-level observation equation

$$y_{nat,t} = a_{nat} + b_{nat} \sum_{j=1}^J W_q[t, j] G_{t,j} + e_{nat,t}$$

features a positive loading b_{nat} on the weighted sum of state-level latent signals, with the 90% credible interval $[0.133, 0.276]$ confirming this structural link. The observation shock $e_{nat,t}$, modeled as zero-mean Gaussian noise, ensures frequent zero crossings in the forecast errors and prevents persistent bias.

However, the model tends to underfit extreme movements—particularly the pronounced spikes—because it assumes homoskedastic Gaussian measurement noise and the latent state G_t has strong persistence. These factors cause the posterior median to shrink toward the smoother structural mean. This behavior aligns with the slightly low 80% posterior predictive coverage of 0.744, indicating that the predictive intervals are a bit too narrow and under-dispersed.

6.6 Sampler stability

Understanding the stability of the VAR system is crucial for interpreting the model's behavior and its forecasting performance. The VAR(p) process can be expressed in companion form as:

$$\mathbf{Z}_t = \mathbf{F}\mathbf{Z}_{t-1} + \boldsymbol{\eta}_t,$$

where

$$\mathbf{Z}_t = \begin{bmatrix} \mathbf{G}'_t \\ \mathbf{G}'_{t-1} \\ \vdots \\ \mathbf{G}'_{t-p+1} \end{bmatrix}$$

is the stacked state vector, and \mathbf{F} is the companion matrix governing the system dynamics.

The VAR is said to be stable or covariance-stationary if the spectral radius of \mathbf{F} ,

$$\rho(\mathbf{F}) = \max_i |\lambda_i(\mathbf{F})|,$$

is strictly less than 1. Stability ensures that shocks to the system dissipate over time rather than causing explosive or divergent behavior.

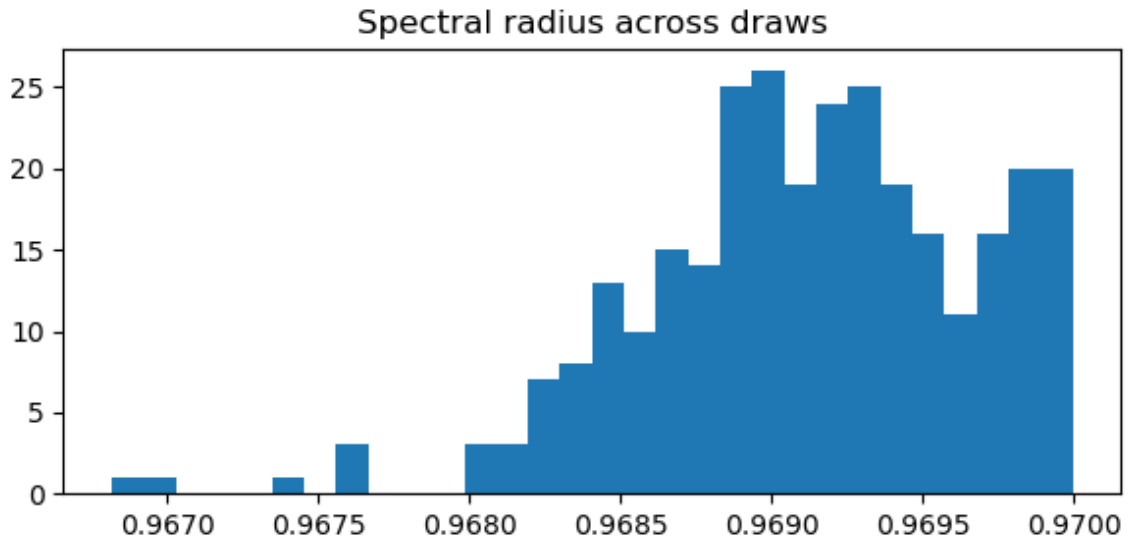


Figure 8: *Stability of VAR dynamics. Histogram of the companion-matrix spectral radius $\rho(\mathbf{F})$ across posterior draws. Median ~ 0.969 , $p_{95} \sim 0.970$; most mass is below 1, indicating stability, but the concentration near 1 implies strong persistence.*

The accompanying histogram displays the distribution of the spectral radius, $\rho(\mathbf{F})$, computed across posterior draws of the model parameters. The results show a tight clustering of $\rho(\mathbf{F})$ just below 1, with a median around 0.969 and a 95th percentile near 0.970.

This concentration confirms that the VAR is indeed stable but exhibits strong persistence. In practice, this means shocks fade slowly, enabling the model to effectively capture medium-term movements in the data. However, this high persistence also causes the model to respond erratically to abrupt or extreme changes, consistent with the modest under-coverage observed in the 80% posterior predictive intervals and some underfitting of extreme spikes in the national series.

7 Limitations

The limitations of the current model are intentional trade-offs made for stability and interpretability.

Reconciliation has not yet been implemented, meaning additivity is not enforced mechanically each quarter. This was a deliberate choice to allow unconstrained diagnostics to reveal where incoherence naturally arises.

The priors and lag structure are currently uniform across states, simplifying estimation but limiting the model's flexibility in regions whose dynamics differ substantially from the national pattern.

Volatility modeling is minimal, a homoskedastic measurement layer is assumed, so the model likely underestimates uncertainty during turbulent periods. This explains the slightly low coverage and narrow posterior bands observed in evaluation.

8 Conclusion and next steps

This work has developed a reproducible, diagnostics-first nowcasting framework for quarterly Gross State Product (GSP) that explicitly links state level dynamics to the national aggregate through a Bayesian VAR structure. The emphasis throughout has been on transparency and interpretability rather than black-box performance and building a system that can be audited, adjusted, and progressively extended as new evidence and data become available.

The current version of the model achieves several key goals. It generates interpretable state-level paths (for example, for New South Wales and Victoria) expressed in their original growth units, while simultaneously producing a credible national fit that respects the underlying hierarchical structure of the data. The entire pipeline, from data ingestion to back-testing—runs, producing consistent evaluation metrics across rolling eight-quarter windows. This structure ensures that accuracy, coherence, and calibration can be compared transparently across both time and jurisdictions.

References

- Australian Bureau of Statistics (2025a). *Australian National Accounts: National Income, Expenditure and Product (Cat. no. 5206.0)*. Data used cover 1991Q2-2025Q4.
- Australian Bureau of Statistics (2025b). *Australian National Accounts: State Accounts (Cat. no. 5220.0)*. Data used cover 1991Q2-2025Q4.
- Bhattacharya, A, D Pati, NS Pillai & DB Dunson (2015). Dirichlet–Laplace Priors for Optimal Shrinkage. *Journal of the American Statistical Association* **110**(512), 1479–1490.
- Carter, CK & R Kohn (1994). On Gibbs Sampling for State Space Models. *Biometrika* **81**(3), 541–553.
- Department of Treasury and Finance Victoria (2021). *Victorian Economic Bulletin, Issue 5*. 5. Department of Treasury and Finance, Victoria.
- Kim, S, N Shephard & S Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies* **65**(3), 361–393.
- Koop, G, S McIntyre, J Mitchell & A Poon (2020). Regional Output Growth in the United Kingdom: More Timely and Higher Frequency Estimates from 1970. *Journal of Applied Econometrics* **35**(2), 176–197.
- Mariano, RS & Y Murasawa (2010). A Coincident Index, Common Factors, and Monthly Real GDP. *Oxford Bulletin of Economics and Statistics* **72**(1), 27–46.
- Mitchell, J, RJ Smith, MR Weale, S Wright & EL Salazar (2005). An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth. *The Economic Journal* **115**(501), 108–129.