Predicting US GDP Growth with a Bayesian Autoregression

Purpose and permitted software

Develop a Bayesian autoregressive model with a conjugate prior by deriving its posterior, programming a posterior-draw routine, and generating both point predictions and predictive distributions for quarterly US real GDP growth. Implement the work in Python, R, Julia, or MATLAB.

Disclose any use of generative AI by identifying the specific parts in which it assisted and explaining its purpose in each part.

Data preparation and initial summaries

Obtain the real GDP series GDPC1 from FRED-QD. Retain observations from the earliest date available through 2019Q4. Construct annualized quarter-to-quarter growth using

$$y_t = 400\big(\log GDP_t - \log GDP_{t-1}\big), \qquad t = 1, \dots, T.$$

Display the growth series in a time plot. Provide its sample mean and standard deviation, together with sample autocorrelations at lags one through four.

Regression specification

Include an intercept in an autoregression of order p,

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2), \tag{1}$$

for $t = p+1, \dots, T$. Define $\boldsymbol{\beta} = (c, \phi_1, \dots, \phi_p)'$ and $x_t = (1, y_{t-1}, \dots, y_{t-p})'$, giving $y_t = x_t'\boldsymbol{\beta} + \varepsilon_t$. In stacked form, write $y = X\boldsymbol{\beta} + \varepsilon$, with y of dimension $n \times 1$, $n = T - p$, and X of dimension $n \times (p+1)$. Treat the initial p observations as given in every calculation, so inference uses the conditional likelihood.

Conjugate prior and numerical settings

Adopt the natural conjugate Normal–Inverse-Gamma specification

$$\sigma^2 \sim IG(a_0, b_0), \qquad \boldsymbol{\beta} \mid \sigma^2 \sim N(\mathbf{b}_0, \sigma^2 V_0),$$

where $p(\sigma^2) \propto (\sigma^2)^{-(a_0+1)} \exp(-b_0/\sigma^2)$. Use the following simple Minnesota-style settings.

- Set $\mathbf{b}_0 = (0, 0, \dots, 0)'$, expressing a prior centred on zero-mean white-noise growth. As an alternative, place the intercept's prior mean at the sample mean of y_t and the first lag's prior mean at 0.5. Explain your preferred choice.
- Choose $V_0 = \text{diag}(100, \lambda^2/1^2, \lambda^2/2^2, \dots, \lambda^2/p^2)$ with $\lambda = 0.5$. This allows considerable freedom for the intercept while shrinking lag coefficients toward zero, with prior variances declining as $1/j^2$ for lag j.
- Take $a_0 = 3$ and $b_0 = 2\hat{s}^2$, where $\hat{s}^2$ denotes the sample variance of y_t. Consequently, $E[\sigma^2] = b_0/(a_0 - 1) = \hat{s}^2$. The prior thus places innovation variance on the scale of unconditional growth variance, with limited certainty.

TriadEssay

Analytical work and posterior implementation

(a) Construct the likelihood. Express $p(y \mid \boldsymbol{\beta}, \sigma^2, X)$ explicitly as a function of the regression coefficients and innovation variance, conditional on the initial observations.

(b) Obtain the posterior distributions. Derive $p(\boldsymbol{\beta}, \sigma^2 \mid y)$ and identify its distributional family. State the posterior of $\boldsymbol{\beta}$ conditional on σ^2 and the marginal posterior of σ^2. Express all distribution parameters using the data and prior hyperparameters. Explain how the posterior mean of $\boldsymbol{\beta}$ relates to both the prior mean and the ordinary least squares (OLS) estimate.

(c) Remove the variance parameter. Integrate σ^2 out of the joint posterior and describe the resulting marginal distribution of $\boldsymbol{\beta}$.

(d) Examine extreme prior choices. Briefly explain the behaviour of the posterior for $\boldsymbol{\beta}$ when $V_0^{-1} \to 0$, corresponding to a flat prior, and when $V_0 \to 0$, corresponding to a dogmatic prior.

(e) Program and summarize posterior sampling. Using the expressions derived in part (b), implement a function whose inputs are the observations $y_{1:T}$, lag length p, and hyperparameters $(\mathbf{b}_0, V_0, a_0, b_0)$. The function must produce

$$\{\boldsymbol{\beta}^{(m)}, \sigma^{2(m)}\}_{m=1}^{M}, \qquad M = 10{,}000,$$

as draws from the posterior.

Fit an AR(4) specification to the entire sample. For every component of $\boldsymbol{\beta}$ and for σ^2, report the posterior mean, posterior standard deviation, and a 90% credible interval. Present the corresponding OLS estimates alongside these summaries. Also show the posterior density of ϕ_1 and its prior density on the same plot.

Point prediction

(f) Choose an estimation endpoint T_0, for example 2007Q4. Predict y_{T_0+h} at horizons $h = 1, \ldots, 8$.

(i) At the one-quarter horizon, derive the posterior predictive expectation $E[y_{T_0+1} \mid y]$ and variance $\mathrm{Var}[y_{T_0+1} \mid y]$. Distinguish the contribution from uncertain parameters from the contribution of the future innovation.

(ii) Beyond one quarter, a closed-form solution is unavailable because future outcomes enter the regressors nonlinearly in $\boldsymbol{\beta}$. Use simulation to obtain the point predictions. For each posterior draw m, apply equation (1) recursively for h periods and draw a new innovation at every step,

$$\varepsilon_{T_0+j}^{(m)} \sim N\big(0, \sigma^{2(m)}\big).$$

Average the simulated values of $y_{T_0+h}^{(m)}$ across draws to obtain point predictions. Tabulate the forecasts and compare them with recursively generated OLS forecasts.

TriadEssay

Predictive distributions and out-of-sample checks

(g) The simulated trajectories from part (f)(ii) sample the posterior predictive distribution $p(y_{T_0+h} \mid y)$. Use them to complete the following analyses.

(i) **Visualize forecast uncertainty.** Produce a fan chart covering horizons one to eight. Include the predictive median and 50%, 68%, and 90% predictive bands, and overlay the actual observations.

(ii) **Run recursive forecasting.** Move the forecast origin T_0 through every quarter from 1999Q4 to one quarter before the sample ends. At each origin, refit the model using observations available through T_0 and obtain the next quarter's predictive distribution. Retain both the point prediction and predictive density from each iteration.

(iii) **Assess forecast performance.** Evaluate the one-step-ahead predictions with the root mean squared forecast error (RMSFE) and the mean log predictive score,

$$\frac{1}{N}\sum_{T_0} \log p(y^{\text{obs}}_{T_0+1} \mid y_{1:T_0}),$$

where N is the number of forecast origins. Estimate the predictive density either by applying a kernel density estimator to the simulated draws or by averaging the Normal densities

$$N(y^{\text{obs}}_{T_0+1};\, x'_{T_0+1}\beta^{(m)}, \sigma^{2(m)})$$

over m. Compare performance with two benchmarks. The no-change benchmark predicts y_{T_0+1} using y_{T_0} and assigns the density $N(y_{T_0}, \hat{s}^2)$. The other benchmark uses the sample mean as its forecast.

(iv) **Check distributional calibration.** Calculate the probability integral transform at each origin,

$$PIT_{T_0} = \Pr(y_{T_0+1} \le y^{\text{obs}}_{T_0+1} \mid y_{1:T_0}),$$

and display the values in a histogram. Describe the histogram expected under well-calibrated predictive densities and assess whether your results exhibit that pattern.

TriadEssay