Vector Autoregressive Models

From AR(p) to VAR(p) with Bayesian Inference

Hedibert F. Lopes

Insper Institute of Education and Research

May 7th 2026

From q AR(p) Models to VAR(p)

Consider q separate AR(p) models:

$$
\begin{aligned}
Y_{t1} &= b_{11} Y_{t-1,1} + b_{12} Y_{t-2,1} + \cdots + b_{1p} Y_{t-p,1} + u_{t1}, \quad u_{t1} \overset{\text{iid}}{\sim} N(0, \sigma^2_{11}) \\
Y_{t2} &= b_{21} Y_{t-1,2} + b_{22} Y_{t-2,2} + \cdots + b_{2p} Y_{t-p,2} + u_{t2}, \quad u_{t2} \overset{\text{iid}}{\sim} N(0, \sigma^2_{22}) \\
&\vdots \\
Y_{tq} &= b_{q1} Y_{t-1,q} + b_{q2} Y_{t-2,q} + \cdots + b_{qp} Y_{t-p,q} + u_{tq}, \quad u_{tq} \overset{\text{iid}}{\sim} N(0, \sigma^2_{qq})
\end{aligned}
$$

Key limitation

q independent AR(p) models $\neq$ VAR(p)

Example: $q = 4$: $(Y_{t1}, \ldots, Y_{t4})$ = (interest rate, inflation, unemployment, industrial production)

What Makes a VAR(p)?

Two key features beyond independent AR(p) systems:

(1) Cross-equation lags: Include lags of *all* q time series in *each* equation.

Overall: q equations $\times$ q regressors $\times$ p lags $= q^2 p$ coefficients.

(2) Joint error distribution:

$$\boldsymbol{u}_t = \begin{pmatrix} u_{t1} \\ u_{t2} \\ \vdots \\ u_{tq} \end{pmatrix} \sim N_q(\mathbf{0}, \Sigma)$$

where Σ is a $(q \times q)$ covariance matrix (not diagonal in general).

Parameter Count

Total number of parameters in a VAR(p):

$$\underbrace{pq^2}_{\text{VAR coefficients}} + \underbrace{\frac{q(q+1)}{2}}_{\Sigma \text{ elements}}$$

Example: $p = 36$, $q = 4$

- Each equation has $q \times p = 4 \times 36 = 144$ regressors
- VAR coefficients: $4 \times 36 \times 4 = 576$
- Σ elements: $\frac{4 \times 5}{2} = 10$
- **Total: 586 parameters**

The Covariance Matrix Σ (for $q = 4$)

$$\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}^2 \end{pmatrix}$$

This symmetric positive-definite matrix has

$$\frac{q(q+1)}{2} \text{ free elements.} \qquad q = 4 \;\Rightarrow\; \frac{4 \times 5}{2} = 10$$

VAR(p) with Gaussian Errors — Matrix Form

Compact representation:

$$\underbrace{\boldsymbol{Y}_t}_{q\times 1} = \underbrace{B_1}_{q\times q}\underbrace{\boldsymbol{Y}_{t-1}}_{q\times 1} + \underbrace{B_2}_{q\times q}\underbrace{\boldsymbol{Y}_{t-2}}_{q\times 1} + \cdots + \underbrace{B_p}_{q\times q}\underbrace{\boldsymbol{Y}_{t-p}}_{q\times 1} + \boldsymbol{u}_t \tag{1}$$

where $\boldsymbol{Y}_t = (Y_{t1}, Y_{t2}, \ldots, Y_{tq})^\top$.

Each B_i, $i = 1, \ldots, p$, has q^2 coefficients $\Rightarrow$ pq^2 total VAR coefficients.

$$\boldsymbol{u}_t \overset{\text{iid}}{\sim} N_q(\mathbf{0}, \Sigma_t) \tag{2}$$

Extension

Allowing Σ_t to vary over time leads to **TVP-BVAR-SV** models (Time-Varying Parameters, Bayesian VAR, Stochastic Volatility).

Structural VAR (SVAR)

Starting from $\boldsymbol{Y}_t \sim N(\mathbf{0}, \Sigma)$, define the **Cholesky-type** decomposition:

$$A\Sigma A^\top = I \quad \Longrightarrow \quad A\boldsymbol{Y}_t \sim N(\mathbf{0}, I)$$

Reduced-form VAR

Posterior: $P(B_1, \ldots, B_p, \Sigma \mid \text{data})$

Estimated via **MCMC**

Structural VAR (SVAR)

Posterior: $P(A_0, B_1, \ldots, B_p \mid \text{data})$

Estimated via **MCMC**

Key distinction

The SVAR identifies structural shocks through the matrix A_0, imposing economic restrictions on the contemporaneous relationships among variables.

Frequentist Estimation and Forecasting

For $t = 1, \dots, T$, OLS yields $\widehat{B}_1, \widehat{B}_2, \dots, \widehat{B}_p,\ \widehat{\Sigma}$.

One-step-ahead predictive distribution:

$$p(Y_{T+1} \mid Y_1, \dots, Y_T) =?$$

Point forecast (plug-in):

$$\widehat{Y}_{T+1} = \widehat{B}_1 Y_T + \widehat{B}_2 Y_{T-1} + \cdots + \widehat{B}_p Y_{T-p+1} + u_{T+1}$$

$$\widehat{Y}_{T|1} = \widehat{E}(Y_{T+1} \mid Y_1, \dots, Y_T), \qquad \widehat{V}(Y_{T+1} \mid Y_1, \dots, Y_T) = \widehat{\Sigma}$$

h-**steps-ahead forecast:** $(\widehat{Y}_{T+1}, \widehat{Y}_{T+2}, \dots, \widehat{Y}_{T+h})$

Recall AR(2): $\theta = (\phi_1, \phi_2, \sigma^2)$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

Two-step-ahead conditional mean:

$$E(Y_t \mid Y_{t-3}, Y_{t-4}) = (\phi_1^2 + \phi_2)\, Y_{t-3} + (\phi_1\phi_2 + \phi_2^2)\, Y_{t-4}$$

Two-step-ahead conditional variance:

$$\begin{aligned} V(Y_t \mid Y_{t-3}, Y_{t-4}) &= (\phi_1^2 + \phi_2)\, \sigma^2 + \sigma^2 \\ &= (\phi_1^2 + \phi_2)^2 + (\phi_1^2 + \phi_2)\, \phi_1^2 + \phi_1^2 + \phi_2^2 \end{aligned}$$

Derived by iterating $Y_t = \phi_1(\phi_1 Y_{t-2} + \phi_2 Y_{t-3} + \varepsilon_{t-1}) + \phi_2 Y_{t-2} + \varepsilon_t$*, etc.*

Summary

- A VAR(p) model couples q time series by including cross-equation lags **and** a full covariance matrix Σ.
- Total parameters: $pq^2 + \frac{q(q+1)}{2}$ — grows quickly with q and p.
- Matrix form: $\boldsymbol{Y}_t = \sum_{i=1}^{p} B_i \boldsymbol{Y}_{t-i} + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim N_q(\mathbf{0}, \Sigma)$.
- Frequentist OLS estimation gives plug-in forecasts; Bayesian MCMC gives full posteriors.
- SVAR adds structural identification via A_0 (the contemporaneous matrix).
- Extensions: TVP-BVAR-SV allows time-varying coefficients and stochastic volatility.