Methods of Macroeconomic Forecasting

Methods Boot Camp – Day 2

KOF ETH Zurich

October 3, 2025, Zurich

Recap of Day 1

- AR models: persistence & mean reversion; h-step forecasts converge to long-run mean; uncertainty widens with horizon.
- VARs: multivariate dynamics; forecasting and IRFs; identification noted but not our focus.
- Forecast evaluation & combination: RMSE/MAE, density forecasts, pooling across models.
- Bayesian basics: posterior $\propto$ prior $\times$ likelihood; priors encode beliefs/regularization.
- MCMC (Gibbs): practical tool to sample posteriors in non-conjugate cases.

Quick Check

- Q1: In an AR(1) with $\phi = 0.9$, does the h-step forecast converge fast or slowly to the mean? Why?
- Q2: Why do we sometimes prefer forecast combination to selecting one “best” model?
- Q3: What does the Gibbs sampler alternate between (conceptually)?

Today's Plan (Day 2)

- Bayesian VARs (BVARs): shrinkage (Minnesota prior), predictive densities, conditional forecasting.
- Structural Equation Models (SEMs): theory-driven systems, identification, Bayesian estimation.
- State-Space Models: measurement/state equations, Kalman filter intuition, dynamic factors.

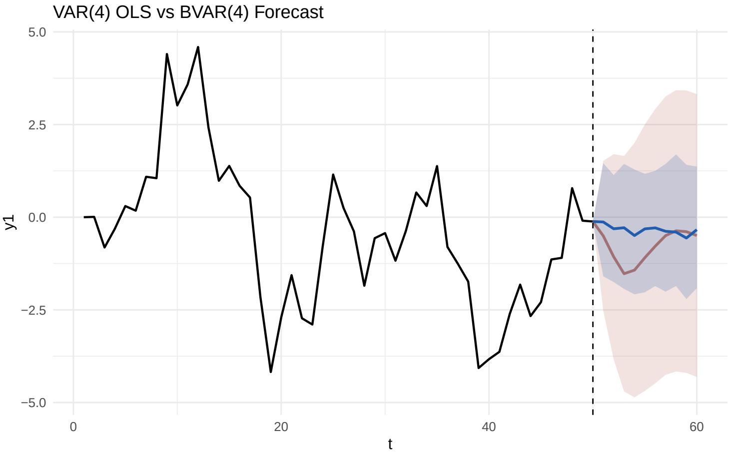
Bayesian VAR

Why Bayesian VARs?

- VARs are heavily parameterized (many coefficients, limited data).
  - Example: $n = 4$ variables, $p = 4$ lags $\Rightarrow$ $4 \times (4 \cdot 4 + 1) = 68$ coefficients, plus covariance parameters.
- Risk: overfitting $\Rightarrow$ unstable coefficients/IRFs, wide forecast bands.
- Bayesian approach introduces prior information to stabilize estimates.
- Key mechanism: shrinkage toward economically sensible values.

Too Many Parameters? Why Bayesian VARs Help

- OLS VAR(4): unstable, wide confidence bands (red).
- BVAR (shrinkage): smoother, tighter density forecasts (blue).



Why BVARs for Forecasting?

- Shrinkage prevents large VARs from becoming overloaded with too many parameters.
- Improve out-of-sample accuracy (less variance, better calibration).
- Naturally produce density forecasts (fan charts, risk analysis).
- Flexible: hyperparameters/hierarchies adapt to datasets.

Minnesota Prior: Intuition

- Macro series are often persistent; random walk is a strong benchmark.
- Prior beliefs:
  - Own first lag is important (close to 1).
  - Higher lags are less important.
  - Cross-variable effects are smaller than own-lag effects.
- Encoded via prior means and variances that shrink coefficients.

Minnesota Prior: Structure

- Prior means (typical choice):

$$\text{Mean}(\phi_{ii}(1)) = 1, \quad \text{Mean}(\phi_{ij}(\ell)) = 0 \quad (i \neq j \text{ or } \ell > 1)$$

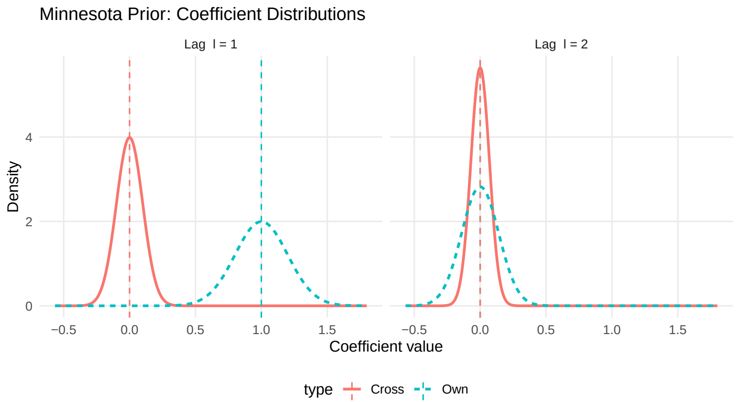
- Prior variances (schematic):

$$\mathrm{Var}\big(\phi_{ij}(\ell)\big) = \lambda_1^2 \, \frac{\sigma_i^2}{\sigma_j^2} \, \frac{1}{\ell^{\lambda_3}} \times \begin{cases} 1, & i = j \\ \lambda_2^2, & i \neq j \end{cases}$$

- Hyperparameters $\lambda_1, \lambda_2, \lambda_3$ control:
  - overall shrinkage, cross-variable shrinkage, lag decay

Minnesota Prior: Visual Example

- Own lag coefficients: prior centered near 1 with tight variance.
- Cross-variable lags: prior centered at 0 with looser variance.
- Higher lags: shrink more strongly.



Minnesota Prior: Role of Hyperparameters

- λ_1: overall tightness
  - Small λ_1: strong shrinkage $\Rightarrow$ forecasts stable.
  - Large λ_1: weak shrinkage $\Rightarrow$ approaches OLS.
- λ_2: cross-variable shrinkage
  - Large λ_2: cross-effects nearly unrestricted.
  - Small λ_2: cross-effects almost shut down.
- λ_3: lag decay
  - Higher order lags shrink more quickly as λ_3 increases.

Bayesian Setup for VAR

- VAR(p):

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}).$$

- Dimensions:
  - $\mathbf{y}_t$: $(n \times 1)$ vector, $\mathbf{c}$: $(n \times 1)$ vector, $\mathbf{\Phi}$: $(n \times n)$ matrix, $\boldsymbol{\varepsilon}$: $(n \times 1)$ vector.
- Stacked regression form (over $t = 1, \ldots, T$):

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \qquad \operatorname{vec}(\mathbf{E}) \sim \mathcal{N}(\mathbf{0},\, \mathbf{\Sigma} \otimes \mathbf{I}_T).$$

- Dimensions:
  - $\mathbf{Y}$: $T \times n$ matrix of dependent variables
  - $\mathbf{X}$: $T \times m$ regressor matrix (lags, constants, exog.)
  - $\mathbf{B}$: $m \times n$ coefficient matrix, $\boldsymbol{\beta} = \operatorname{vec}(\mathbf{B})$ $(mn \times 1)$
  - $\mathbf{\Sigma}$: $n \times n$ covariance matrix of innovations
- Bayesian inference:

$$p(\theta \mid \mathbf{Y}, \mathbf{X}) \;\propto\; p(\theta)\, p(\mathbf{Y} \mid \theta), \text{ with } \theta = \{\boldsymbol{\beta}, \mathbf{\Sigma}\}.$$

Common Priors for VARs

- Diffuse (flat): like OLS, essentially no shrinkage.
- Normal–Inverse Wishart (conjugate): analytical posterior, very fast but somewhat rigid.
- Independent Normal $\times$ Inverse Wishart: more flexible, sampled with Gibbs.
- Minnesota prior: we saw earlier how it encodes shrinkage toward persistence — the forecasting workhorse.

Diffuse Prior (Non-informative)

- Prior: flat / uninformative.
- Posterior = classical OLS with Gaussian errors.
- $\Rightarrow$ Little regularization, forecasts unstable in small samples.
- Useful as baseline comparison.

Conjugate NIW Prior

- Prior: $\boldsymbol{\beta} \mid \Sigma \sim \mathcal{N}(\beta_0, \Sigma \otimes \Lambda_0^{-1}),\ \ \Sigma \sim \mathcal{IW}(V_0, v_0)$.
- Posterior has closed form (conjugacy).
- Very fast updates $\Rightarrow$ used in large MCMC exercises.
- Limitation: separable covariance structure may be restrictive.

Independent Normal $\times$ Inverse Wishart Prior

- Prior factorizes: coefficients $\sim \mathcal{N}(\beta_0, \Lambda_0^{-1})$, covariance $\sim \mathcal{IW}(V_0, v_0)$.
- Allows separate control over coefficient priors and variance priors.
- Posterior not conjugate $\Rightarrow$ Gibbs sampling required.
- More flexible than NIW; widely used in Bayesian VAR literature.

Why Computation Matters

- Closed-form posteriors exist only for very special priors (e.g. NIW).
- With more flexible priors, no closed form $\Rightarrow$ need simulation.
- Markov Chain Monte Carlo (MCMC) methods let us approximate the posterior.
- Most common in BVARs: Gibbs sampling.

Gibbs Sampler for BVAR

1. Initialize $(B^{(0)}, \Sigma^{(0)})$.
2. For $r = 1, \ldots, R$:
   2.1 Draw $B^{(r)}$ given $\Sigma^{(r-1)}$ and the data.
   2.2 Draw $\Sigma^{(r)}$ given $B^{(r)}$ and the data.
3. After burn-in, keep draws for:
   - Forecast densities
   - Impulse responses

(Full conditional formulas in appendix.)

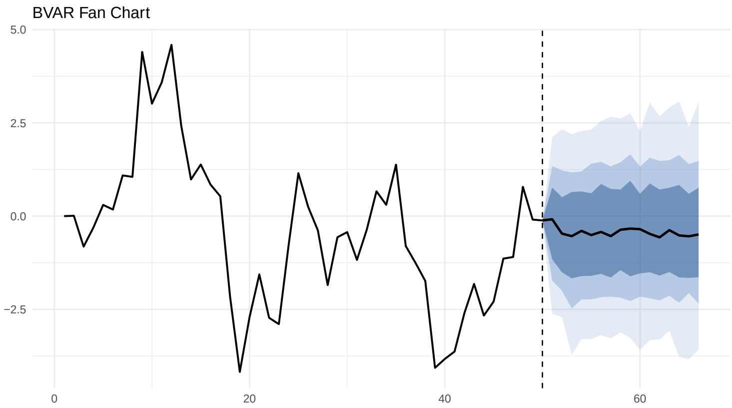
Practical Aspects of Gibbs Sampling

- Requires many draws depending on dimension.
- First iterations (burn-in) discarded.
- Convergence diagnostics important.
- After convergence, the sample approximates the posterior very well.
- Modern BVAR toolboxes handle all this automatically.

Forecasting with BVARs

- Forecast distribution integrates over both:
  - Parameter uncertainty (posterior draws of B, Σ).
  - Shock uncertainty (future ε_{t+h}).
- For each posterior draw:
  1. Draw parameters (B, Σ).
  2. Simulate future shocks $\varepsilon_{t+1}, \ldots, \varepsilon_{t+h}$.
  3. Iterate VAR forward to get forecast path.
- Many draws $\Rightarrow$ predictive density.

Forecast Fan Charts



- Fan charts show predictive density (e.g. 50%, 70%, 90% intervals).
- Standard in central bank communication.

Conditional Forecasts with BVARs

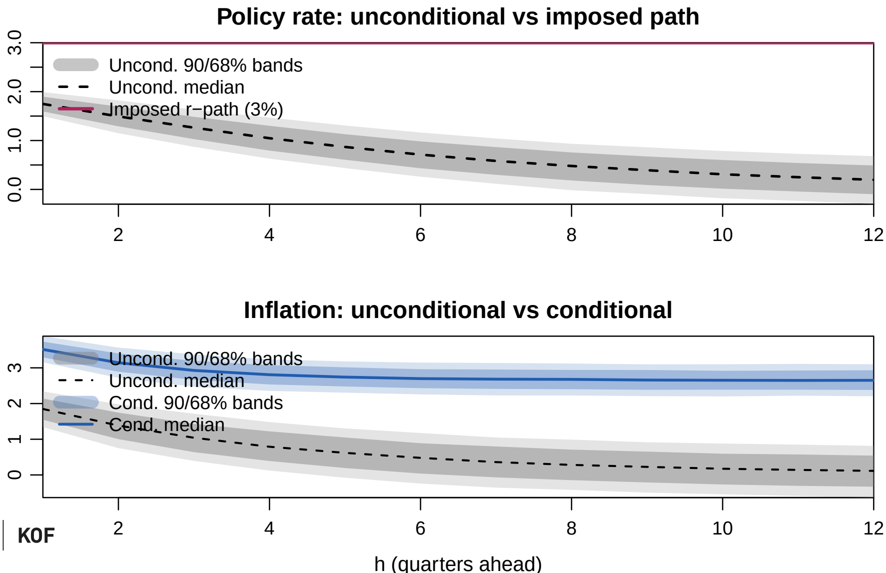
- Unconditional: project all variables forward from history.
- Conditional: impose paths on selected variables (e.g. interest rate, oil price), let others adjust consistently.
- Intuition: “What happens to GDP and inflation if rates follow this scenario?”
- Widely used in central banks for policy scenario analysis.

Conditional Forecasts: Algorithm

- Step 1: Draw parameters (B, Σ) from posterior.
- Step 2: Specify restrictions on selected variables (linear conditions, e.g. policy rate path).
- Step 3: Draw shocks from their conditional distribution such that the simulated paths satisfy the restrictions.
- Step 4: Other variables evolve consistently $\Rightarrow$ conditional predictive density.

Waggoner & Zha (1999), *Conditional Forecasts in Dynamic Multivariate Models*.

Example: Conditional Forecast



Takeaway: Forecasting with BVARs

- BVAR forecasts are densities, not just point predictions.
- Fan charts communicate uncertainty clearly.
- Conditional forecasts allow scenario and policy analysis.
- This is why central banks rely heavily on BVARs.

Structural Equation Models (SEMs)

Why SEMs in Forecasting?

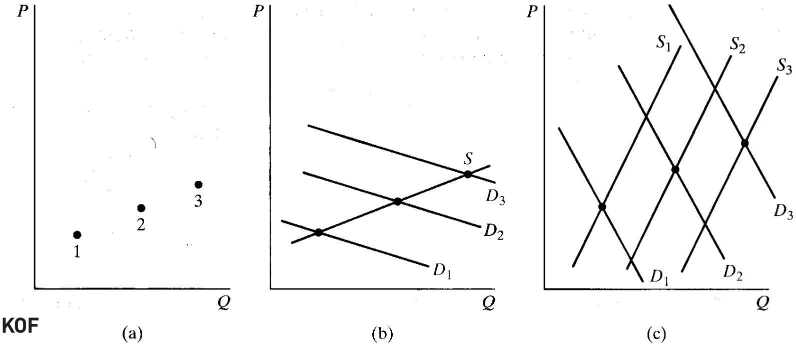
- **Transparent causal links:** coefficients have direct policy meaning ($x \to y$).
- **Scenario design:** can plug in exogenous assumptions (interest rates, fiscal policy, oil).
- **Accounting consistency:** identities ensure GDP = C+I+G+NX holds by construction.
- **Complement to VARs/SSMs:**
  - VARs/BVARs: short-run joint forecasts, IRFs, densities.
  - SEMs: policy scenarios, long-run narrative, easy interpretation.

SEMs vs. VARs

- VARs / BVARs
  - All variables endogenous; dynamics captured via lags.
  - Reduced-form errors correlated; identification via restrictions on shocks.
  - Focus: forecasting accuracy, shock propagation.
- SEMs
  - Equation-level causal links $x \to y$.
  - Rely on instruments, exclusion restrictions, and exogeneity.
  - Focus: policy multipliers, scenario analysis, narrative forecasts.

Illustrative SEM: Supply & Demand

- Demand: $q_{d,t} = \alpha_1 p_t + \alpha_2 x_t + \varepsilon_{d,t}$
- Supply: $q_{s,t} = \beta_1 p_t + \varepsilon_{s,t}$
- Equilibrium: $q_{d,t} = q_{s,t} = q_t$
- p_t, q_t endogenous, x_t exogenous.
- OLS inconsistent since p_t is correlated with shocks.
- **Solution:** use x_t as an instrument $\Rightarrow$ Two-Stage Least Squares.



Two-Stage Least Squares (2SLS)

1. First stage: $p_t = \pi_1 x_t + u_t \;\Rightarrow\; \hat{p}_t$
2. Second stage: $q_t = \beta_1 \hat{p}_t + e_t \;\Rightarrow\; \widehat{\beta}_1$

Instrument conditions:

- Relevance: Cov$(x_t, p_t) \neq 0$
- Exogeneity: Cov$(x_t, \varepsilon_{d,t}) =$ Cov$(x_t, \varepsilon_{s,t}) = 0$

With only one instrument: supply identified, demand not identified.

A Small Macro SEM

$$
\begin{aligned}
c_t &= \alpha_0 + \alpha_1 y_t + \alpha_2 c_{t-1} + \varepsilon_{t1} \\
i_t &= \beta_0 + \beta_1 r_t + \beta_2 (y_t - y_{t-1}) + \varepsilon_{t2} \\
y_t &= c_t + i_t + g_t \quad \text{(identity)}
\end{aligned}
$$

- Endogenous: c_t, i_t, y_t
- Exogenous: r_t, g_t
- Predetermined: c_{t-1}, y_{t-1}
- Identities ensure accounting consistency.

A Small Macroeconomic Model (Matrix Form)

- Define the vectors $\boldsymbol{y}_t' = \begin{bmatrix} c_t & i_t & y_t \end{bmatrix}$ and $\boldsymbol{x}_t' = \begin{bmatrix} 1 & r_t & g_t & c_{t-1} & y_{t-1} \end{bmatrix}$
- Then the system can be written in matrix form as:

$$\boldsymbol{y}_t'\boldsymbol{\Gamma} = \boldsymbol{x}_t'\boldsymbol{B} + \boldsymbol{\varepsilon}_t',$$

where

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -\alpha_1 & -\beta_2 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \alpha_0 & \beta_0 & 0 \\ 0 & \beta_1 & 0 \\ 0 & 0 & 1 \\ \alpha_2 & 0 & 0 \\ 0 & -\beta_2 & 0 \end{bmatrix}, \quad \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ 0 \end{bmatrix}$$

General Framework

- The structural form of the model is

$$
\begin{aligned}
\gamma_{11} y_{t1} + \gamma_{21} y_{t2} + ... + \gamma_{N1} y_{tN} &= \beta_{11} x_{t1} + ... + \beta_{K1} x_{tK} + \varepsilon_{t1} \\
\gamma_{12} y_{t1} + \gamma_{22} y_{t2} + ... + \gamma_{N2} y_{tN} &= \beta_{12} x_{t1} + ... + \beta_{K2} x_{tK} + \varepsilon_{t2} \\
\vdots \quad &= \quad \vdots \\
\gamma_{1N} y_{t1} + \gamma_{2N} y_{t2} + ... + \gamma_{NN} y_{tN} &= \beta_{1N} x_{t1} + ... + \beta_{KN} x_{tK} + \varepsilon_{tN}
\end{aligned}
$$

- N equations and N endogenous variables, x also includes intercept and lagged dependent variables

General Framework in Matrix Form

- The system can also be written in matrix form

$$
\begin{bmatrix} y_{1t} & \cdots & y_{Nt} \end{bmatrix}
\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1N} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2N} \\ & & \vdots & \\ \gamma_{N1} & \gamma_{N2} & \cdots & \gamma_{NN} \end{bmatrix}
= \begin{bmatrix} x_{1t} & \cdots & x_{Kt} \end{bmatrix}
\begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1N} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2N} \\ & & \vdots & \\ \beta_{K1} & \beta_{K2} & \cdots & \beta_{KN} \end{bmatrix}
+ \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} & \cdots & \varepsilon_{Nt} \end{bmatrix}
$$

- Thus

$$
\boldsymbol{y}_t' \boldsymbol{\Gamma} = \boldsymbol{x}_t' \boldsymbol{B} + \boldsymbol{\varepsilon}_t'
$$

- The underlying theory will imply restrictions on $\boldsymbol{\Gamma}$ and $\boldsymbol{B}$

General Framework in Matrix Form (contd.)

- The reduced form of the model is

$$\begin{aligned}\boldsymbol{y}_t' &= \boldsymbol{x}_t'\boldsymbol{B}\boldsymbol{\Gamma}^{-1} + \boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{-1} \\ &= \boldsymbol{x}_t'\boldsymbol{\Pi} + \boldsymbol{\nu}_t'\end{aligned}$$

- It follows that the reduced form errors $\boldsymbol{\nu}_t' = \boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{-1}$ have

$$E[\boldsymbol{\nu}_t] = \mathbf{0} \text{ and } E[\boldsymbol{\nu}_t\boldsymbol{\nu}_t'] = (\boldsymbol{\Gamma}^{-1})'\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1} = \boldsymbol{\Omega}$$

which implies that

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Omega}\boldsymbol{\Gamma}$$

Identification in SEMs

- Reduced form delivers $(\mathbf{\Pi}, \mathbf{\Omega})$.
- Structural form requires $(\mathbf{\Gamma}, \boldsymbol{B}, \mathbf{\Sigma})$.
- More unknowns than reduced-form parameters $\Rightarrow$ restrictions needed.
- Typical restrictions:
  - Normalization: one coefficient per equation normalized to 1.
  - Exclusions: some variables absent from certain equations.
  - Assumptions on error covariance structure.
- **Order condition (necessary):** # instruments $\geq$ # endogenous regressors.
- **Rank condition (sufficient):** excluded instruments must provide independent variation.

Dynamic SEMs & Multipliers

- Structural dynamic SEM:
$$\boldsymbol{y}_t'\boldsymbol{\Gamma} = \boldsymbol{x}_t'\boldsymbol{B} + \boldsymbol{y}_{t-1}'\boldsymbol{\Phi} + \boldsymbol{u}_t'$$
- Reduced form:
$$\boldsymbol{y}_t' = \boldsymbol{x}_t'\boldsymbol{\Pi} + \boldsymbol{y}_{t-1}'\boldsymbol{\Theta} + \boldsymbol{v}_t', \qquad \boldsymbol{\Pi} = \boldsymbol{B}\boldsymbol{\Gamma}^{-1},\ \boldsymbol{\Theta} = \boldsymbol{\Phi}\boldsymbol{\Gamma}^{-1}$$
- Short-run impact multipliers: $\boldsymbol{\Pi}$
- Dynamic multipliers: $\boldsymbol{\Pi}\boldsymbol{\Theta}^s$
- Long-run multipliers: $\boldsymbol{\Pi}(\boldsymbol{I} - \boldsymbol{\Theta})^{-1}$ (if $\rho(\boldsymbol{\Theta}) < 1$).

Forecasting with SEMs

- **Scenario-based forecasting:**
  - Plug in exogenous paths (interest rates, fiscal, oil prices).
  - Model delivers endogenous responses (consumption, investment, GDP).
- **Density forecasts:**
  - Combine posterior draws of parameters with scenarios.
  - Produce fan charts.
- **Key advantage:** SEMs are transparent, scenario-friendly, and accounting-consistent.

State-Space Models

Why State-Space Models?

- Many macro quantities are unobserved: output gap, latent trends, time-varying parameters.
- State-space models link observed data to hidden states that evolve over time.
- Unified framework for:
  - Trend–cycle decompositions
  - Dynamic factors
  - Time-varying parameters
  - Missing data handling

The State-Space Representation of a Dynamic System

- The vector of variables $\boldsymbol{y}_t$ observed at date t can described in terms of a possibly unobserved state vector $\boldsymbol{\alpha}_t$
- The state-space representation of $\boldsymbol{y}$ is then given by

$$
\begin{aligned}
\underset{n\times 1}{\boldsymbol{y}_t} &= \underset{n\times k}{\boldsymbol{A}}\underset{k\times 1}{\boldsymbol{x}_t} + \underset{n\times r}{\boldsymbol{H}}\underset{r\times 1}{\boldsymbol{\alpha_t}} + \underset{n\times 1}{\boldsymbol{w}_t} && \text{observation equation} \quad (1)\\
\underset{r\times 1}{\boldsymbol{\alpha_t}} &= \underset{r\times m}{\boldsymbol{B}}\underset{m\times 1}{\boldsymbol{z}_t} + \underset{r\times r}{\boldsymbol{F}}\underset{r\times 1}{\boldsymbol{\alpha_{t-1}}} + \underset{r\times 1}{\boldsymbol{v}_t} && \text{state equation} \quad (2)
\end{aligned}
$$

$\boldsymbol{A}$, $\boldsymbol{H}$, $\boldsymbol{B}$ and $\boldsymbol{F}$ are matrices of parameters, $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ are vectors of exogenous variables, $E(\boldsymbol{w}_t\boldsymbol{w}_t') = \boldsymbol{R}$, $E(\boldsymbol{v}_t\boldsymbol{v}_t') = \boldsymbol{Q}$ and $E(\boldsymbol{v}_t\boldsymbol{w}_t') = \boldsymbol{0}$, where $\boldsymbol{Q}$ and $\boldsymbol{R}$ are $r\times r$ and $n\times n$ matrices

Example: AR(p) as a State-Space Model

- AR(p): $\quad y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t.$
- Choose state $\alpha_t = [\, y_t,\, y_{t-1},\, \ldots,\, y_{t-p+1} \,]'$.
- Observation: $\quad y_t = [1\ 0\ \cdots\ 0]\, \alpha_t.$
- State:

$$\alpha_t = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \varepsilon_t \sim N(0, \sigma^2).$$

Example: Unobserved Components (Trend + Cycle)

- Decompose y_t into trend τ_t and cycle c_t:

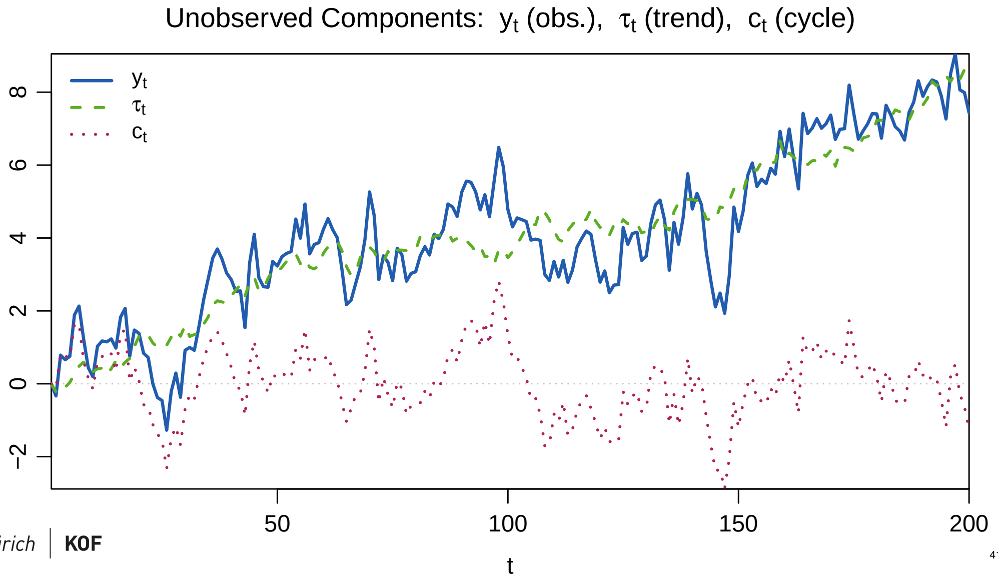
$$y_t = \tau_t + c_t, \qquad \tau_t = \delta + \tau_{t-1} + \nu_t, \qquad c_t = \phi c_{t-1} + \eta_t,$$

  with $|\phi| < 1$, $\begin{bmatrix} \nu_t \\ \eta_t \end{bmatrix} \sim N(0, \Sigma)$.

- State: $\alpha_t = [\,\tau_t,\ c_t\,]'$; observation: $y_t = [1\ 1]\alpha_t$.
- State transition:

$$\alpha_t = \begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} \delta \\ 0 \end{bmatrix} + \begin{bmatrix} \nu_t \\ \eta_t \end{bmatrix}.$$

Example: UC with Simulated Data



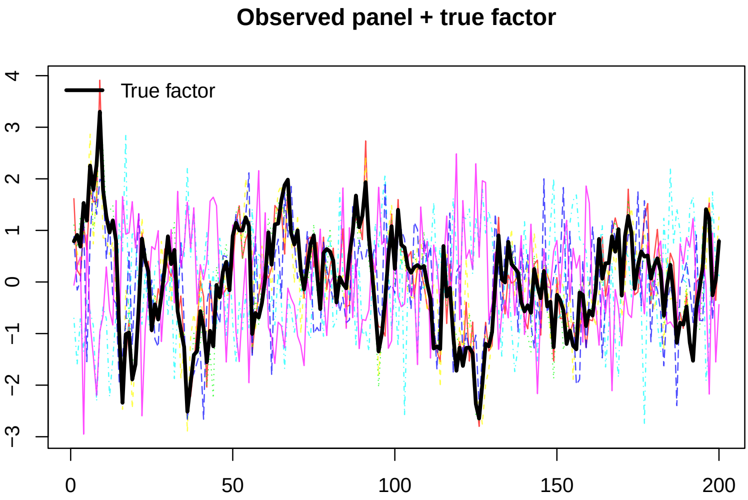
Example: Dynamic Factor Model (Activity Index)

- One common factor f_t drives n observables:

$$y_{it} = c_i + \lambda_i f_t + \nu_{it}, \quad i = 1, \ldots, n; \;\; f_t = \phi f_{t-1} + \eta_t.$$

- Observation: $y_t = A + H\alpha_t + w_t$ with $A = [c_i]$, $H = [\lambda_i]$, $\alpha_t = f_t$.
- Idiosyncratic: $w_t \sim N(0, R)$ with $R = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.
- Identification: fix sign/scale (e.g., $\lambda_1 > 0$, $\sigma_f^2 = 1$) or triangular H for $r > 1$.

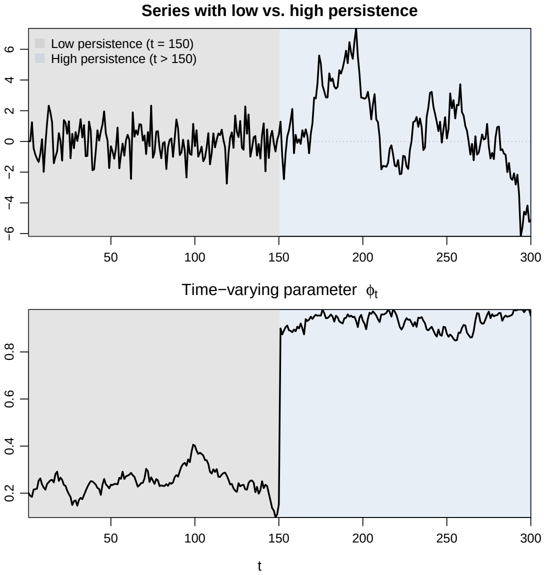
Example: DFM with Simulated Data



Example: Time-Varying Parameters (TVP-AR)

- Observation equation: $y_t = \phi_t \, y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_y^2)$.
- State equation: $\phi_t = \phi_{t-1} + \nu_t, \quad \nu_t \sim N(0, \sigma_\phi^2)$.
- Write as SSM with state $\alpha_t = \phi_t$, observation matrix $H_t = y_{t-1}$ (time-varying H).

Example: TVP-AR with Simulated Data



Bayesian Estimation of State-Space Models

- Unknowns: parameters
$$\Theta = \{\boldsymbol{H}, \boldsymbol{F}, \boldsymbol{R}, \boldsymbol{Q}, (\text{optionally } \boldsymbol{A}, \boldsymbol{B})\},$$
and the latent states $\boldsymbol{\alpha}_{1:T}$.
- Goal: posterior distribution
$$p(\Theta, \boldsymbol{\alpha}_{1:T} \mid \boldsymbol{Y}).$$
- Use Gibbs sampling:
  - Draw $\boldsymbol{\alpha}_{1:T} \mid \Theta, \boldsymbol{Y}$ via FFBS (forward-filtering, backward-sampling).
  - Draw $\Theta \mid \boldsymbol{\alpha}_{1:T}, \boldsymbol{Y}$ from conjugate blocks (e.g., Normal for coefficients, inverse-Wishart for covariances) or via Metropolis–Hastings.

Forward-Filtering, Backward-Sampling (FFBS)

- Forward (Kalman filter): compute $\boldsymbol{\alpha}_{t|t-1}, \boldsymbol{P}_{t|t-1}$ and $\boldsymbol{\alpha}_{t|t}, \boldsymbol{P}_{t|t}$ for $t = 1, \ldots, T$.
- Backward (sampling): draw $\boldsymbol{\alpha}_T \sim \mathcal{N}(\boldsymbol{\alpha}_{T|T}, \boldsymbol{P}_{T|T})$, then for $t = T-1, \ldots, 1$

$$\boldsymbol{\alpha}_t \sim \mathcal{N}\left(\boldsymbol{\alpha}_{t|t} + \boldsymbol{J}_t(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_{t+1|t}),\;\; \boldsymbol{P}_{t|t} - \boldsymbol{J}_t \boldsymbol{P}_{t+1|t} \boldsymbol{J}_t'\right),$$

  where

$$\boldsymbol{J}_t = \boldsymbol{P}_{t|t} \boldsymbol{F}' (\boldsymbol{P}_{t+1|t})^{-1}.$$

- Iterate within a Gibbs loop with parameter updates.

Summary of the Kalman Filter

Given the initial conditions $\boldsymbol{\alpha}_{0|0}$, $\boldsymbol{P}_{0|0}$ and the parameter matrices $(\boldsymbol{A})$, $(\boldsymbol{B})$, $\boldsymbol{H}$, $\boldsymbol{F}$, $\boldsymbol{R}$ and $\boldsymbol{Q}$ we iterate through the following steps for $t = 1, 2, ..., T$

Forecasting steps

$$\begin{aligned}
\boldsymbol{\alpha}_{t|t-1} &= \boldsymbol{F}\boldsymbol{\alpha}_{t-1|t-1} \\
\boldsymbol{P}_{t|t-1} &= \boldsymbol{F}\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{Q} \\
\boldsymbol{y}_{t|t-1} &= \boldsymbol{H}\boldsymbol{\alpha}_{t|t-1} \\
\boldsymbol{S}_{t|t-1} &= \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}' + \boldsymbol{R}
\end{aligned}$$

Updating steps

$$\begin{aligned}
\boldsymbol{\alpha}_{t|t} &= \boldsymbol{\alpha}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}'\boldsymbol{S}_{t|t-1}^{-1}(\boldsymbol{y}_t - \boldsymbol{y}_{t|t-1}) \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}'\boldsymbol{S}_{t|t-1}^{-1}\boldsymbol{H}\boldsymbol{P}_{t|t-1}
\end{aligned}$$

Mixed-Frequency Approaches

Why Mixed-Frequency Models?

- Forecast targets often observed at low frequency (e.g., quarterly GDP, quarterly inflation).
- But many predictors are available at higher frequency (monthly surveys, daily financial data).
- Problem: Standard time-series models require matching frequencies $\Rightarrow$ throw away information or aggregate predictors.
- Solution: Mixed-frequency models use high-frequency indicators *directly*, without discarding data.
- Two main approaches:
  - State-space: treat missing low-frequency values as latent, handle via Kalman filter.
  - MIDAS regressions: regress directly on high-frequency lags with parsimonious lag polynomials.

MIDAS Regression: Basic Form

- Schematic regression:

$$y_t = \beta_0 + \sum_{k=0}^{K} \beta(k;\theta)\, x_{t-k/m} + \varepsilon_t$$

  - y_t: low-frequency target (e.g., quarterly GDP).
  - $x_{t-k/m}$: high-frequency regressor (e.g., monthly PMI, $m = 3$ months per quarter).
  - $\beta(k;\theta)$: lag weights governed by a few parameters (e.g., exponential Almon).
- Avoids estimating a separate β for each high-frequency lag $\Rightarrow$ parsimonious and stable.
- Estimation: nonlinear least squares or Bayesian methods.
- Widely used for nowcasting: combining daily/monthly indicators with quarterly targets in real time.

References: Ghysels, Santa-Clara, Valkanov (2007); Ghysels et al. (2016). R package: `midasr`.

State-Space Mixed-Frequency Models

- Alternative: embed the mixed-frequency structure into a state-space model.
- Intuition:
  - Low-frequency series treated as observed only at some dates.
  - Missing values in between are handled as latent states.
  - Kalman filter/smoother integrates information from all available high-frequency predictors.
- Flexibility: can handle multiple predictors, dynamic interactions, and structural restrictions.
- Applications: central banks often use state-space mixed-frequency VARs for forecasting GDP and inflation.

Reference: Mariano & Murasawa (2003); Schorfheide & Song (2015).

Intuition: Quarterly GDP vs. Monthly Indicators

- Quarterly GDP is sparse: only one observation every three months.
- Monthly surveys/indicators provide more frequent signals.
- Mixed-frequency methods combine them:
- $\rightarrow$ “fill in the gaps” between quarterly data using higher-frequency information.

Mixed-Frequency VARs: Intuition

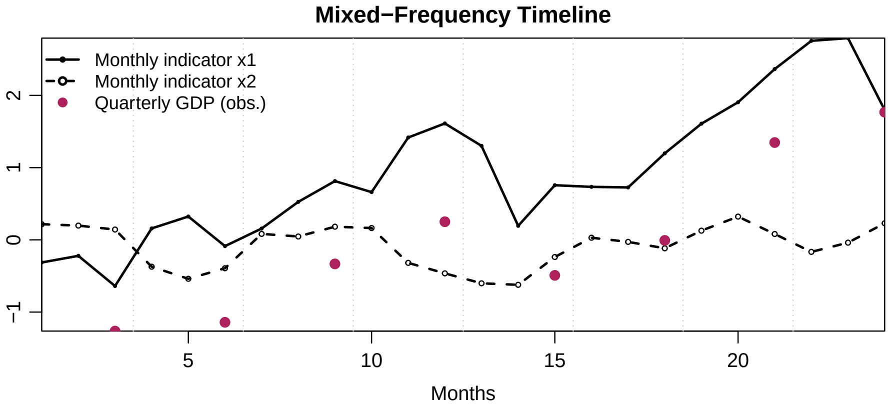
- Generalization of VAR models to allow variables sampled at different frequencies.
- Example: quarterly GDP, monthly inflation, daily financial indicators in one system.
- State-space form:
  - Observation equation handles missing values for low-frequency series.
  - Transition equation is the usual VAR dynamics.
- Estimation typically Bayesian (Minnesota priors, Gibbs sampling).

Reference: Schorfheide & Song (2015, Journal of Business and Economic Statistics).

Why MF-VARs?

- Captures interactions between multiple variables at mixed frequencies.
- Consistent forecasting framework:
  - Use monthly and quarterly data without aggregation.
  - Generate density forecasts for quarterly targets.
- Natural extension of the widely used VAR framework.
- Flexible: can incorporate priors, structural restrictions, or time variation.

MF-VAR Intuition: Timeline



- Quarterly GDP “fills in” only once per quarter.
- Monthly variables provide extra observations in between.
- State-space representation reconciles them within the VAR structure.

Appendix

BVAR with Diffuse (Noninformative) Prior

- Model: $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \;\; \mathrm{vec}(\mathbf{E}) \sim \mathcal{N}(\mathbf{0},\, \boldsymbol{\Sigma} \otimes \mathbf{I}_T)$.
- Diffuse prior: $p(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(n+1)/2}$.
- Posterior:

$$\mathrm{vec}(\widehat{\mathbf{B}}) = \mathrm{vec}\Big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\Big), \quad \widehat{\mathbf{E}} = \mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}.$$

$$\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \mathbf{Y}, \mathbf{X} \sim \mathcal{N}\Big(\mathrm{vec}(\widehat{\mathbf{B}}),\, \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}\Big),$$

$$\boldsymbol{\Sigma} \mid \mathbf{Y}, \mathbf{X} \sim \mathcal{IW}\Big(\widehat{\mathbf{E}}'\widehat{\mathbf{E}},\, T - m\Big).$$

- Intuition: with a flat prior, the Bayesian posterior coincides with classical OLS uncertainty.

Conjugate NIW Prior for VAR

- Prior:

$$\boldsymbol{\beta} \mid \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\beta}_0,\ \boldsymbol{\Sigma} \otimes \boldsymbol{\Lambda}_0^{-1}), \qquad \boldsymbol{\Sigma} \sim \mathcal{IW}(\mathbf{V}_0,\ v_0).$$

- Posterior:

$$\boldsymbol{\Lambda}_1 = \mathbf{X}'\mathbf{X} + \boldsymbol{\Lambda}_0, \qquad \mathbf{B}_1 = \boldsymbol{\Lambda}_1^{-1}\big(\mathbf{X}'\mathbf{Y} + \boldsymbol{\Lambda}_0\,\mathbf{B}_0\big), \quad \boldsymbol{\beta}_1 = \mathrm{vec}(\mathbf{B}_1).$$

$$\mathbf{V}_1 = (\mathbf{Y} - \mathbf{X}\mathbf{B}_1)'(\mathbf{Y} - \mathbf{X}\mathbf{B}_1) + (\mathbf{B}_1 - \mathbf{B}_0)'\boldsymbol{\Lambda}_0(\mathbf{B}_1 - \mathbf{B}_0) + \mathbf{V}_0, \qquad v_1 = v_0 + T.$$

$$\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \mathbf{Y}, \mathbf{X} \sim \mathcal{N}\big(\boldsymbol{\beta}_1,\ \boldsymbol{\Sigma} \otimes \boldsymbol{\Lambda}_1^{-1}\big), \qquad \boldsymbol{\Sigma} \mid \mathbf{Y}, \mathbf{X} \sim \mathcal{IW}(\mathbf{V}_1, v_1).$$

- Pros: conjugacy $\Rightarrow$ very fast updates; Cons: separable covariance structure may be restrictive.

Independent Normal $\times$ Inverse–Wishart Prior

- Prior factorizes:

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0,\ \boldsymbol{\Lambda}_0^{-1}), \qquad \boldsymbol{\Sigma} \sim \mathcal{IW}(\mathbf{V}_0,\ v_0).$$

- Conditionals (closed form) $\Rightarrow$ Gibbs sampling:

$$\underbrace{\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \mathbf{Y}, \mathbf{X}}_{\text{Normal}} \sim \mathcal{N}(\boldsymbol{\beta}_1,\ \boldsymbol{\Lambda}_1^{-1}), \qquad \boldsymbol{\Lambda}_1 = (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X}) + \boldsymbol{\Lambda}_0,$$

$$\boldsymbol{\beta}_1 = \boldsymbol{\Lambda}_1^{-1}\Big[(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X})\,\mathrm{vec}(\widehat{\mathbf{B}}) \ + \ \boldsymbol{\Lambda}_0\,\boldsymbol{\beta}_0\Big],$$

$$\underbrace{\boldsymbol{\Sigma} \mid \mathbf{B}, \mathbf{Y}, \mathbf{X}}_{\text{Inverse–Wishart}} \sim \mathcal{IW}\Big(\mathbf{V}_1,\ v_1\Big), \quad \mathbf{V}_1 = (\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB}) + \mathbf{V}_0, \quad v_1 = v_0 + T.$$

- Gibbs steps: alternate draws of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ from these conditionals.

Gibbs Sampler for Independent Normal $\times$ IW Prior

1. Initialize $(\mathbf{B}^{(0)}, \boldsymbol{\Sigma}^{(0)})$.
2. For $r = 1, \ldots, R$:
   2.1 Draw $\boldsymbol{\beta}^{(r)} \sim \mathcal{N}(\boldsymbol{\beta}_1, \boldsymbol{\Lambda}_1^{-1})$ using $\boldsymbol{\Sigma}^{(r-1)}$ and the formulas on previous slide.
   2.2 Reshape $\boldsymbol{\beta}^{(r)} \mapsto \mathbf{B}^{(r)}$.
   2.3 Draw $\boldsymbol{\Sigma}^{(r)} \sim \mathcal{IW}(\mathbf{V}_1, v_1)$ using $\mathbf{B}^{(r)}$.
3. After burn-in, keep draws for density forecasts and IRFs.