Methods of Macroeconomic Forecasting

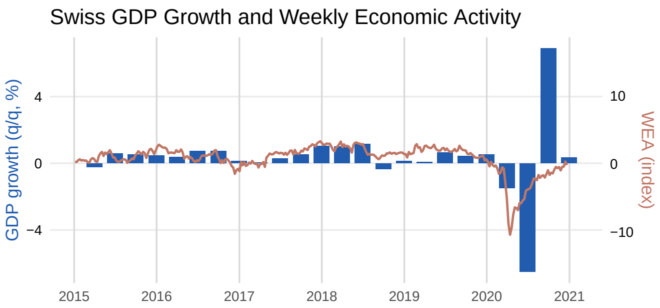
Methods Boot Camp – Day 1
KOF ETH Zurich
October 2, 2025, Zurich

Course Overview

- Hands-on introduction to modern time series forecasting methods used in central banks and research institutes.
- Emphasis on both theory and practice: lectures paired with computer labs in R.
- Core methods: vector autoregressive and simultaneous equation systems, state-space models, mixed-frequency nowcasting.
- You will apply these methods in a group forecasting project and present your results to the class.

Why Forecasting Matters: COVID-19

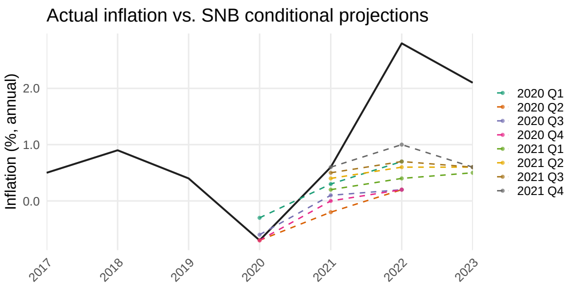
- During COVID, official GDP arrived with long delays.
- Policymakers needed a timely proxy for economic activity.
- SECO's Weekly Economic Activity (WEA) index provided weekly updates.
- Composite of high-frequency indicators (mobility, payments, energy).



Source: SECO.

Why Forecasting Matters: Inflation Surge

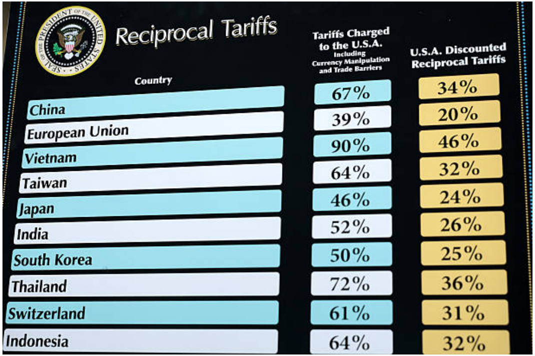
- After the pandemic, inflation surged far beyond many forecasts.
- Forecast errors exposed limits of traditional models.
- Highlights the importance of uncertainty bands and model choice.



Source: SNB

Why Forecasting Matters: Trade Policy Shocks

- Trade policy shocks such as the Trump tariffs disrupted trade flows.
- Such shocks are hard to anticipate with standard models.
- Forecasting requires scenario analysis.



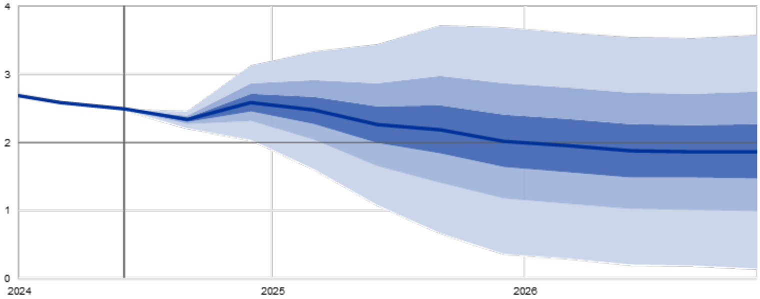
Learning Outcomes

After completing this course, you will be able to:

- Implement forecasting models (VARs, SEMs, state-space, mixed-frequency) in R.
- Produce both point and density forecasts.
- Evaluate forecasts systematically using standard metrics.
- Design conditional policy scenarios (e.g., oil prices, exchange rates, monetary/fiscal policy).
- Communicate results effectively to both technical peers and policy audiences.

Example Learning Outcome: Density Forecasts

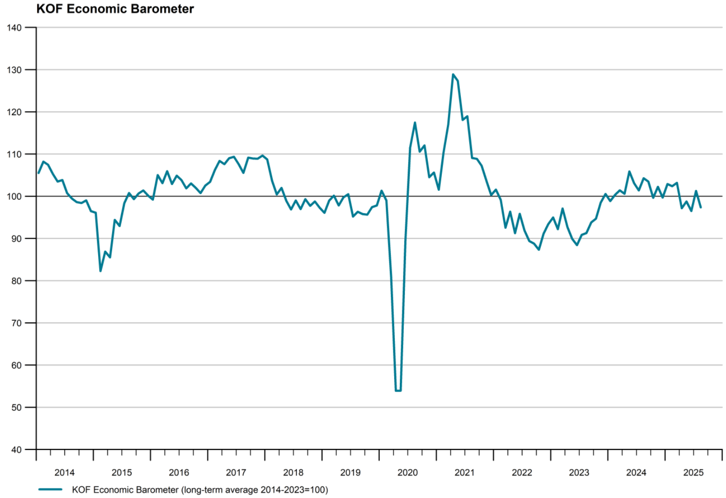
- Beyond point forecasts, density forecasts communicate uncertainty.
- Fan charts are commonly used in central banks to show predictive distributions.



Source ECB, September 2024 staff projections.

Example Learning Outcome: Economic Indicators

- Construct composite economic indicators from multiple series.
- Indicators help track the state of the economy in real time.



ETH zürich | KOF

Scope & Further Reading

- Due to time constraints, many methods will be covered only at a high level.
- We focus on hands-on intuition and applications, not full technical derivations.
- For deeper study, see:

Topic	Reference
Time series basics	Hamilton (1994): *Time Series Analysis* Lütkepohl (2005): *New Introduction to Multiple Time Series Analysis*
SEMs	Green (2003): *Econometric analysis, Chapter 15*
Bayesian econometrics	Koop (2003): *Bayesian Econometrics* Lancaster (2004): *Introduction to Modern Bayesian Econometrics*
State-space models	Durbin & Koopman (2012): *Time Series Analysis by State Space Methods* Kim & Nelson (1999): *State-Space Models with Regime Switching*
Forecast evaluation	Diebold (2015): *Forecasting in Economics, Business, Finance and Beyond*

Instructors

Meet Your Instructors

Samad Sarferaz

Responsible for lectures

LEE G 302

`sarferaz@kof.ethz.ch`

Office hours on appointment

Lena Will

Responsible for labs

LEE G 219

`will@kof.ethz.ch`

Office hours on appointment

Merlin Scherer

Responsible for labs

LEE G 219

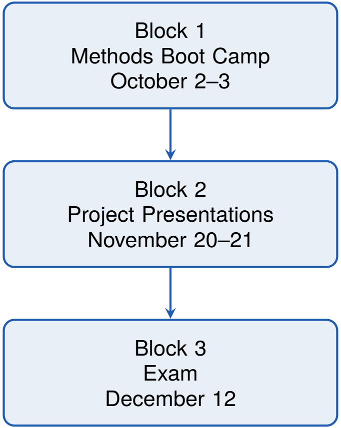
`scherer@kof.ethz.ch`

Office hours on appointment

Logistics & Performance Assessment

ETH zürich KOF

Course Roadmap & Logistics



- Format:
  - Block 1: lectures + computer labs
  - Block 2: student presentations in small groups
  - Block 3: written computer exam
- Materials: slides, lab scripts, datasets via Git repository

Methods Boot Camp (Block 1)

Day 1 – Lectures

- Autoregression (AR)
- Vector autoregression (VAR)
- Bayesian basics
- Forecast evaluation and combination

Day 1 – Computer Lab

- Reproducibility
- Forecasting baseline

Day 2 – Lectures

- Simultaneous Equation Model (SEM)
- State-space models
- Mixed-frequency methods
- Bayesian VARs (BVARs)

Day 2 – Computer Lab

- Forecasting with BVARs, SEMs, and state space models

Project Presentations (Block 2)

- Work in small groups of 2–3 people.
- Forecasting exercise: apply methods from class to Swiss data.
- Goals:
  - Design, implement, and critically assess multivariate forecasting models.
  - Gain experience presenting results to technical and policy audiences.
- Each project focuses on one specific forecasting technique.
- Tasks:
  - Understand method and communicate the idea behind.
  - Apply it to Swiss data (know your data).
  - Present and defend your results.

Project Presentations (Reproducibility)

- Reproducibility is essential in both academic research and policy forecasting.
- All code must be made publicly available in our GitHub repository (see project descriptions PDF).
- Each team member must contribute:
  - At least two meaningful commits per person.
  - Demonstrating shared responsibility for the project.
- Deliverables:
  - Slides and reproducible code in the GitHub repo.

Structure of a Good Project Presentation

- Motivation & Question: why does this forecast matter?
- Methodology: chosen model and estimation approach
- Data: variables, frequency, sample period
- Results & Forecast Evaluation: forecasts and assessment (details on next slide)
- Interpretation: economic meaning and intuition
- Takeaways: main findings and policy relevance
- Reproducibility: GitHub repository with code & data

30–45 minutes, focus on clarity and communication – not on every technical detail.

Forecast Expectations

- Each project should present at least:
  - A one-step and one-year ahead point forecast for GDP growth, inflation, and exchange rate.
  - One of the following three extensions
    - Scenario forecast (e.g., conditional on oil prices, exchange rates, or policy).
    - Density forecast (e.g., fan chart).
    - Macroeconomic indicator (e.g., diffusion index).
- Forecast evaluation:
  - Root mean squared errors (RMSE) or mean absolute error (MAE).
  - All models should be compared against an AR(2) benchmark.

Goal: show forecasts, assess their quality, and discuss what we learn from them.

Performance Assessment

- Project Presentation (60% of grade)
  - Group work (2–3 students), reproducibility via Git.
  - Evaluation: methods, results, communication, teamwork.
  - Heavier weight: reflects that the core skills of the course are hands-on implementation and communication of forecasts.
- Final Exam (40% of grade)
  - Duration: 60 minutes.
  - Format: mix of theory questions and short R-based tasks.
  - Focus: individual understanding and interpretation of results.

Project Presentations (Evaluation Criteria)

- Modeling and implementation in R (25%) Correct application of methods; transparent workflow.
- Clarity of communication (25%) Well-structured slides; clear delivery; appropriate for both technical and policy audiences.
- Forecast evaluation (20%) Point and density forecast accuracy; comparison with benchmark model.
- Design of extension and economic interpretation (15%) Linking forecasts to policy-relevant questions.
- Reproducibility and teamwork (15%) Public Git repository; meaningful contributions from all members.

Final Exam: What to Expect

- Covers the full Methods Boot Camp (AR, VAR, BVAR, SEM, state-space, MIDAS, forecast evaluation).
- Emphasis on:
  - Understanding model intuition and forecasting applications.
  - Interpreting outputs (IRFs, fan charts, density forecasts).
  - Explaining differences between methods and their trade-offs.
- You will **not** be asked to reproduce full derivations.

Project Topics

Topics covered in class

- 1. Bayesian VARs
- 2. Bayesian SEMs
- 3. Dynamic Factor Models
- 4. Time-Varying Parameter Models
- 5. Mixed-Frequency Models
- 6. Forecast Combinations

Machine Learning Topics

- 7. Forecasting with Transformers
- 8. Forecasting with Random Forests
- 9. Forecasting with Gaussian Processes

Pick Your Own Topic

- Propose a topic of your choice.
- Please email us to discuss in advance.

A detailed document with all topics and suggested literature will be uploaded to our Git repository soon after this method boot camp.

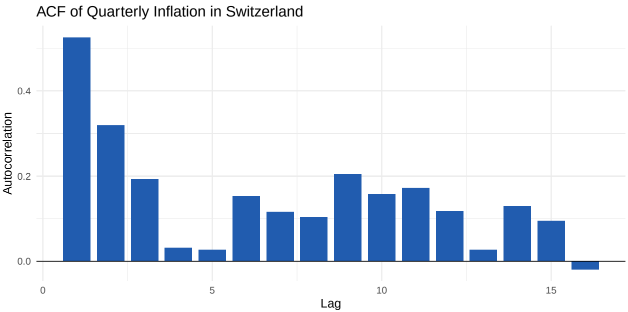
Univariate Time Series Models

Why Time Series Models?

- Many macroeconomic series show persistence: shocks today influence the future.
- Forecasting requires exploiting this serial correlation.
- Example: GDP, unemployment, or inflation move slowly and are correlated with their own past.
- $\Rightarrow$ We need a simple model that captures persistence.

From Persistence to Modeling

- Data show persistence: shocks fade only gradually.
- The autocorrelation function (ACF) confirms it.
- $\Rightarrow$ How can we capture this formally?



AR(1): Capturing Persistence

- Simplest model of persistence: AR(1)

$$y_t = c + \phi y_{t-1} + \varepsilon_t$$

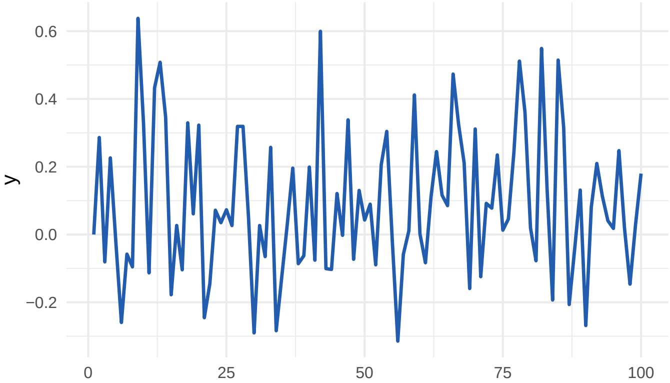
- c: intercept, ε_t: shock with mean 0.
- If $|\phi| < 1$, the process is stationary with long-run mean

$$\mu = \tfrac{c}{1-\phi}.$$

- Intuition: today's value depends on yesterday's value plus a shock.
- Persistence and mean reversion are determined by ϕ.

AR(1): Simulated Paths

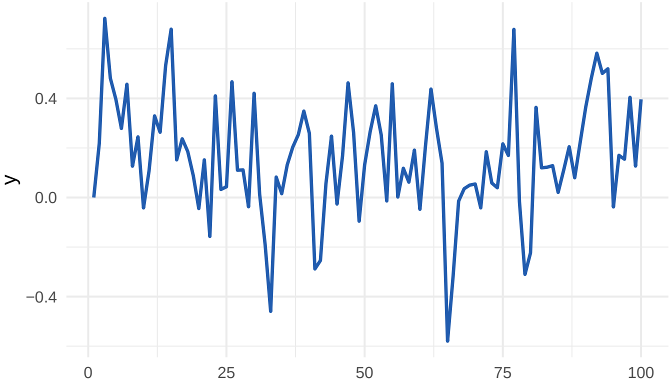
Model: $y_t = c + \phi y_{t-1} + \varepsilon_t$



$c = 0.1,\ \phi = 0$

AR(1): Simulated Paths

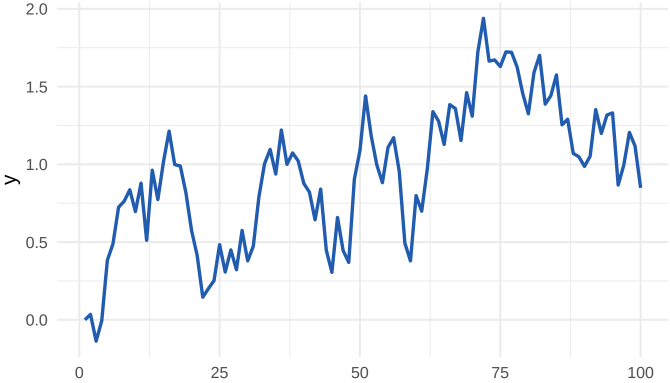
Model: $y_t = c + \phi y_{t-1} + \varepsilon_t$



$c = 0.1, \, \phi = 0.5$

AR(1): Simulated Paths

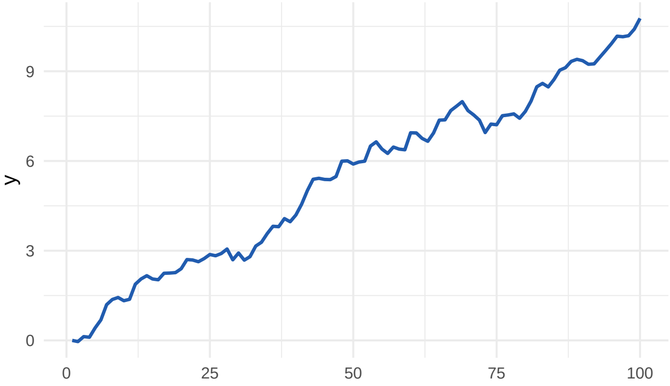
Model: $y_t = c + \phi y_{t-1} + \varepsilon_t$



$c = 0.1,\ \phi = 0.9$

AR(1): Simulated Paths

Model: $y_t = c + \phi y_{t-1} + \varepsilon_t$



$c = 0.1, \phi = 1$

AR(1): Impulse Response Function

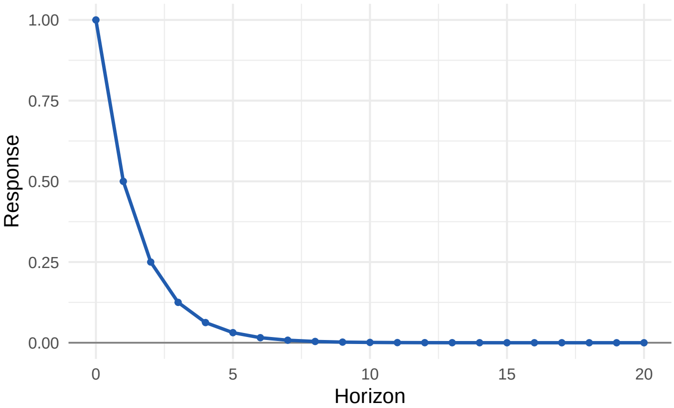
$$y_t = \phi y_{t-1} + \varepsilon_t, \qquad \psi_j \equiv \frac{\partial \mathbb{E}[y_{t+j}]}{\partial \varepsilon_t} = \phi^j$$

One-time shock at t:

- $|\phi| < 1$: geometric decay.
- $\phi = 1$: permanent effect.
- $|\phi| > 1$: explosive; $\phi < 0$ alternates in sign.

AR(1): IRFs across regimes

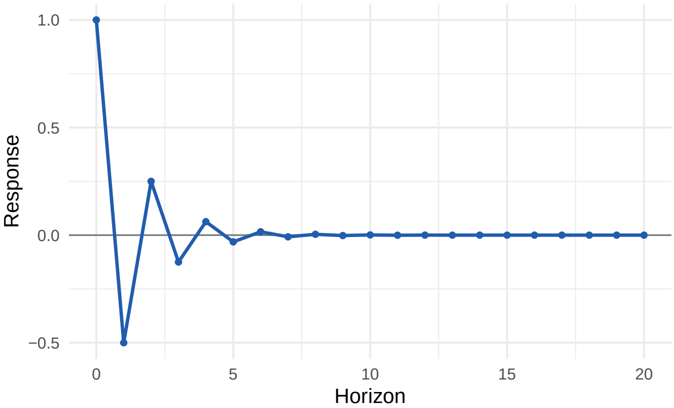
Model: $y_t = \phi y_{t-1} + \varepsilon_t$, IRF $\psi_j = \phi^j$



$\phi = 0.5$ (stationary: geometric decay)

AR(1): IRFs across regimes

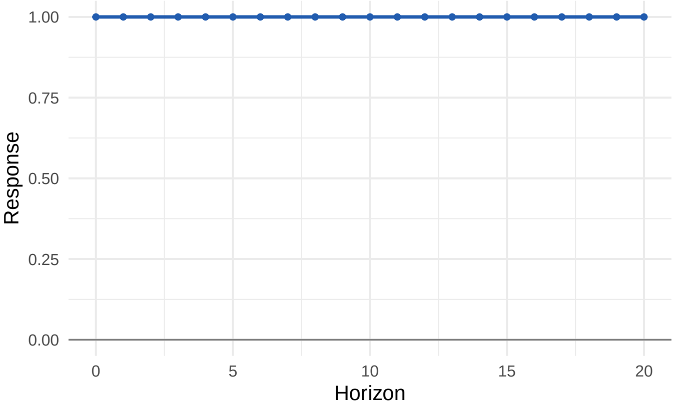
Model: $y_t = \phi y_{t-1} + \varepsilon_t$, IRF $\psi_j = \phi^j$



$\phi = -0.5$ (stationary: alternating decay)

AR(1): IRFs across regimes

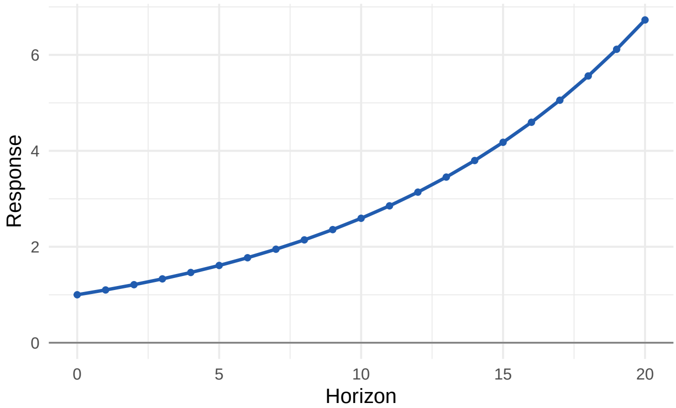
Model: $y_t = \phi y_{t-1} + \varepsilon_t$, IRF $\psi_j = \phi^j$



$\phi = 1$ (unit root: permanent effect)

AR(1): IRFs across regimes

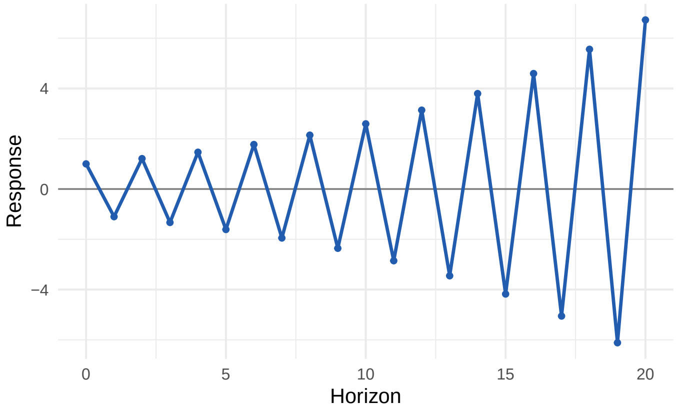
Model: $y_t = \phi y_{t-1} + \varepsilon_t$, IRF $\psi_j = \phi^j$



$\phi = 1.1$ (explosive: geometric growth)

AR(1): IRFs across regimes

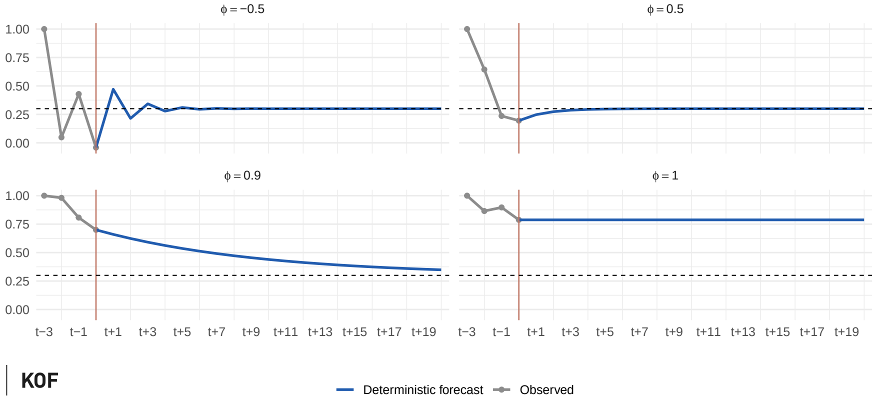
Model: $y_t = \phi y_{t-1} + \varepsilon_t$, IRF $\psi_j = \phi^j$



$\phi = -1.1$ (explosive: oscillating growth)

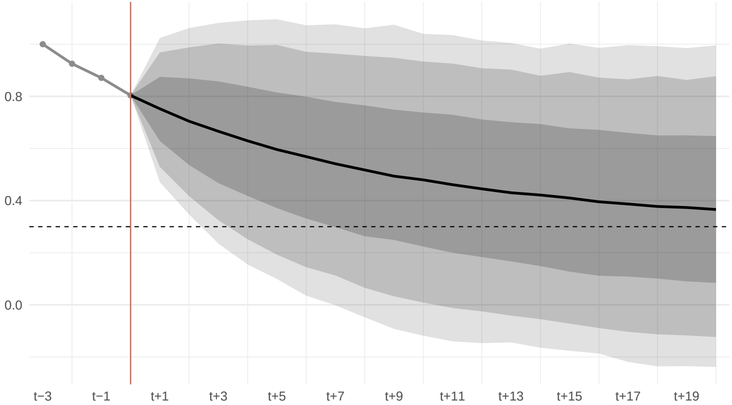
AR(1): Forecast Profiles

- One-step: $\hat{y}_{t+1|t} = c + \phi y_t$.
- h-step: $\hat{y}_{t+h|t} = \mu + \phi^h (y_t - \mu)$, with $\mu = \dfrac{c}{1-\phi}$ (if $|\phi| < 1$).
- As $h \uparrow$, $\phi^h \downarrow \Rightarrow$ forecast drifts to μ.



Forecast Uncertainty in AR(1)

- Forecast error variance grows with horizon h.
- Approaches the unconditional variance $\frac{\sigma^2}{1-\phi^2}$ as $h \to \infty$.

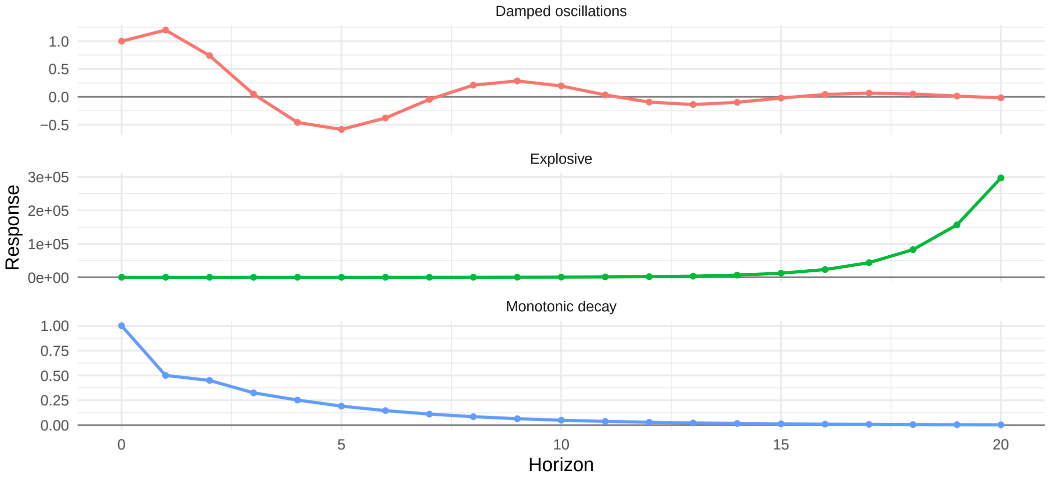


Higher-Order AR Models

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

- Capture richer dynamics (e.g., damped cycles with AR(2)).
- Interpretation still about persistence and mean reversion.

AR(2): Dynamic Patterns in the IRF



Moving Average (MA) Models

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \qquad \varepsilon_t \sim \text{i.i.d. } (0, \sigma^2)$$

- ε_t is the innovation (one-step-ahead forecast error).
- MA models capture short-lived autocorrelation in the data.
- Fingerprint: ACF cuts off at lag q.
- Impulse response: directly given by coefficients $(1, \theta_1, \ldots, \theta_q)$; beyond lag q, the effect is zero.

ARMA and ARIMA Models

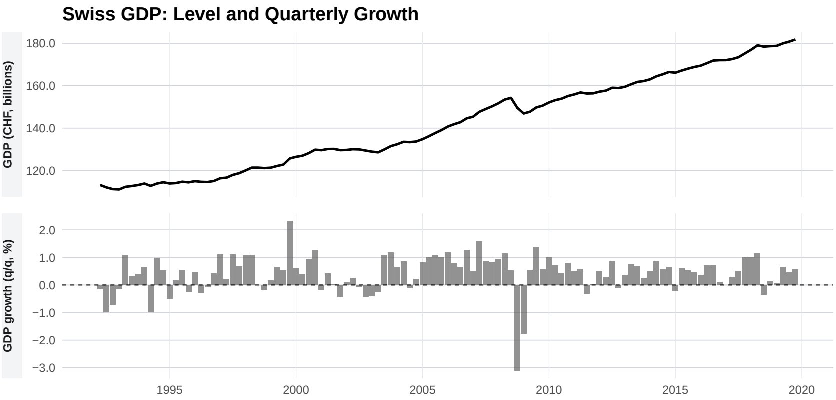
- ARMA(p, q): combines two mechanisms
  - AR part: persistence via lags of y_t.
  - MA part: short-memory noise smoothing via lags of shocks.
- If the series is non-stationary, difference the data d times:

$$\text{ARIMA}(p, d, q) \quad \equiv \quad (1 - L)^d y_t.$$

- Example: Random walk = ARIMA$(0, 1, 0)$
  - No AR or MA terms, just differencing once.
  - Shocks accumulate $\Rightarrow$ permanent effects.
  - No mean reversion (contrast with AR(1) where shocks fade).

ARIMA in Practice: GDP Example

- GDP level: looks like a random walk with drift $\Rightarrow$ non-stationary, shocks have permanent effects.
- GDP growth: fluctuations around a stable mean $\Rightarrow$ stationary, mean-reverting.
- ARIMA models bridge the two: difference non-stationary series until they look stationary.



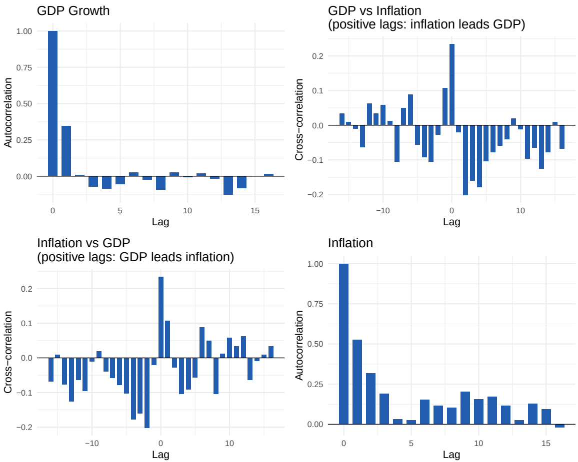
Source: SECO.

Model Diagnostics

- AIC (Akaike Information Criterion): balances fit and complexity, tends to favor richer models (good for forecasting).
- BIC (Bayesian Information Criterion): stronger penalty for complexity, tends to favor more parsimonious models.
- Both start from the log-likelihood ℓ (goodness of fit).
- Then add a penalty for the number of parameters k (model complexity).

From Univariate to Multivariate Models

- So far: forecast one variable from its own lags.
- In macro, variables co-move and predict each other.



Vector Autoregressive Models

Why VARs?

- Until the 1970s, large simultaneous equation models were standard in macro.
- Criticism: these models relied on "incredible assumptions" (exogeneity).
- Vector autoregressions (VARs) emerged as an alternative:
  - Treat all variables as endogenous.
  - Focus on the dynamic effects of structural shocks.

From AR to VAR

- Recall a univariate autoregression:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t.$$

- Extend to a vector of variables:

$$\boldsymbol{y}_t = \begin{bmatrix} \text{GDP} \\ \text{Exchange rate} \\ \text{Inflation} \end{bmatrix}.$$

- Goal: capture their joint dynamics.

Vector Autoregression (VAR)

- A VAR(p) with n variables is

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t,$$

  where

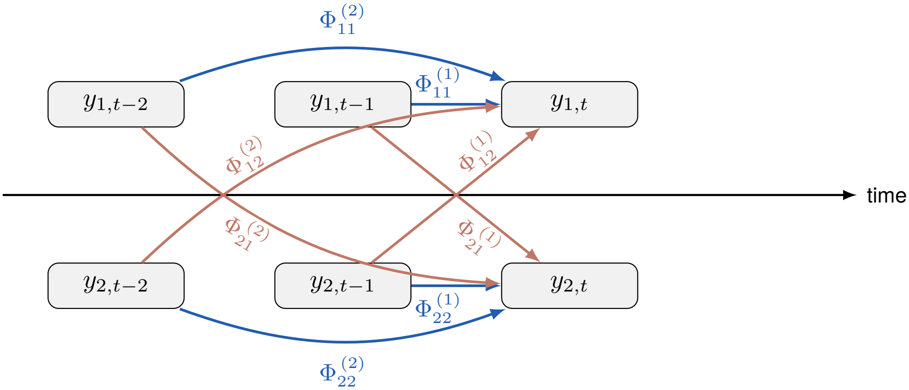
  - $\boldsymbol{y}_t$: $n \times 1$ vector of variables at time t
  - $\boldsymbol{c}$: vector of constants
  - $\boldsymbol{\Phi}_j$: $n \times n$ coefficient matrices ($j = 1, \ldots, p$)
  - $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$: $n \times 1$ error vector with covariance matrix $\boldsymbol{\Omega}$

A VAR(2) Example with $n = 2$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \mathbf{\Phi}_1 \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \mathbf{\Phi}_2 \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

- $y_{1,t}$ depends on its own lags *and* on past $y_{2,t}$ values.
- $y_{2,t}$ depends on its own lags *and* on past $y_{1,t}$ values.
- Intuition: dynamic interactions across variables.
- Example: GDP today depends on past GDP and past inflation; inflation today depends on its own past and past GDP.

VAR(2): Dynamic Interactions (clean labels & arced lag-2)



Stability of VARs

- A process is stationary if its mean and covariance do not depend on time.
- For a VAR, stability requires the eigenvalues of the companion matrix to lie inside the unit circle.
- Intuition: shocks should die out over time, not explode.

Forecasting with VARs

- One-step ahead:
$$\widehat{\mathbf{y}}_{T+1|T} = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_T + \cdots + \mathbf{\Phi}_p \mathbf{y}_{T-p+1}.$$
- Multi-step ahead: use previously forecasted values:
$$\widehat{\mathbf{y}}_{T+h|T} = \mathbf{c} + \mathbf{\Phi}_1 \widehat{\mathbf{y}}_{T+h-1|T} + \cdots + \mathbf{\Phi}_p \widehat{\mathbf{y}}_{T+h-p|T}.$$
- (Bayesian) VAR forecasts widely used in policy institutions for short- to medium-run projections.

Multi-Step Forecasts: VAR(1) — Graphical Intuition

$$\longrightarrow \mathbf{y}_T \xrightarrow{\mathbf{c}\,+\,\Phi\,\mathbf{y}_T} \widehat{\mathbf{y}}_{T+1|T} \xrightarrow{\mathbf{c}\,+\,\Phi\,\widehat{\mathbf{y}}_{T+1|T}} \widehat{\mathbf{y}}_{T+2|T} \xrightarrow{\mathbf{c}\,+\,\Phi\,\widehat{\mathbf{y}}_{T+2|T}} \widehat{\mathbf{y}}_{T+3|T} \longrightarrow \cdots$$

last observed

multi-step recursion

- VAR(1) recursion: $\widehat{\mathbf{y}}_{T+1|T} = \mathbf{c} + \Phi\mathbf{y}_T, \;\; \widehat{\mathbf{y}}_{T+2|T} = \mathbf{c} + \Phi\widehat{\mathbf{y}}_{T+1|T}, \ldots$
- In general: $\widehat{\mathbf{y}}_{T+h|T} = \left(\sum_{j=0}^{h-1} \Phi^j\right)\mathbf{c} + \Phi^h \mathbf{y}_T$.

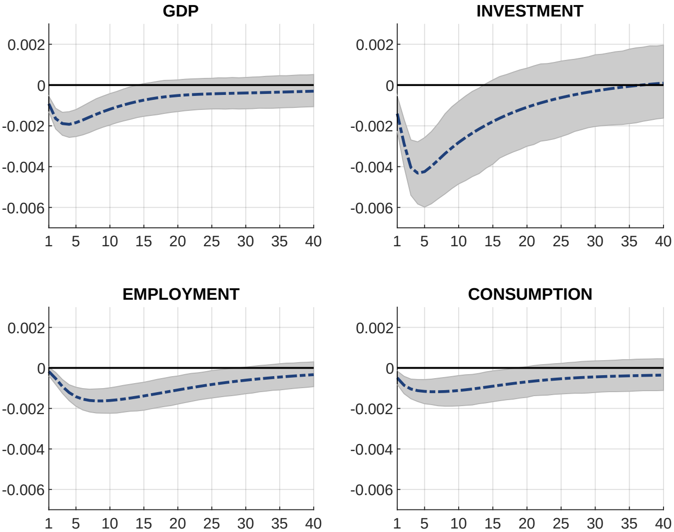
Impulse Response Functions

- A shock today affects variables in future periods through the VAR dynamics.
- Impulse response function (IRF): the path of a variable after a one-unit shock to one equation's innovation.
- Example questions: how does an uncertainty policy shock affect the economy?

Computing IRFs (Intuition)

- Set initial lags (state) and introduce a one-unit shock in one innovation.
- Simulate the system forward with other innovations set to zero.
- The responses over horizon $s = 0, 1, 2, \ldots$ form the IRF sequence.

Impulse Responses to an Uncertainty Shock in G7 Countries



VARs: Shocks and Identification (not covered here)

- VARs are often used to study structural shocks (e.g., monetary, fiscal, uncertainty).
- Identification requires assumptions beyond the VAR (examples: recursive ordering, long-run restrictions, sign restrictions).
- These methods are essential for causal analysis.
- In this course: focus is on forecasting, not structural interpretation. $\rightarrow$ Identification procedures are out of scope.

Further Reading: Lütkepohl, H. & Kilian, L. (2017): *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Introduction to Bayesian Analysis

Why Bayesian Methods?

- Frequentist inference: parameters are fixed, data vary.
- Bayesian inference: parameters are random, data are observed.
- Priors allow us to incorporate expert knowledge or beliefs.
- Bayesian analysis delivers full distributions (posteriors) instead of only point estimates.
- Widely used in modern forecasting: Bayesian VARs, state-space models, etc.

Bayesian Inference

- The joint distribution of data $\boldsymbol{y}$ and parameters $\boldsymbol{\theta}$:

$$p(\boldsymbol{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta})\, p(\boldsymbol{y}|\boldsymbol{\theta}) = p(\boldsymbol{y})\, p(\boldsymbol{\theta}|\boldsymbol{y})$$

- $p(\boldsymbol{\theta})$: prior distribution.
- $p(\boldsymbol{y}|\boldsymbol{\theta})$: likelihood.
- $p(\boldsymbol{\theta}|\boldsymbol{y})$: posterior distribution.
- $p(\boldsymbol{y})$: marginal likelihood.

Bayes' Theorem

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta})\, p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})} \quad \propto \quad p(\boldsymbol{\theta})\, p(\boldsymbol{y}|\boldsymbol{\theta}) \tag{1}$$

- Prior $\times$ Likelihood $\quad \rightarrow \quad$ Posterior.
- Analytical solutions possible with conjugate priors.
- Otherwise, we rely on simulation methods (MCMC).

Linear Regression Model

- Model setup:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_T)$$

- Dimensions:
  - $\mathbf{y}$: $(T \times 1)$ vector, $\mathbf{X}$: $(T \times k)$ matrix, $\boldsymbol{\beta}$: $(k \times 1)$ vector
- Priors will be specified for $(\boldsymbol{\beta}, \sigma^2)$.

Prior Choices in Bayesian Regression

- Noninformative prior (Jeffreys): $p(\beta, \sigma^2) \propto \sigma^{-2}$.
  - Posterior coincides with OLS sampling distribution.
- Conjugate Normal–Gamma prior: $\beta|\sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 V_0),\ \sigma^{-2} \sim \mathcal{G}(a_0, b_0)$.
  - Posterior is also Normal–Gamma (closed form).
- *Note:* More flexible priors (e.g. independent Normal $\times$ Inverse-Gamma) are possible, but break conjugacy $\Rightarrow$ require MCMC methods (covered later).

Posterior in the Linear Regression (non-informative prior)

- Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_T)$$

- With Jeffreys' prior $p(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}$, the posterior is:

$$\boldsymbol{\beta} \mid \sigma^2, \mathbf{y} \sim \mathcal{N}\big(\hat{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\big), \qquad \sigma^2 \mid \mathbf{y} \sim \mathcal{IG}\big(\tfrac{T-k}{2}, \tfrac{SSE}{2}\big)$$

- Where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \qquad SSE = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

- Intuition: Posterior coincides with OLS distribution when prior is flat.

Posterior in the Linear Regression (conjugate prior)

- Conjugate prior: Normal–Inverse-Gamma

$$\beta \mid \sigma^2 \sim \mathcal{N}(\underline{\beta}, \sigma^2 \underline{\mathbf{V}}), \quad \sigma^2 \sim \mathcal{IG}\left(\frac{\underline{\nu}}{2}, \frac{\underline{s}^2}{2}\right)$$

- Posterior is also Normal–Inverse-Gamma:

$$\beta \mid \sigma^2, \mathbf{y} \sim \mathcal{N}(\overline{\beta}, \sigma^2 \overline{\mathbf{V}}), \qquad \sigma^2 \mid \mathbf{y} \sim \mathcal{IG}\left(\frac{\overline{\nu}}{2}, \frac{\overline{s}^2}{2}\right)$$

- Posterior hyperparameters:

$$\overline{\mathbf{V}} = \left(\underline{\mathbf{V}}^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}, \quad \overline{\beta} = \overline{\mathbf{V}}\left(\underline{\mathbf{V}}^{-1}\underline{\beta} + \mathbf{X}'\mathbf{y}\right)$$

$$\overline{\nu} = \underline{\nu} + T, \quad \overline{s}^2 = \underline{s}^2 + SSE + (\hat{\beta} - \underline{\beta})' \left(\underline{\mathbf{V}}^{-1} + \mathbf{X}'\mathbf{X}\right)(\hat{\beta} - \underline{\beta})$$

- Where $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad SSE = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$.
- Interpretation:
  - $\overline{\beta}$ = weighted average of prior mean and OLS estimate.
  - Conjugacy $\Rightarrow$ closed-form posteriors, easy simulation.

Posterior Distributions in Regression

- With noninformative prior (Jeffreys):
  - Posterior for β: Normal around OLS estimate.
  - Posterior for σ^2: Inverse-Gamma around residual variance.
- With conjugate Normal–Inverse-Gamma prior:
  - Posterior remains Normal–Inverse-Gamma (closed form).
  - Prior beliefs update seamlessly with data.
- With non-conjugate priors (e.g. independent Normal × Inverse-Gamma):
  - No closed-form posterior.
  - Motivation: we need simulation methods (MCMC) to approximate the posterior.

Why Simulation Methods?

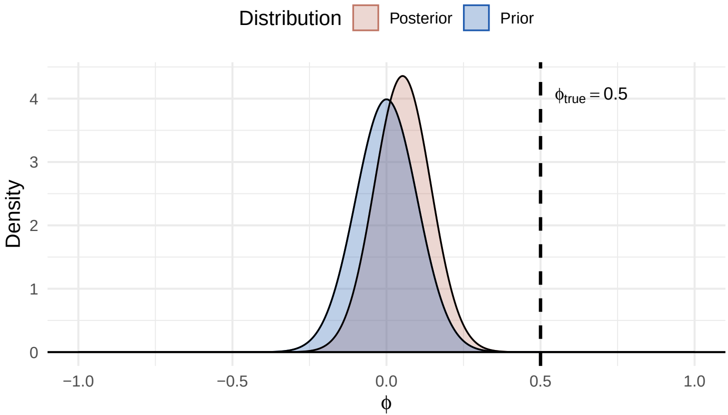
- As we saw, many useful priors (non-conjugate) lead to posteriors without closed form. $\Rightarrow$ We need simulation methods.
- Exact evaluation of the posterior often impossible:
  - High-dimensional parameter spaces.
  - Complex models (VARs, SEMs, state-space).
- Solution: Markov Chain Monte Carlo (MCMC) methods.
- Goal: generate draws from the posterior distribution.
- Key approaches:
  - Gibbs sampling
  - Metropolis–Hastings algorithm

Bayesian Updating: An Illustration

- Before turning to simulation, let's first see how priors and data interact in a simple AR(1) example.
- Suppose we want to estimate the coefficient ϕ of the AR(1) model.
- Prior: reflects beliefs before seeing the data (e.g. $\phi \sim \mathcal{N}(0, \tau^2)$).
- Likelihood: information from the data.
- $\rightarrow$ Posterior: compromise between prior and data.
- $\rightarrow$ Posterior is shifted toward the data when we have many observations.
- $\rightarrow$ With little data, the prior dominates.

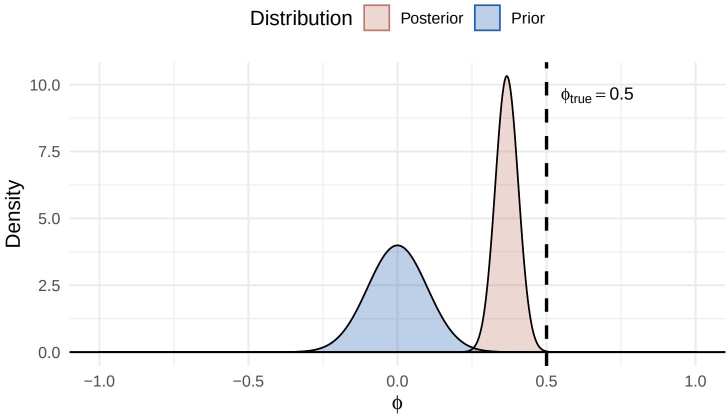
Bayesian Updating for AR(1): Few Observations

- Prior strongly influences the posterior when the sample is small ($n = 20$ and $\tau = 0.1$).



Bayesian Updating for AR(1): Many Observations

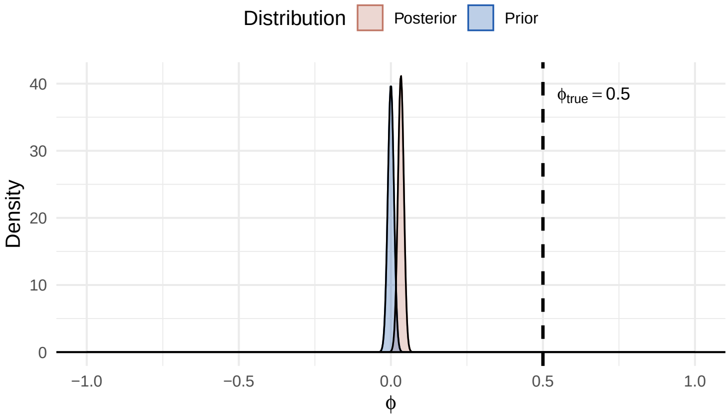
- Prior has little influence when n is large ($n = 500$ and $\tau = 0.1$).



ETH zürich | KOF

Bayesian Updating for AR(1): Tight Prior

- With many observations ($n = 500$), a tight prior ($\tau = 0.01$) still shapes the posterior.



Gibbs Sampling: Illustration

- Suppose parameter vector can be split: $\theta = [\theta_1, \theta_2]$.
- Algorithm:
  1. Draw $\theta_1^{(r)} \sim p(\theta_1 | \theta_2^{(r-1)}, y)$
  2. Draw $\theta_2^{(r)} \sim p(\theta_2 | \theta_1^{(r)}, y)$
- Repeat until convergence.
- Under weak conditions, draws converge to the true posterior distribution.
- In practice: discard burn-in period and keep post-burn-in draws.

Gibbs Sampling: Illustration with AR(1)

- Now we return to the AR(1) model, this time to illustrate how Gibbs sampling generates draws from the posterior.
- Stack data: $\mathbf{x} = (y_1, \ldots, y_{T-1})'$, $\mathbf{y} = (y_2, \ldots, y_T)'$.
- Model: $\mathbf{y} = \phi\,\mathbf{x} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\, \sigma^2 \mathbf{I}_{T-1})$.
- Priors (independent): $\phi \sim \mathcal{N}(\mu_0, \tau_0^2)$, $\sigma^2 \sim \mathcal{IG}(a_0, b_0)$.
- Full conditionals:

$$\phi \mid \sigma^2, \mathbf{y} \sim \mathcal{N}\Big(\mu_n,\ \tau_n^2\Big),$$

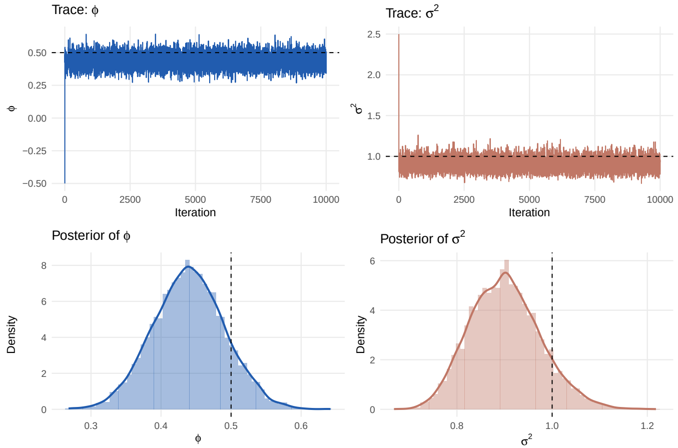
$$\tau_n^2 = \left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}\mathbf{x}'\mathbf{x}\right)^{-1}, \qquad \mu_n = \tau_n^2\left(\frac{\mu_0}{\tau_0^2} + \frac{1}{\sigma^2}\mathbf{x}'\mathbf{y}\right),$$

$$\sigma^2 \mid \phi, \mathbf{y}\ \sim\ \mathcal{IG}\Big(a_0 + \frac{T-1}{2},\ \ b_0 + \frac{1}{2}\,(\mathbf{y} - \phi\mathbf{x})'(\mathbf{y} - \phi\mathbf{x})\Big).$$

- Gibbs steps: sample ϕ from the Normal, then σ^2 from the Inverse-Gamma; iterate.

Gibbs Sampling for AR(1): Convergence & Posterior

- After burn-in, chains mix well and posteriors concentrate near the truth.



ETH zürich | KOF

Simulation as the Common Tool

- Simulation methods (MCMC) are the backbone of modern Bayesian econometrics.
- Applications in this course:
  - Bayesian VARs (BVARs)
  - Simultaneous Equation Models (SEMs)
  - State-Space Models
- We will primarily use the Gibbs sampler.

Forecast Evaluation

Why Evaluate Forecasts?

- Policymakers need reliable point and density forecasts.
- Evaluation disciplines model choice and hyperparameters (e.g., BVAR shrinkage).
- We must mimic real use: real-time and horizon-specific.

Real-Time Evaluation Protocol

At each vintage t:

1. Use information set $\mathcal{I}_t$ (data available at t) to re-estimate the model.
2. Produce h-step ahead forecasts: point $\hat{y}_{t+h|t}$ and (if available) predictive density $p(y_{t+h}|\mathcal{I}_t)$.
3. Store forecasts. When y_{t+h} is released, compute loss/score.

Aggregate by horizon h: for a set $\mathcal{T}_h$, compute average loss/score

$$\frac{1}{|\mathcal{T}_h|}\sum_{t\in\mathcal{T}_h}\ell\big(y_{t+h},\hat{y}_{t+h|t}\big) \quad \text{or} \quad \frac{1}{|\mathcal{T}_h|}\sum_{t\in\mathcal{T}_h}s(p(\cdot|\mathcal{I}_t);y_{t+h})\,.$$

(Use expanding or rolling window; keep the same vintages across models.)

Point Forecasts: RMSE and MAE

Definitions: RMSE = Root Mean Squared Error, MAE = Mean Absolute Error

Errors and losses

$$e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}, \qquad \text{SE}_{t+h|t} = e^2_{t+h|t}, \qquad \text{AE}_{t+h|t} = |e_{t+h|t}|.$$

Aggregate over real-time vintages $\mathcal{T}_h$:

$$\mathrm{RMSE}(h) = \sqrt{\frac{1}{|\mathcal{T}_h|} \sum_{t \in \mathcal{T}_h} e^2_{t+h|t}}, \qquad \mathrm{MAE}(h) = \frac{1}{|\mathcal{T}_h|} \sum_{t \in \mathcal{T}_h} |e_{t+h|t}|.$$

When to prefer which?

- RMSE penalizes large errors more (quadratic loss); sensitive to outliers.
- MAE is robust; aligns with median forecasts.

Computing Evaluation Metrics in Real Time (Step-by-Step)

1. Build a vintage loop: $t = t_0, \ldots, t_1$.
2. At each t, construct $\mathcal{I}_t$ (respect data release lags/revisions).
3. Refit model on $\mathcal{I}_t$; compute forecasts $\hat{y}_{t+h|t}$ (and densities if applicable).
4. When y_{t+h} becomes available, append the chosen loss/score to horizon-h lists.
5. After the loop, report average metric (e.g., $\mathrm{RMSE}(h)$ or $\mathrm{MAE}(h)$) per horizon h.

Density Forecasts: Log Score

Definition (proper scoring rule)

$$\mathrm{LS}(h) = \frac{1}{|\mathcal{T}_h|} \sum_{t \in \mathcal{T}_h} \log p(y_{t+h} \mid \mathcal{I}_t) \, .$$

Higher is better. Rewards calibrated, sharp predictive densities.

Gaussian predictive density $y_{t+h} \mid \mathcal{I}_t \sim N(\mu_{t+h|t}, \sigma^2_{t+h|t})$:

$$\log p(y_{t+h}|\mathcal{I}_t) = -\tfrac{1}{2} \left[\log(2\pi\sigma^2_{t+h|t}) + \frac{(y_{t+h} - \mu_{t+h|t})^2}{\sigma^2_{t+h|t}} \right] .$$

- Alternative: CRPS (Continuous Ranked Probability Score), interpretable as a distributional analogue of MAE (lower is better).

  Note: if the forecast is a single point, CRPS reduces exactly to MAE.

Practical Checklist

- Real-time discipline: use true vintages; respect publication lags/revisions.
- Same information set across models at each t.
- Horizon-specific metrics (don't mix $h = 1$ with $h = 4$).
- Always compare against a benchmark model (e.g. random walk, AR(1), simple VAR).
- Document window choice (expanding vs rolling), start date, sample length.

Forecast Combination

Why Combine Forecasts?

- No single model is best at all times, horizons, or regimes.
- Combining forecasts can:
  - reduce model risk (hedge against misspecification),
  - stabilize performance across horizons,
  - often improve accuracy out-of-sample.
- Empirical regularity: simple combinations (e.g., equal weights) often work remarkably well.

Setup

- Suppose we have M competing forecasts for y_{t+h}:

$$\hat{y}_{t+h|t}^{(m)}, \quad m = 1, \ldots, M.$$

- A linear combination forecast:

$$\hat{y}_{t+h|t}^{c} = \sum_{m=1}^{M} w_m \hat{y}_{t+h|t}^{(m)},$$

  with weights w_m (typically $\sum_m w_m = 1$).
- Goal: choose weights to improve forecast accuracy.

Weighting Schemes

- Equal Weights
  - $w_m = 1/M$ for all m.
  - Simple, robust benchmark; often surprisingly hard to beat.
- Variance–Covariance Weights
  - Use past forecast errors to estimate error covariance matrix Σ.
  - Down-weight noisy or highly correlated forecasts.
- Regression-Based Weights
  - Estimate weights by regressing y_{t+h} on competing forecasts.
  - Flexible, adapts to relative model performance.

Variance–Covariance Method: Bivariate Case I

- Combine two unbiased forecasts:

$$\hat{y}^c = \omega \hat{y}^a + (1-\omega)\hat{y}^b, \quad \omega \in [0,1].$$

- Forecast errors:

$$e^c = \omega e^a + (1-\omega)e^b, \quad e^m = y - \hat{y}^m.$$

- If e^a and e^b are uncorrelated:

$$\sigma_c^2 = \omega^2 \sigma_a^2 + (1-\omega)^2 \sigma_b^2.$$

Variance–Covariance Method: Bivariate Case II

- Minimizing σ_c^2 w.r.t. ω gives:

$$\omega^\star = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2}.$$

- Intuition: more weight goes to the forecast with lower error variance.
- Allowing for correlated errors ($\sigma_{ab} \neq 0$):

$$\omega^\star = \frac{\sigma_b^2 - \sigma_{ab}}{\sigma_a^2 + \sigma_b^2 - 2\sigma_{ab}}.$$

- Positive correlation reduces the benefit of combining forecasts.

Variance–Covariance Method: General Case

- With K forecasts, let Σ be the $K \times K$ error covariance matrix.
- Problem: choose weights ω to minimize forecast error variance

$$\min_{\omega} \; \omega' \Sigma \omega \quad \text{s.t. } \mathbf{1}'\omega = 1.$$

- Solution:

$$\omega^{\star} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}.$$

- Special case (diagonal Σ): inverse-variance weighting.
- Intuition: down-weight noisy and highly correlated forecasts.

Regression-Based Combination

- Idea: use historical data to let the data "speak" about optimal weights.
- Regress realizations on competing forecasts:

$$y_{t+h} = \alpha + \sum_{m=1}^{M} w_m \hat{y}^{(m)}_{t+h|t} + u_{t+h}.$$

- Estimate weights w_m by OLS on a real-time evaluation sample.
- Special cases:
  - If $\alpha = 0$ and $\sum_m w_m = 1$, the solution nests the variance–covariance method.
  - Without constraints, OLS can assign negative weights or weights summing $\neq 1$.

Toy Example: Combining Forecasts

Suppose we have 4 real-time forecasts for GDP growth one-quarter ahead:

Vintage t	Realized y_{t+1}	AR $\hat{y}^{AR}_{t+1\|t}$	BVAR $\hat{y}^{BVAR}_{t+1\|t}$
1	1.2	1.5	1.0
2	0.8	1.1	0.7
3	1.0	1.4	0.9
4	0.5	0.9	0.6

With $e^m = y - \hat{y}^m$:

$$e^{AR} = (-0.3, -0.3, -0.4, -0.4), \qquad e^{BVAR} = (0.2, 0.1, 0.1, -0.1).$$

Toy Example: Evaluation

Compute RMSE:

$$\mathrm{RMSE}_{AR} = \sqrt{\frac{0.3^2+0.3^2+0.4^2+0.4^2}{4}} \approx 0.35, \qquad \mathrm{RMSE}_{BVAR} = \sqrt{\frac{0.2^2+0.1^2+0.1^2+0.1^2}{4}} \approx 0.13.$$

Equal-weight combination:

$$\hat{y}^{comb} = 0.5\,\hat{y}^{AR} + 0.5\,\hat{y}^{BVAR},$$

$$e^{comb} = (-0.05, -0.10, -0.15, -0.25) \Rightarrow \mathrm{RMSE} \approx 0.16.$$

Result: combination improves AR substantially; close to BVAR.

Toy Example: Regression Weights

Regression yields weights near BVAR, e.g.

$$y_{t+1} = 0.3\,\hat{y}^{AR}_{t+1|t} + 0.7\,\hat{y}^{BVAR}_{t+1|t}.$$

With these weights, $\mathrm{RMSE} \approx 0.10$ (better than either model alone).

Takeaway: Regression combination can tilt toward the stronger model and improve accuracy.

Bayesian Model Averaging (BMA)

- Idea: treat the *model itself* as random. Each model has a posterior probability given the information set.
- Let $\mathcal{I}_t$ = information set at time t (data observed up to t).
- Posterior model probability:

$$P(M_k|\mathcal{I}_t) = \frac{P(\mathcal{I}_t|M_k)\,P(M_k)}{\sum_{j=1}^{K} P(\mathcal{I}_t|M_j)\,P(M_j)}.$$

- Combined predictive density:

$$p(y_{t+h}|\mathcal{I}_t) = \sum_{j=1}^{K} P(M_j|\mathcal{I}_t)\,p(y_{t+h}|\mathcal{I}_t, M_j).$$

- Intuition:
  - Forecasts are averaged using posterior model probabilities as weights.
  - Naturally accounts for parameter and model uncertainty.
- Practical note: elegant but computationally heavier than equal weights or regression.

BMA: Illustration

Suppose we have three competing models for inflation one-quarter ahead. Posterior probabilities are computed given the information set $\mathcal{I}_t$ (all data up to t):

Model	Posterior Probability $P(M_k \mid \mathcal{I}_t)$
AR(1)	0.20
BVAR	0.30
SEM	0.50

Combined predictive density:

$$p(y_{t+1} \mid \mathcal{I}_t) = 0.2\, p(y_{t+1} \mid \mathcal{I}_t, \text{AR(1)}) \; + \; 0.3\, p(y_{t+1} \mid \mathcal{I}_t, \text{BVAR}) \; + \; 0.5\, p(y_{t+1} \mid \mathcal{I}_t, \text{SEM}).$$

Takeaway: Models are averaged with probabilities reflecting how well each explains the data.

Forecast Combination: Practical Notes

- Equal weights: robust benchmark.
- Variance–covariance: optimal if Σ estimated well, but unstable with few vintages.
- Regression method: adapts to relative performance, but risk of overfitting with short evaluation samples.
- BMA: theoretically elegant; computationally intensive for many models.
- Always compare to the simple equal-weight average.