COURSE HANDBOOK

Nowcasting with High-Frequency Data

Nowcasting with High-Frequency Data | Collected sessions

Thursday, 30 July 2026

Contents of this handbook

PART I

Session 1 · The Nowcasting Problem, State Space Models & the Kalman Filter

---

Suppose it is 8:29 on a release morning. Our current estimate of this quarter's GDP growth is 2.0 percent. At 8:30 a new indicator arrives below the model's expectation. The number is relevant, but it is noisy. How far should the GDP estimate move? How much narrower should its uncertainty band become? And how can we explain the revision afterward?

Those are the questions for this session. The Kalman filter supplies the calculation, but we will not begin with the general matrix formula. We will first work through one release in a scalar model. The matrix version will then be a generalization of a calculation we already understand.

By the end of the session, you should be able to:

1. distinguish the period an observation describes from the date on which it becomes available;
2. compute a one-observation Kalman update and interpret the innovation, its variance, and the gain;
3. write a linear Gaussian state space model and run its prediction and update steps;
4. explain how missing observations and mixed frequencies enter a real-time data vintage; and
5. turn a filtered state into a quarterly GDP nowcast, an uncertainty band, and a release-by-release news decomposition.

The session assumes familiarity with matrix multiplication, transposes, variances and covariances, conditional expectation, and the multivariate normal distribution. The Gaussian conditioning result used to derive the filter is stated when we need it.

1 GDP Publication Lags, Revisions, and Nowcasting

1.1 Why GDP Must Be Estimated Before Its First Release

Real GDP is a quarterly measure. In the United States, the advance estimate is released about four weeks after a quarter ends and is revised afterward. By the middle of a quarter, the latest official GDP observation therefore describes an economy that is already several months old.

The gap matters because other decisions do not wait for the national accounts. Central banks meet, firms change production plans, and markets reprice assets while the current quarter is still under way. During that interval, evidence arrives from sources such as employment reports, industrial production, retail sales, weekly unemployment claims, electricity use, and payment data. Each indicator covers only part of the economy, and each contains noise, but together they say something about current activity.

Nowcasting is the statistical problem of using those partial measurements to estimate the present (Giannone et al., 2008). If y_Q is GDP growth in quarter Q, and Ω_r is the information available on release date r, the nowcast is

$$\widehat{y}_{Q|r} = \mathbb{E}[y_Q \mid \Omega_r]. \tag{1}$$

The subscript r matters. A nowcast is always attached to an information date. "Our estimate for 2027Q1" is incomplete unless we also say whether it was made in January, March, or after the quarter ended but before GDP was released.

The terminology follows the target's position relative to the release calendar. A **forecast** concerns a future period. A **nowcast** concerns a period under way. A **backcast** concerns a completed period whose official value has not yet been released. In practice, the same model produces all three. Only the horizon changes.

That classification tells us *which* period is being estimated. It does not yet say what the forecaster was allowed to know. For that we need a second clock.

1.2 Reference Periods, Release Dates, and Data Vintages

Real-time macroeconomic data have two dates:

- The **reference period**, written τ, is the week, month, or quarter the observation describes.
- The **release date**, written r, is the date on which that observation or revision enters the forecaster's information set.

For example, a retail-sales number may describe July, first appear in August, and be revised in September. It belongs to July in the economic model, but it must be absent from every data **vintage** constructed before its August release. Neither date can replace the other.

We write $x_{i,\tau}^{(r)}$ for the value of series i, referring to period τ, as it appeared in vintage r. Table 1 records a small example.

Table 1. An illustrative vintage ledger. In a row such as retail sales, holding the series $i = \text{RS}$ and reference period $\tau = \text{Jul}$ fixed while moving from r_0 to r_2 traces $x_{i,\tau}^{(r)}$ across release dates.

Series and reference period	Model notation	r_0: before release	r_1: first release	r_2: later vintage
July retail sales	$x_{\text{RS,Jul}}^{(r)}$	missing	0.4%	0.2%
July industrial production	$x_{\text{IP,Jul}}^{(r)}$	missing	missing	−0.1%
Previous-quarter GDP	$x_{\text{GDP},Q-1}^{(r)}$	−1.0%	−1.0%	−0.8%

Moving from one column to the next can do two things. A previously empty cell can receive its first value, and an older value can be revised. Both events change Ω_r. A production nowcasting system therefore stores the reference period, release timestamp, vintage value, and transformation applied to each observation.

This distinction also clarifies real-time evaluation. Using predictor values that were unavailable on date r is look-ahead bias[1]. The value used later to score the nowcast is a separate choice. We may score against GDP's first release if the objective is to predict the initial print, or against a later vintage if the objective is to estimate activity as eventually measured. Both targets can be informative, but they answer different questions.

Once the two clocks are separate, we can introduce the objects that carry information through the model. The next table is a map; the prose and examples that follow will give each symbol a job.

1.3 Notation Reference for Data Vintages, States, and Updates

Table 2 is a reference for the notation used across the session. Each symbol is introduced again where it enters the argument; there is no need to memorize the table before continuing.

Table 2. Notation used throughout Session 1.

Symbol	Meaning
τ	Reference period of a state or observation

[1] Look-ahead bias occurs when information from after a forecast origin leaks into model fitting, transformation, variable selection, or prediction. It makes historical performance look better than a real-time forecaster could have achieved.

Symbol	Meaning
r	Release date, and therefore the date defining a data vintage
$x_{i,\tau}^{(r)}$	Value of series i for reference period τ as available in vintage r
$Y_s^{(r)}$	Observations through reference period s that are available in vintage r
$a_{\tau\|s}^{(r)},\ P_{\tau\|s}^{(r)}$	Conditional mean and covariance of state α_τ given $Y_s^{(r)}$
$W_\tau^{(r)}$	Matrix selecting the observations available for period τ in vintage r
T, R, Z	State transition, state-shock loading, and measurement loading matrices
Q, H	State-shock and measurement-error covariance matrices
$Z_\tau^{(r)},\ H_\tau^{(r)}$	Measurement matrix and error covariance after selecting the observed rows
$T(Q),\ Z_Q$	Final monthly reference period for quarter Q and the row mapping the state into quarterly GDP
v_τ, F_τ, K_τ	Innovation, innovation covariance, and Kalman gain
$q, \pi_\tau, \bar{\pi}, \bar{K}$	Signal-to-noise ratio, scaled predicted variance, and their steady-state values in the local level model
J_τ	Smoother gain mapping later information back to state α_τ
$\tilde{v}_\tau$	Standardized innovation used for model diagnostics

The central object is the **state** α_τ: a compact summary of the model at reference period τ. The quantity $a_{\tau|s}^{(r)}$ is its conditional mean[2], and $P_{\tau|s}^{(r)}$ is its conditional covariance[3]. The bar in $\tau \mid s$ can be read as "state dated τ, given information through s."

[2]A conditional mean is an average taken after restricting attention to what is known in a stated information set. Here it is the model's minimum-mean-squared-error point estimate of the state given $Y_s^{(r)}$.

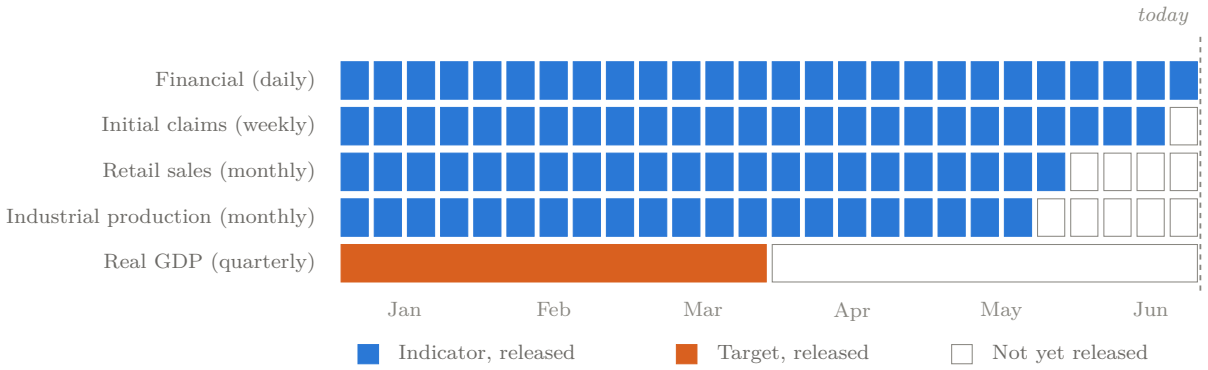
[3]A covariance measures how two random quantities vary together. A covariance matrix places the variance of each state component on its diagonal and pairwise covariances off the diagonal.

We now turn from the ledger to the shape it creates in a panel. Because series do not share a release schedule, the last observed reference period is not the same in every row.

1.4 Asynchronous Release Schedules and the Ragged Edge

Series are released on different schedules. On a given date, one monthly series may be available through June, another through July, and weekly claims through the previous Saturday. The lower-right corner of a data matrix is consequently incomplete. This is the **ragged edge** (Bańbura et al., 2011). Figure 1 shows the resulting staircase in one data vintage.

Figure 1. The ragged edge. The figure shows one vintage. Rows are ordered by publication lag, so their last available observations do not line up. GDP for the current quarter is still missing.



Three separate difficulties are visible here. The indicators have different frequencies, their latest observations arrive at different times, and their published histories can be revised. A statistical model can handle the first two once the vintage has been assembled correctly. It cannot recover release dates that were lost during data collection.

The data problem is now clear: information is partial, asynchronous, and revisable. To see how a model should react to one new piece of that information, we temporarily set the large panel aside and study the smallest useful update.

2 Scalar Kalman Updates in the Local Level Model

2.1 State Variation Versus Measurement Error

Before using a panel of indicators, consider one **latent** measure of current activity, α_τ, observed through one noisy indicator, y_τ:

$$\begin{aligned} \alpha_{\tau+1} &= \alpha_\tau + \eta_\tau, \quad \eta_\tau \sim \mathcal{N}(0, Q), \\ y_\tau &= \alpha_\tau + \varepsilon_\tau, \quad \varepsilon_\tau \sim \mathcal{N}(0, H). \end{aligned} \tag{2}$$

This is the **local level model**. The state follows a random walk[4]. Its **shock** η_τ allows economic conditions to change between periods. The **measurement error** ε_τ represents the part of the indicator that does not measure aggregate activity: sampling error, sector-specific movements, and other idiosyncratic variation.

The distinction between Q and H is central:

- A large Q says the underlying economy can change quickly.
- A large H says the indicator is a noisy reading of that economy.

Both make observations volatile, but they imply different responses. If Q is large, recent data deserve more weight because the state may have moved. If H is large, any single observation deserves less weight because much of its movement may be measurement noise.

Figure 2 separates the two sources of volatility. Increasing Q makes the unobserved path itself less stable, so an older estimate becomes stale more quickly. Increasing H spreads observations farther around a given path, so a single reading is less trustworthy.

The distinction tells us qualitatively how much attention a release deserves. The next example turns that intuition into a numerical update.

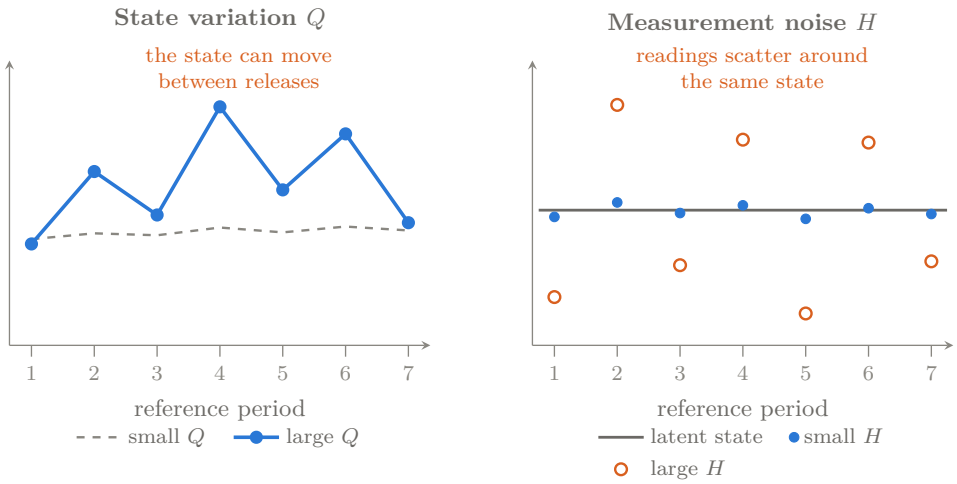
2.2 Numerical Example: State Update After an Indicator Release

Return to the release morning. To keep the units interpretable, suppose the indicator has already been transformed into annualized[5] GDP-growth units. Just before the release, the model's distribution for current activity is

[4] A random walk sets the next level equal to the current level plus a new mean-zero disturbance. Its changes can be stationary even though the level is not, and its uncertainty grows as the prediction horizon lengthens.

[5] An annualized rate reports the rate that would obtain over a year if the within-year pace continued. For quarter-on-quarter growth, the exact conversion is $(1 + g_q)^4 - 1$; multiplying by four is a common small-rate approximation.

Figure 2. Two reasons an indicator can move. A larger Q means genuine state movement and raises the value of recent data. A larger H means a noisier reading of a given state and lowers the value of any one observation. The observed series alone cannot identify which explanation is correct without a model.



$$\alpha \mid \Omega_{r^-} \sim \mathcal{N}(a, P), \qquad a = 2.0, \quad P = 0.25.$$

Read this statement as follows: conditional on everything known just before the release, α is normally distributed with center $a = 2.0$ and variance $P = 0.25$. The expression is a probability model for what the latent activity rate could be, not a claim that the activity rate itself is observed.

This scaling is only for the hand calculation. In an empirical model, the measurement loading Z maps an indicator in its own units into the units of the state.

The prior mean[6] is 2.0 percent, and its standard deviation[7] is $\sqrt{0.25} = 0.5$ percentage points. The new indicator reads $y = 1.2$. Its measurement-error variance is $H = 0.75$.

The calculation has three parts. First compute the **innovation**, the observed release minus the model's expectation:

$$v = y - a = 1.2 - 2.0 = -0.8. \tag{3}$$

Next compute the **innovation variance** and the scalar **Kalman gain**:

[6]The prior mean is the model's mean estimate immediately before the new observation is incorporated; "prior" here means prior to this update, not necessarily a subjective belief.

[7]A standard deviation is the square root of a variance and is therefore expressed in the same units as the quantity being measured.

$$
\begin{aligned}
F &= P + H = 1.00, \\
K &= \frac{P}{P + H} = 0.25.
\end{aligned}
\tag{4}
$$

Finally update the state estimate and its variance:

$$
\begin{aligned}
a^{+} &= a + Kv = 2.0 + 0.25(-0.8) = 1.8, \\
P^{+} &= (1 - K)P = 0.75(0.25) = 0.1875.
\end{aligned}
\tag{5}
$$

The indicator came in 0.8 percentage points below expectation. The model passes one quarter of that **surprise** into the state estimate, so the nowcast falls by 0.2 percentage points. The posterior standard deviation falls from 0.50 to $\sqrt{0.1875} \approx 0.43$ because the release provided information.

This example is intentionally small, but it contains the three quantities that will recur throughout the course. Table 3 collects their numerical values and interpretations.

Table 3. The three objects in a Kalman update.

Quantity	Question it answers	Value in the example
Innovation v	How different was the release from the model's expectation?	-0.8
Innovation variance F	How uncertain was that surprise expected to be?	1.00
Gain K	How much of the surprise should revise the state?	0.25

Two qualifications prevent common misreadings. First, K is a weight on the surprise, not on the level of the released number. Second, a release equal to its expectation has $v = 0$, so it does not move the point estimate[8], but it can still reduce uncertainty. Learning that an expected value actually occurred is information.

An update answers what the new release tells us about the current state. Between releases, however, the state itself continues to evolve. That requires the other half of the recursion: prediction.

[8] A point estimate is one reported value, such as the conditional mean, used to summarize an entire estimated distribution. It should be accompanied by an uncertainty measure when decisions depend on its precision.

2.3 State Prediction and Variance Growth Between Releases

After the update, the state moves to the next reference period according to Equation 2. Conditional on current information,

$$a_{\tau+1|\tau} = a_{\tau|\tau}, \qquad P_{\tau+1|\tau} = P_{\tau|\tau} + Q. \tag{6}$$

The mean remains unchanged because the random-walk shock has mean zero. The variance increases by Q because the next state can move before it is observed. Prediction and updating therefore have different effects: prediction adds uncertainty about state evolution; an informative observation removes some uncertainty.

Equation Equation 6 is read one line at a time. The first equality says that today's best level is also tomorrow's best level because the expected shock is zero. The second says that tomorrow's uncertainty equals the uncertainty left after today's update plus the variance of the as-yet unseen shock. The conditioning subscript $\tau + 1 \mid \tau$ means "about period $\tau + 1$, using data only through period τ."

For a fixed target date, the mere passage of calendar time does not reveal new information. Suppose no data arrive between Monday and Tuesday. Monday's forecast distribution for a Friday target already allowed for the state shocks that might occur before Friday. Advancing the model one day increases the current-state variance and shortens the remaining horizon by one step; those effects offset. The target distribution changes when the information set changes, not because the clock moved.

2.4 The Signal-to-Noise Ratio and the Steady-State Kalman Gain

The numerical example used one choice of Q and H. To understand the whole family of choices, it is convenient to reduce the two variances to one relative measure. The **signal-to-noise ratio** compares variation in the signal[9] with variation in the noise[10]:

$$q = \frac{Q}{H}.$$

Large q means that changes in the state are large relative to measurement noise. Small q means that observations are noisy relative to the pace at which the state moves. Writing the predicted variance as

[9] A signal is the model component that carries information about the object of interest; in the local level model it is the evolving state α_τ.

[10] Noise is variation that obscures the object of interest. Here it is the measurement error ε_τ.

$\pi_\tau = P_{\tau|\tau-1}/H$, its recursion[11] is

$$\pi_{\tau+1} = \frac{\pi_\tau}{\pi_\tau + 1} + q.$$

At a steady state[12], set $\pi_{\tau+1} = \pi_\tau = \bar{\pi}$. Then

$$\bar{\pi} = \frac{\bar{\pi}}{\bar{\pi}+1} + q \quad \Longrightarrow \quad \bar{\pi}(\bar{\pi}+1) = \bar{\pi} + q(\bar{\pi}+1) \quad \Longrightarrow \quad \bar{\pi}^2 - q\bar{\pi} - q = 0.$$

The quadratic formula gives one negative root and the positive root

$$\bar{\pi} = \frac{q + \sqrt{q^2+4q}}{2}, \qquad \bar{K} = \frac{\bar{\pi}}{\bar{\pi}+1}. \tag{7}$$

Why does the sequence reach that root? Let $g(\pi) = \pi/(\pi+1) + q$. For $\pi \geq 0$, the map is increasing and

$$0 < g'(\pi) = \frac{1}{(\pi+1)^2} \leq 1.$$

For $q > 0$, the inequality is strict after the first iteration. Moreover, $g(\pi) - \pi$ is positive below $\bar{\pi}$ and negative above it. An iterate therefore moves toward $\bar{\pi}$ without crossing it, and each subsequent gap shrinks. The variance recursion converges to the unique nonnegative fixed point, so the gain converges to $\bar{K}$.

When q is large, $\bar{K}$ is near one and the estimate responds strongly to new observations. When q is small, the filter averages across more observations. These statements are exact for the scalar local level model. In a multivariate system, gains can be positive or negative and depend on the full covariance structure, so it is safer to interpret them one release and one target at a time.

Figure 3 plots the steady-state gain against q. The four marked values correspond to Exercise 4.

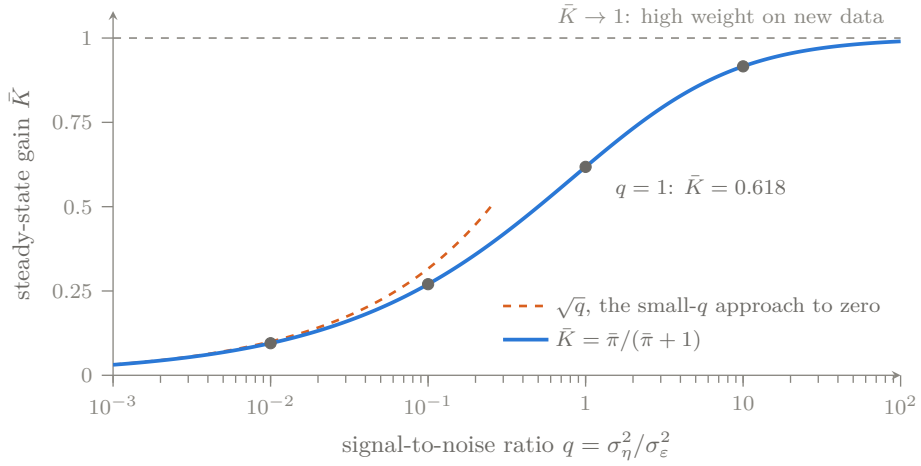
2.5 Missing Observations: Prediction Without a Measurement Update

The steady-state calculation presumes regular observations. A release calendar immediately breaks that rhythm, so we next ask what the filter does when no measurement is available.

[11] A recursion defines the next value from the current one. Once an initial value is supplied, repeated application generates the entire sequence.

[12] A steady state of a recursion is a value left unchanged by another application of its updating map. It concerns the convergence of the variance and gain here, not a claim that the realized economy stops moving.

Figure 3. Signal, noise, and the steady-state gain. A larger Q/H gives new observations more influence. The marked values correspond to Exercise 4.



If y_τ is missing, there is no innovation to compute. The update is skipped, and the model carries its predicted distribution forward. In the local level model, each missing period adds another Q to the state variance.

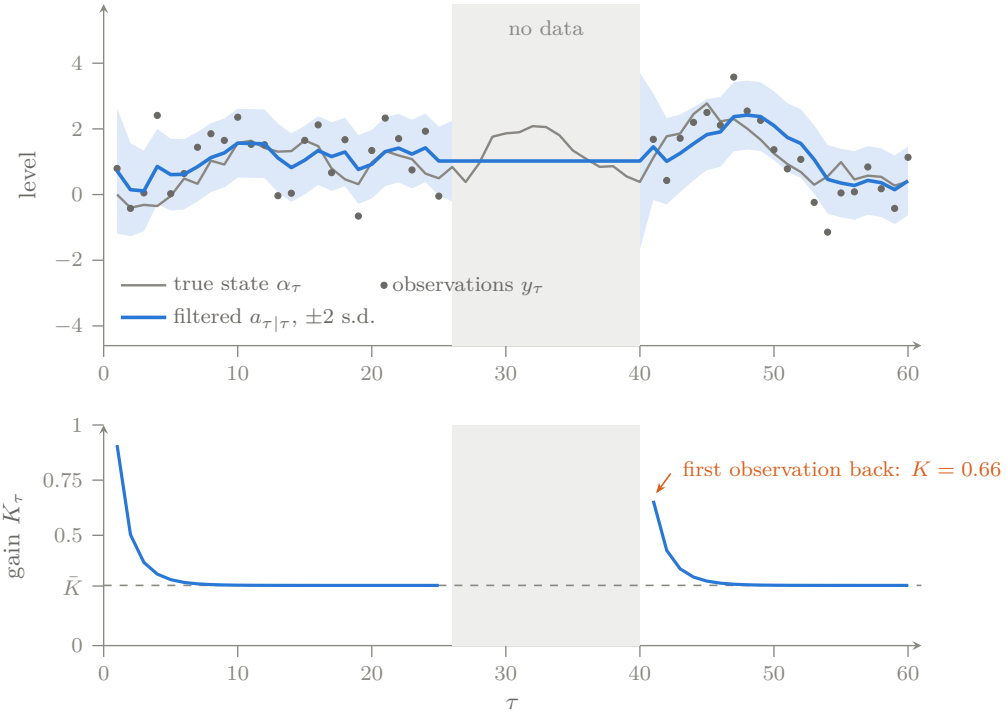
Figure 4 shows the consequences of a run of missing observations: the point estimate remains at its last updated value, uncertainty grows, and the first observation after the gap receives a larger gain.

The outage is a useful introduction to missing data, but it is not yet a full ragged edge. A real panel contains many series, and a release may describe a period earlier than the date on which it arrives. We now generalize the model while keeping those two clocks separate.

3 Linear Gaussian State-Space Models

The local level model answered one release-morning question with one state and one indicator. A production nowcast must keep track of several latent forces, their lags, and many observations at once. We therefore need notation that scales without changing the logic of the scalar calculation. This section defines that notation and states the conditions under which it has the probabilistic interpretation used later.

Figure 4. A local level model during a data outage. Observations are withheld in the shaded interval. The state estimate stays at its last value while uncertainty grows. Once observations resume, the first update receives more weight because the prior is less precise.



3.1 State and Measurement Equations: Matrices and Assumptions

A linear Gaussian[13] state space model has two equations. The **state equation** describes how an $m \times 1$ latent state vector evolves across reference periods:

$$\alpha_{\tau+1} = T\alpha_\tau + R\eta_\tau, \qquad \eta_\tau \sim \mathcal{N}(0, Q). \tag{8}$$

Read Equation 8 from right to left. Start with the current state α_τ, carry its predictable part forward through T, and add the new disturbance η_τ through the loading matrix R. Thus T describes persistence and cross-state dynamics, while RQR' describes the uncertainty introduced on the way to $\tau + 1$.

The **measurement equation** links a $p \times 1$ observation vector to the state:

$$y_\tau = Z\alpha_\tau + \varepsilon_\tau, \qquad \varepsilon_\tau \sim \mathcal{N}(0, H). \tag{9}$$

[13] "Linear" means the state and observations enter through matrix multiplication and addition; "Gaussian" means the initial state and disturbances are normally distributed. Together they make every joint and conditional distribution in the filter Gaussian.

Read Equation 9 as a measurement recipe. The row structure of Z converts the same state α_τ into the model's expected value for each observed series; ε_τ is what remains series-specific. The state equation moves through time. The measurement equation moves from an unobserved state to data at a given time.

Here T governs state dynamics, R loads shocks into the state, and Z maps the state into observed series. The covariance matrices Q and H describe state and measurement noise. Intercepts and deterministic regressors[14] can be added to either equation; we omit them to keep the filtering notation readable. Any system matrix may also vary with τ.

The assumptions used in this session are:

A1 (State noise). The sequence $\{\eta_\tau\}$ is independent over time. This makes each η_τ genuinely new state information, so the prediction variance needs only the current Q. If state shocks are serially correlated, the state must usually be enlarged to model that dependence; otherwise the filter understates what is predictable.

A2 (Measurement noise). The sequence $\{\varepsilon_\tau\}$ is independent over time. This says that any persistence in an observed series is explained by the state rather than by leftover measurement error. If the errors are serially correlated, their dynamics should be modeled as additional states or the innovations will remain forecastable.

A3 (Independence). State shocks, measurement errors, and the initial state are mutually independent at all leads and lags. This removes cross-terms such as $\mathrm{Cov}(R\eta_\tau, \varepsilon_{\tau+1})$ from the covariance blocks used below. If economic shocks also alter reporting error, the standard gain formula is incomplete unless that covariance is included explicitly.

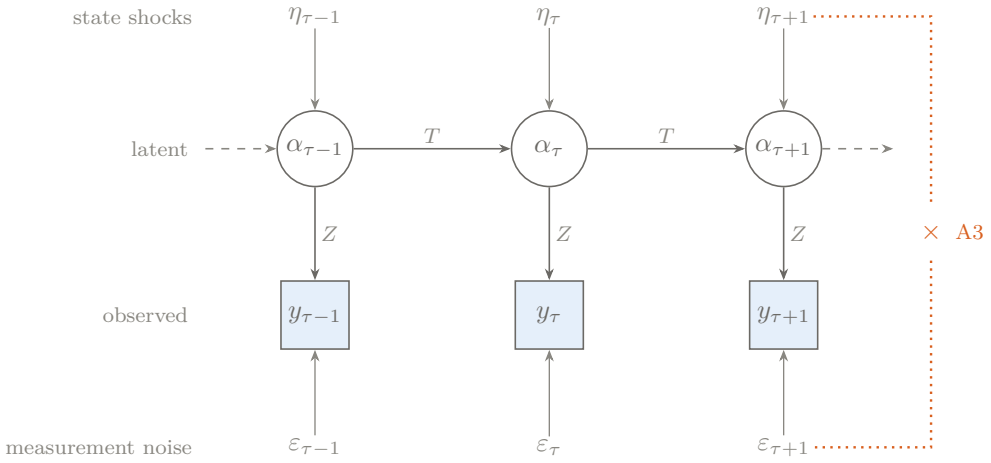
A4 (Initial state). $\alpha_1 \sim \mathcal{N}(a_{1|0}, P_{1|0})$. This supplies the distribution needed to compute the first innovation and begin the recursion. A poor but proper initialization fades as informative data arrive; a missing or internally inconsistent one leaves the first update undefined.

Together, A1, A2, A3, and A4 let means and covariances carry all information needed for the conditional distribution. The normality assumption makes those conditional distributions Gaussian. The same recursions remain best linear predictors under weaker distributional assumptions, but their interpretation as the full conditional distribution then requires care. These assumptions are therefore both conveniences and diagnostic claims: the innovations later tell us when they are inadequate.

Figure 5 displays the dependence structure implied by Equation 8, Equation 9, and assumption A3.

[14] A deterministic regressor is treated as known rather than generated by the stochastic state equation. Examples include an intercept, a calendar dummy, a pre-specified trend, or an intervention indicator.

Figure 5. The state space model. The transition matrix T carries the state from one reference period to the next. The measurement matrix Z connects each state to observations from that period. The crossed orange path marks assumption A3: no arrow connects a state shock to a measurement error.



Within a single data vintage r, arrange the observations by reference period and write

$$Y_s^{(r)} = \{y_1^{(r)}, \dots, y_s^{(r)}\}.$$

The filtered mean[15] and covariance are

$$a_{\tau|s}^{(r)} = \mathbb{E}[\alpha_\tau \mid Y_s^{(r)}], \qquad P_{\tau|s}^{(r)} = \mathrm{Var}(\alpha_\tau \mid Y_s^{(r)}). \tag{10}$$

The vintage superscript is essential when comparing release dates. During one filter pass it clutters every expression, so the derivation below suppresses it. The time subscripts then refer only to economic reference periods.

This notation can represent much more than an underlying level. The practical question is what the state must remember so that one step of the model contains all information needed for the next.

3.2 State Vectors for Autoregressive Lags and Model Components

The local level model uses one state. Larger states allow the model to remember lags, separate trends from cycles, and combine several common forces.

[15] A filtered mean is the conditional mean of a state after incorporating observations available through the filter date. "Filtered" distinguishes it from a prediction made before the current observation and a smoothed estimate that also uses later observations.

An AR(p) process[16] provides a simple example. Stack the current value and its lags[17], $\alpha_\tau = (y_\tau, y_{\tau-1}, \dots, y_{\tau-p+1})'$, and use

$$T = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad Z = (1, 0, \dots, 0).$$

The first row of T contains the forecasting equation. The rows underneath move each lag down one position. This companion form is the standard way to put finite-order dynamics into a first-order state equation.

The same notation will return in later sessions. Table 4 identifies the state components used by several models in the course.

Table 4. The state vector is chosen to retain the information needed for the next prediction.

Model	What the state contains	Where it returns
Local level	an underlying level	this session
Dynamic factor model	common factors and their lags	Session 2
Trend and seasonal model	level, slope, and seasonal states	Session 3
Time-varying regression	coefficients that change over time	later nonlinear and instability work

Session 2 addresses the main modeling question omitted today: when dozens of series measure activity, how should Z, the factor dynamics, and the noise covariances be estimated? For this session, take the system matrices as known while learning what the filter does with them.

4 The Multivariate Kalman Filter

The model has now separated two tasks: the state equation describes how the economy moves, and the measurement equation describes how indicators observe it. The filter combines them. Rather than present

[16] An autoregressive process predicts a variable from its own past values and a new disturbance. AR(p) means that the forecasting equation uses the previous p values.

[17] A lag is an earlier value indexed relative to the current period: $y_{\tau-1}$ is the first lag, $y_{\tau-2}$ the second, and so on.

its matrix formulas as a new algorithm to memorize, we derive them from one familiar question: if two random vectors move together and one is observed, what does that observation tell us about the other?

4.1 Gaussian Derivation of the Kalman Update

The update in Equation 5 is an application of one conditional-normal formula[18].

Lemma 4.1 (Gaussian conditioning and linear projection)**.** *If*

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \right),$$

with Σ_{yy} *invertible*[19]*, then*

$$x \mid y \sim \mathcal{N}\big(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y),\ \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\big)\,.$$

Here is how to read Lemma 4.1. The first displayed block says that x and y are jointly normal, with their means stacked above a four-block covariance matrix. The diagonal blocks are the uncertainty in x and y separately; the off-diagonal blocks record how they move together. After y is observed, the notation $x \mid y$ asks for the distribution of the unobserved block x conditional on that realized value.

The conditional mean has a familiar structure: prior mean plus a coefficient times a surprise. That coefficient, $B = \Sigma_{xy}\Sigma_{yy}^{-1}$, is the linear projection[20] coefficient. Indeed, the residual

$$u = x - \mu_x - B(y - \mu_y)$$

satisfies $\mathrm{Cov}(u, y) = \Sigma_{xy} - B\Sigma_{yy} = 0$. Joint normality turns that zero covariance into independence, so observing y leaves only the residual uncertainty in u. This gives both the adjusted mean and the subtracted covariance in Lemma 4.1. The amount of uncertainty removed depends on the covariance structure, not on whether the realized y happens to be high or low.

The Kalman update follows once we identify x with the next state and y with its new observation. We

[18] A conditional-normal formula gives the distribution of one block of a jointly normal vector after another block has been observed. Its mean is a linear adjustment and its covariance does not depend on the observed numerical value.

[19] A square matrix is invertible when no nonzero direction is redundant: there is a matrix A^{-1} satisfying $A^{-1}A = I$. A singular Σ_{yy} means some linear combination of y has zero variance and must be removed or handled with a generalized method.

[20] A linear projection is the least-squares linear predictor of one random quantity from another. Its residual is uncorrelated with every predictor used in the projection.

first need their predicted joint covariance.

4.2 Matrix Prediction and Measurement-Update Equations

Suppose the current filtered distribution is

$$\alpha_\tau \mid Y_\tau \sim \mathcal{N}(a_{\tau|\tau}, P_{\tau|\tau}).$$

The **prediction step** applies the state equation:

$$\begin{aligned} a_{\tau+1|\tau} &= T a_{\tau|\tau}, \\ P_{\tau+1|\tau} &= T P_{\tau|\tau} T' + RQR'. \end{aligned} \tag{11}$$

Both lines follow directly from Equation 8 and the assumptions:

$$\begin{aligned} \mathbb{E}[T\alpha_\tau + R\eta_\tau \mid Y_\tau] &= T a_{\tau|\tau}, \\ \mathrm{Var}(T\alpha_\tau + R\eta_\tau \mid Y_\tau) &= T P_{\tau|\tau} T' + RQR'. \end{aligned}$$

The first equality uses the mean-zero shock. The second propagates existing state uncertainty through T and then adds the independent uncertainty of the new shock. A covariance cross-term would appear here without A3.

Before observing $y_{\tau+1}$, the joint conditional distribution of state and observation is

$$\begin{pmatrix} \alpha_{\tau+1} \\ y_{\tau+1} \end{pmatrix} \Big| \, Y_\tau \sim \mathcal{N}\left(\begin{pmatrix} a_{\tau+1|\tau} \\ Z a_{\tau+1|\tau} \end{pmatrix}, \begin{pmatrix} P_{\tau+1|\tau} & P_{\tau+1|\tau} Z' \\ Z P_{\tau+1|\tau} & Z P_{\tau+1|\tau} Z' + H \end{pmatrix} \right).$$

Read this display as the input required by Lemma 4.1. The top row describes the unobserved state, the bottom row its not-yet-seen measurement. The four covariance blocks follow from the two model equations. The state has predicted covariance $P_{\tau+1|\tau}$. The observation has covariance $Z P_{\tau+1|\tau} Z' + H$: uncertainty about the state is carried through Z, and measurement noise adds H. The cross-covariance is $P_{\tau+1|\tau} Z'$ because A3 makes the measurement error independent of the state.

Define the innovation and its covariance:

$$v_{\tau+1} = y_{\tau+1} - Za_{\tau+1|\tau}, \qquad F_{\tau+1} = ZP_{\tau+1|\tau}Z' + H. \tag{12}$$

In Lemma 4.1 set $x = \alpha_{\tau+1}$, $y = y_{\tau+1}$, $\Sigma_{xy} = P_{\tau+1|\tau}Z'$, and $\Sigma_{yy} = F_{\tau+1}$. Substitution gives the Kalman gain and the **update step**:

$$\begin{aligned} K_{\tau+1} &= P_{\tau+1|\tau}Z'F_{\tau+1}^{-1}, \\ a_{\tau+1|\tau+1} &= a_{\tau+1|\tau} + K_{\tau+1}v_{\tau+1}, \\ P_{\tau+1|\tau+1} &= (I - K_{\tau+1}Z)P_{\tau+1|\tau}. \end{aligned} \tag{13}$$

Read the three lines as "weight, revise, learn." The first line computes the weight placed on each direction of the surprise. The second adds the weighted innovation to the predicted state. The third removes the state uncertainty resolved by that observation. These are the scalar formulas from Section 2.2 with matrices in place of numbers. The gain is now an $m \times p$ matrix. Its jth column tells us how a surprise in observation j revises each component of the state, after accounting for the covariance among all observations in that update.

The innovation has conditional mean zero[21] and covariance $F_{\tau+1}$. Under the stated Gaussian assumptions[22], innovations from different reference periods are independent[23]. Serial correlation[24] in fitted innovations is therefore evidence that the model has omitted predictable structure.

4.3 Kalman Filtering with Partially Observed Data

The derivation used a complete observation vector, but the resulting recursion does not require one. At each reference period we form the same three objects from whatever rows are available. Algorithm 1 is best read as a repeated conversation between model and data: predict what should arrive, compare that prediction with what did arrive, revise, and move forward.

[21] Conditional mean zero means that, before a release is seen, its expected forecast error is zero given the model's current information. It does not mean every realized innovation is small.

[22] The Gaussian assumptions are joint normality of the initial state and disturbances, together with the stated linear equations and independence conditions. They make uncorrelated Gaussian innovations independent.

[23] Random quantities are independent when learning one does not change the distribution of the other. Independence is stronger than zero covariance outside the Gaussian case.

[24] Serial correlation is correlation between a series and its own earlier values. In innovations it signals that some predictable time pattern remains.

Algorithm 1. Kalman filter for one vintage

Start from $a_{1|0}$ and $P_{1|0}$. For $\tau = 1, \dots, n$:

1. Form the innovation v_τ, innovation covariance F_τ, and gain K_τ using the observations available for reference period τ.
2. Update $a_{\tau|\tau}$ and $P_{\tau|\tau}$. If no observation is available for that period, retain the predicted mean and covariance.
3. Predict $a_{\tau+1|\tau}$ and $P_{\tau+1|\tau}$.

Store the predicted moments and innovations. They are used later for the likelihood, diagnostics, and smoother.

The stored objects matter as much as the final state. Without the predicted moments we cannot explain a release surprise, evaluate the likelihood, or run a backward smoother. Before adding the complication of a ragged panel, a short numerical pass gives us a transparent check that every stage has been coded in the intended order.

4.4 Two-Period Numerical Unit Test for the Local Level Model

Matrix notation can conceal simple timing mistakes. This example returns to the scalar model so that each number can be verified by hand and later used as a unit test for a general implementation.

Consider the local level model with $Q = 0.1024$, $H = 1$, $a_{1|0} = 0$, and $P_{1|0} = 10$. Let the first two observations be $y_1 = 2.4$ and $y_2 = 1.9$.

Table 5 applies the filter twice and records every intermediate quantity.

Table 5. Two local level filter updates. The rounded value 2.182 is used in the second innovation.

Quantity	$\tau = 1$	$\tau = 2$
$v_\tau = y_\tau - a_{\tau\|\tau-1}$	2.400	−0.282
$F_\tau = P_{\tau\|\tau-1} + H$	11.000	2.011
$K_\tau = P_{\tau\|\tau-1}/F_\tau$	0.909	0.503
$a_{\tau\|\tau}$	2.182	2.040
$P_{\tau\|\tau}$	0.909	0.503
$P_{\tau+1\|\tau} = P_{\tau\|\tau} + Q$	1.011	0.605

The large first gain reflects the diffuse[25] starting variance. After the first observation, the state is much more precisely estimated, so the second release receives less weight. This table is a useful unit test for any first implementation of the filter.

4.5 Initialization and Numerically Stable Covariance Updates

Every recursion needs a place to start. In the Kalman filter, the first innovation is $y_1 - Za_{1|0}$ and its variance depends on $P_{1|0}$. Without an initial mean and covariance, neither quantity exists and the first release cannot be weighted. The choice of **initialization** says what the model knows before seeing the sample.

For a **stationary** state, a common choice is its unconditional mean and covariance. That choice starts the state in the same distribution implied by its long-run dynamics, so the early filter does not contain an artificial transient. After demeaning, $a_{1|0} = 0$, and the covariance solves the discrete Lyapunov equation[26]

$$P_{1|0} = TP_{1|0}T' + RQR'. \tag{14}$$

Read Equation 14 as a balance condition: the covariance before a transition must equal the covariance obtained after propagating it through T and adding one period of shock variance. A stable transition has a unique finite balance point.

For a **nonstationary** state such as a random walk, no unconditional covariance exists. A **diffuse** initialization represents weak prior information by assigning very large uncertainty to the unknown initial level. Exact diffuse methods handle the limiting case analytically; a large finite variance is a convenient approximation but should be checked for numerical sensitivity (Durbin and Koopman, 2012, ch. 5). Early diffuse periods require special treatment in the likelihood[27] and should usually be omitted from residual diagnostics.

Three implementation practices are worth adopting from the beginning:

1. **Solve rather than invert.** To obtain $F_\tau^{-1}v_\tau$, solve $F_\tau x = v_\tau$. A factorization-based solve is faster and less sensitive to rounding error than explicitly constructing an inverse.
2. **Restore covariance symmetry.** A true covariance satisfies $P = P'$, but floating-point arithmetic

[25] Diffuse means that the initial state is assigned very weak prior information, represented here by a variance much larger than the measurement-error variance. The first observation therefore receives nearly all the weight.

[26] A discrete Lyapunov equation has the form $P = TPT' + S$. For a stable T, its unique solution is the unconditional covariance generated when each period propagates existing uncertainty and adds new shock covariance S.

[27] A likelihood is the joint density of the observed data viewed as a function of the model parameters. Diffuse initial conditions change the density contribution of the earliest observations.

can leave tiny asymmetric entries after many updates. Replacing P with $(P + P')/2$ removes that numerical artifact without changing the intended matrix.

3. **Use the Joseph covariance update when needed.** The Joseph covariance form[28] is

$$P_{\tau|\tau} = (I - K_\tau Z_\tau)P_{\tau|\tau-1}(I - K_\tau Z_\tau)' + K_\tau H_\tau K_\tau'. \tag{15}$$

With the optimal gain, Equation 15 is algebraically equivalent to the shorter covariance update. Its sum-of-quadratic-forms representation is more resistant to a computed covariance becoming indefinite. Initialization and numerical care do not change the economic model, but they determine whether the recursive calculation faithfully represents it.

5 GDP Nowcasts from Real-Time Data Vintages

The filter is now complete, but a filter is not yet a GDP nowcasting system. We still have to translate a release calendar into observed matrix rows, reconcile monthly latent growth with a quarterly target, and extract the target's mean and variance. These are not bookkeeping details. Each step determines which economic quantity the model is estimating and which information it was allowed to use.

5.1 Observed-Row Selection for a Data Vintage

Fix a release date r. Suppose only some components[29] of $y_\tau^{(r)}$ are available. Let $W_\tau^{(r)}$ select[30] those components.

For example, if a three-series vector contains observations for series 1 and 3 but not series 2, then

$$W_\tau^{(r)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The measurement equation used in that vintage is

[28] The Joseph form writes posterior covariance as the sum of propagated prior covariance and propagated measurement-error covariance. Each term is positive semidefinite, which makes the expression more resistant to a computer producing an impossible negative variance.

[29] The components of a vector are its individual entries. Here each component of y_τ is one series measured for reference period τ.

[30] Selection here is a deterministic matrix operation that keeps rows available in a vintage. It is distinct from statistical sample selection, in which inclusion may depend on the data-generating process.

$$y_\tau^{obs,(r)} = W_\tau^{(r)} Z\alpha_\tau + W_\tau^{(r)}\varepsilon_\tau,$$

with

$$Z_\tau^{(r)} = W_\tau^{(r)} Z, \qquad H_\tau^{(r)} = W_\tau^{(r)} H (W_\tau^{(r)})'. \tag{16}$$

The first row of $W_\tau^{(r)}$ takes the dot product with the original vector and returns its first component; the second returns its third. Multiplying Z on the left keeps the corresponding measurement rows. Multiplying H on both sides keeps their variances *and* the covariance between them. Equation Equation 16 is therefore the same measurement model, reduced to what could actually be observed in vintage r.

Algorithm 1 then uses only the observed rows. If a period contains no observed series, it receives a prediction step but no update.

This is computationally convenient, but it does not make the data problem trivial. The selection matrix must be built from an accurate release calendar because $W_\tau^{(r)}$ is a claim about what was knowable on date r. If a row is included one day early, the filter uses future information; if it is included one month late or assigned to the wrong reference period, the innovation is matched to the wrong state.

A value removed after looking at its magnitude is also not automatically an innocent missing observation; that decision can introduce selection. Outliers[31] may instead require an intervention variable[32], a heavy-tailed[33] error model, or a documented rule fixed before evaluation.

When a new vintage arrives, it can fill a cell at the ragged edge or revise an older cell. The simplest reliable production procedure is to rebuild the vintage matrix and rerun the filter with fixed parameters. More specialized algorithms[34] can update old observations without a full pass, but they must preserve the same reference-period and release-date logic. Otherwise a computational shortcut can apply a revision to the wrong economic period or make it appear available before its actual release.

Row selection solves the "which observations?" problem. It does not yet solve the "which frequency?" problem: the state may evolve monthly while GDP is observed quarterly.

[31] Outliers are observations far from the bulk of the data or from a model's predictive distribution. They may be genuine extreme events, recording errors, or evidence that the assumed error distribution is too thin-tailed.

[32] An intervention variable is a pre-specified regressor that isolates an unusual event, such as a strike, shutdown, or definitional break, rather than forcing the regular state dynamics to explain it.

[33] A heavy-tailed error distribution assigns more probability to extreme deviations than a normal distribution. Student-t errors are a common example.

[34] Examples include disturbance-smoothing revision updates, fixed-lag smoothers, and incremental or square-root filters that update stored factorizations. They save work by reusing earlier calculations, but require careful bookkeeping when a historical observation is revised.

5.2 Monthly-to-Quarterly GDP Aggregation

Suppose the model runs at monthly frequency, and g_τ is latent monthly log GDP growth. Quarterly GDP is a flow: the quarterly level is the sum of the three monthly levels. Let L_τ be the monthly level, $\ell_\tau = \log L_\tau$, and $g_\tau = \ell_\tau - \ell_{\tau-1}$. A log-linear approximation[35] around three equal monthly levels gives

$$\log(L_\tau + L_{\tau-1} + L_{\tau-2}) \approx \log 3 + \tfrac{1}{3}(\ell_\tau + \ell_{\tau-1} + \ell_{\tau-2}).$$

Subtract the corresponding approximation for the preceding quarter:

$$y_\tau^Q \approx \tfrac{1}{3}\left[(\ell_\tau + \ell_{\tau-1} + \ell_{\tau-2}) - (\ell_{\tau-3} + \ell_{\tau-4} + \ell_{\tau-5})\right].$$

Replacing level differences with the monthly growth rates and collecting like terms yields

$$y_\tau^Q = \tfrac{1}{3}g_\tau + \tfrac{2}{3}g_{\tau-1} + g_{\tau-2} + \tfrac{2}{3}g_{\tau-3} + \tfrac{1}{3}g_{\tau-4}, \tag{17}$$

where y_τ^Q is quarter-on-quarter log growth assigned to the final month of the quarter (Mariano and Murasawa, 2003). This is a mixed-frequency[36] measurement equation, not an interpolation[37].

If the published target is expressed at a seasonally adjusted annual rate, the model and target must be put in consistent units before estimation or reporting. Otherwise a coefficient or forecast error compares a quarterly rate with a number roughly four times as large; the model can absorb that unit error into a loading, but the reported nowcast, uncertainty, and evaluation loss will not have the intended interpretation.

Figure 6 displays the five aggregation weights. The tent shape is not a decorative pattern: it counts how many monthly levels on opposite sides of the quarterly comparison contain each growth rate. The middle growth rate affects all three month-to-month paths between the adjacent quarterly averages, while the rates at either end affect only one.

To implement Equation 17, include g_τ and four lags in the state and give GDP the measurement row

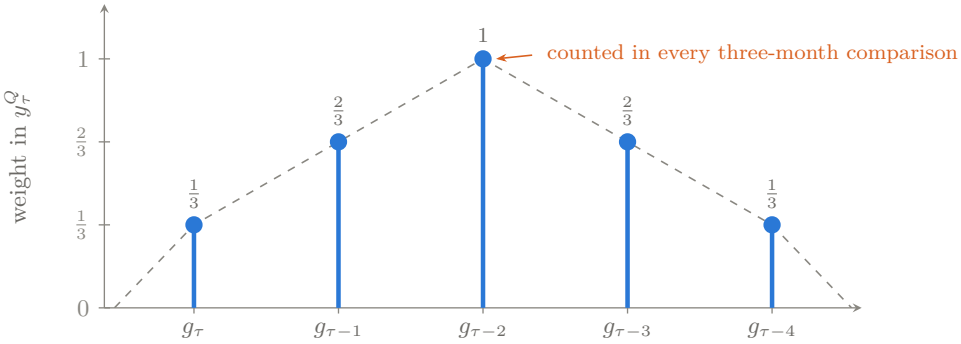
$$Z_Q = (\tfrac{1}{3}, \tfrac{2}{3}, 1, \tfrac{2}{3}, \tfrac{1}{3}).$$

[35] A log-linear approximation replaces a nonlinear relationship with its first-order Taylor expansion in logarithms. It is accurate for modest changes around the expansion point and turns multiplicative growth relationships into weighted sums.

[36] A mixed-frequency model relates variables observed or evolving at different intervals, such as monthly indicators and quarterly GDP, without first forcing all of them onto one artificial frequency.

[37] Interpolation fills unobserved values between observed endpoints by a chosen rule. The aggregation here instead defines how a monthly latent process maps economically into an observed quarterly flow.

Figure 6. Monthly growth contributions to quarterly GDP growth. The middle monthly growth rate affects more terms in the comparison of adjacent quarterly averages, producing the five tent-shaped weights.



Why four lags? The oldest term in Equation 17 is $g_{\tau-4}$. A first-order state equation can use only what its current state stores, so the vector must retain the current monthly growth rate and the four earlier rates needed by the GDP row. Fewer lags would discard a contribution before the quarterly observation could use it.

That row exists only for quarter-ending reference months. Even then, its value remains missing in vintage r until the GDP release date. This is where the two clocks matter: the observation is modeled at the quarter's reference date but becomes available later.

The same aggregation rule does not apply to every series. Flows may be summed or averaged over a period; stocks are often measured at a point in time; rates and indexes require definitions appropriate to the series. Mixed-frequency modeling begins with those economic choices, not with a generic interpolation rule.

With the vintage rows and frequency mapping in place, the final step is to ask for the distribution of the GDP measurement implied by the filtered state.

5.3 GDP Point Nowcast and Conditional Variance

Let $T(Q)$ denote the final monthly reference period needed to construct quarter Q. At vintage r, filter the available data through the latest reference period and predict the state to $T(Q)$ when necessary:

$$\begin{aligned} a_{s+1|r} &= T a_{s|r}, \\ P_{s+1|r} &= T P_{s|r} T' + RQR'. \end{aligned} \tag{18}$$

Each application of Equation 18 moves the state one month closer to the quarter-ending reference period.

The mean follows T; the covariance follows T and accumulates one more block of shock uncertainty RQR'.

The notation $s \mid r$ here emphasizes that the state is dated by reference period while the conditioning information is dated by vintage. The GDP nowcast and its model-implied conditional variance are

$$\begin{aligned} \widehat{y}_{Q|r} &= Z_Q a_{T(Q)|r}, \\ \mathrm{Var}(y_Q \mid \Omega_r) &= Z_Q P_{T(Q)|r} Z_Q' + H_Q. \end{aligned} \tag{19}$$

The first line reads the GDP target from the predicted state using the aggregation row Z_Q. The second carries state uncertainty into GDP units and adds any GDP-specific measurement variance H_Q. It is the matrix version of "estimate plus uncertainty," evaluated for one target quarter and one release date.

As releases enter Ω_r, the point estimate changes when their innovations are nonzero. Under a fixed linear Gaussian model with known parameters, adding observations weakly reduces the conditional variance. Calendar time without new information does not by itself narrow the band.

The variance in Equation 19 is conditional on the chosen model and its parameters. A published interval based on estimated parameters usually omits some parameter uncertainty. Bootstrap or Bayesian procedures can incorporate that additional source of uncertainty, at extra computational cost[38].

5.4 What the Real-Time Layer Adds

Section 5 has added three pieces around the Kalman recursion. The release calendar determines $W_\tau^{(r)}$ and therefore which rows may update the state. The aggregation equation determines how a monthly state becomes a quarterly flow. The target row then converts the resulting state distribution into a point nowcast and uncertainty band. A correct filter applied to an incorrect calendar or frequency mapping is still an incorrect nowcast.

Those steps produce a revised number. The next question is the one a user of the nowcast will ask first: *why did it change?*

[38] For example, a parametric bootstrap repeatedly simulates a dataset, re-estimates the model, and recomputes the nowcast; a Bayesian analysis draws parameters and states from their joint posterior. Both replace one filter pass with hundreds or thousands of estimation-and-filtering passes, which is why the added uncertainty assessment costs computation.

6 Release News and Nowcast Revisions

Operational nowcasting is not complete when the dashboard number moves. A central bank, firm, or market analyst needs to connect that revision to the releases that arrived. The same covariance calculation used by the filter gives that attribution; no second forecasting model is required.

Let Ω_{old} be the information set before a release and Ω_{new} the information set afterward. Let v collect the innovations in the new or revised observations, computed using the old vintage. For a target y^*, Gaussian conditioning gives

$$\mathbb{E}[y^* \mid \Omega_{new}] - \mathbb{E}[y^* \mid \Omega_{old}] = \operatorname{Cov}(y^*, v \mid \Omega_{old}) \operatorname{Var}(v \mid \Omega_{old})^{-1} v. \tag{20}$$

Write the row vector before v as w'. Then

$$\Delta \text{ nowcast} = \sum_i w_i v_i. \tag{21}$$

Each contribution has two parts:

- v_i, the difference between a release and the model's expectation for that release;
- w_i, the way that surprise maps into the target after accounting for its precision and its covariance with the other releases.

In the scalar release from Section 2.2, $v = -0.8$, $w = K = 0.25$, and the nowcast contribution is -0.2. The news decomposition[39] is therefore not an additional model. It records the same update in target units.

If several observations arrive together, their joint covariance belongs in the calculation. Suppose retail sales and card spending contain partly the same consumption news. A joint block recognizes that overlap and does not treat both signals as independent evidence. Sequential processing reaches the same final state in a linear model with exact arithmetic, but the first release processed temporarily receives credit for information shared with the second. Reversing the order reverses that intermediate credit. A joint block defines one common baseline and produces a simultaneous attribution that does not depend on an arbitrary ordering.

The model's expectation is not necessarily the survey consensus quoted in a news report. A release can surprise market economists and still be close to the model's forecast, or the reverse. A **nowcast**

[39] A decomposition expresses one total as a sum of interpretable parts. Here the total nowcast revision is split into release-level contributions whose sum reproduces the revision.

commentary should say which expectation defines its surprise.

The decomposition is exact[40] conditional[41] on a fixed[42] linear model and fixed parameters. If the model is re-estimated between vintages, part of the nowcast change comes from new parameter estimates. A complete **production report** should show that component separately rather than assigning it to the data releases.

Attribution assumed the system matrices were already known. To build an empirical model, we must choose them from data and then ask whether the resulting innovations behave as the model says they should.

7 Likelihood Estimation and Innovation Diagnostics

Filtering answers what the state would be *if* T, Z, Q, H were known. In an empirical model they are not. We need a criterion for choosing parameters and a way to check whether the fitted system has left predictable information behind. The innovation sequence supplies both: its conditional density forms the likelihood, and its standardized behavior forms the diagnostics.

7.1 Prediction-Error Likelihood for State-Space Parameters

So far, T, Z, Q, H have been treated as known. In an empirical model they contain unknown parameters θ. The filter provides a convenient likelihood because it computes the conditional distribution of each observation given earlier observations.

The chain rule expands a joint density into successive conditional densities:

$$\begin{aligned} p(y_1, \dots, y_n \mid \theta) &= p(y_n \mid Y_{n-1}, \theta) p(y_{n-1} \mid Y_{n-2}, \theta) \cdots p(y_1 \mid \theta) \\ &= \prod_{\tau=1}^{n} p(y_\tau \mid Y_{\tau-1}, \theta). \end{aligned}$$

The first line peels off the last observation, then the next-to-last, until only the first remains. The second line writes that same product compactly. The filter has already computed the mean and variance of every

[40] Exact means the reported contributions add to the model's computed nowcast change, apart from numerical rounding. It does not mean the model itself is true.

[41] Conditional means "holding stated objects given." Here the attribution is computed given the model structure, parameters, and old information set.

[42] Fixed parameters are held at the same numerical values before and after the release. If they are re-estimated, the before-and-after mapping from data to the target also changes.

factor in the product, including the smaller observation blocks created by missing data.

For observed rows at period τ, the conditional mean is $Z_\tau a_{\tau|\tau-1}$, the error is v_τ, and its covariance is F_τ. The Gaussian **log-likelihood** is

$$\log L(\theta) = -\frac{1}{2}\sum_{\tau=1}^{n}\left[p_\tau \log(2\pi) + \log|F_\tau| + v_\tau' F_\tau^{-1} v_\tau\right], \tag{22}$$

where p_τ is the number of observed rows. The three terms have distinct roles. The first is the normal-density constant for the observed dimension. The log determinant penalizes a model that declares a very wide predictive distribution. The quadratic form penalizes a realized innovation that is large relative to that distribution. A parameter vector fits well only by balancing calibration and sharpness.

A numerical optimizer is a search routine wrapped around the filter. It proposes a value of θ, runs the filter, records Equation 22, and uses the change in the objective (and, when available, its derivatives) to choose a new proposal. Iteration stops when further parameter changes no longer improve the objective materially. This is useful because the state estimates depend on the parameters and the parameters' likelihood depends on the state estimates; the optimizer closes that loop without requiring a closed-form solution.

Parameters with sign restrictions[43] should be transformed during optimization. If a standard deviation is written $\sigma = \exp(\gamma)$, the optimizer may search over any real γ while every proposed σ remains positive. Without the transformation, trial steps can produce impossible covariances and make the filter fail.

Large **dynamic factor models** require additional identification restrictions and often use two-step[44] or EM estimators[45] Those belong in Session 2, after the factor model and its normalizations have been introduced. For now, the important point is that the innovations serve two roles: they update the state and evaluate how well a proposed parameter vector predicts the data.

A high likelihood is not proof that the model's forecast errors have the structure we assumed. That claim has observable implications, which we check next.

[43] A sign restriction limits a parameter to be positive, negative, or nonnegative. Variances and standard deviations, for example, cannot be negative.

[44] A two-step estimator first estimates factors and parameters by a simpler method, often principal components and regression, then holds those estimates fixed while applying the Kalman filter to the ragged panel.

[45] An expectation-maximization (EM) estimator alternates between computing expected latent-state sufficient statistics with a smoother and maximizing parameters given those expectations. Each iteration does not lower the likelihood, though convergence can be slow.

7.2 Innovation Diagnostics for Bias, Scale, Dependence, and Stability

The likelihood compresses fit into one number. Diagnostics unpack it. A model can have an acceptable objective value while consistently missing in one direction, understating recession uncertainty, or leaving a lag pattern in its errors. Standardization puts innovations from different series and dates onto a common reference scale so those failures can be seen.

Under a correctly specified model, innovations have conditional mean zero and covariance F_τ. A convenient standardized residual[46] is obtained from a matrix square root[47] such as the Cholesky factor[48] If $F_\tau = L_\tau L_\tau'$, define

$$\tilde{v}_\tau = L_\tau^{-1} v_\tau. \tag{23}$$

Read Equation 23 as a solve, not as an instruction to construct L_τ^{-1}. Compute $\tilde{v}_\tau$ from $L_\tau \tilde{v}_\tau = v_\tau$. Because

$$\mathrm{Var}(\tilde{v}_\tau \mid Y_{\tau-1}) = L_\tau^{-1} F_\tau L_\tau^{-\prime} = I,$$

the resulting components have unit variance and no contemporaneous covariance under the model. A value of 2 now means a surprise two conditional standard deviations from zero, regardless of the original series' units.

For a correctly specified linear Gaussian model, the available components of $\tilde{v}_\tau$ should behave like independent standard normal draws. Useful checks include:

- **Centering**. A nonzero mean indicates a systematic prediction bias.
- **Scale**. An average squared residual far from one indicates miscalibration, but it does not identify a unique parameter at fault. The problem may involve H, Q, omitted dynamics, outliers, or instability.
- **Serial dependence**. Autocorrelation in residuals indicates remaining predictable structure. Autocorrelation in squared residuals suggests changing volatility.
- **Distributional shape**. A QQ plot reveals whether a few episodes, such as recessions or shutdowns, dominate departures from normality.
- **Stability**. Residual plots by date can show a model that fit one period but not another.

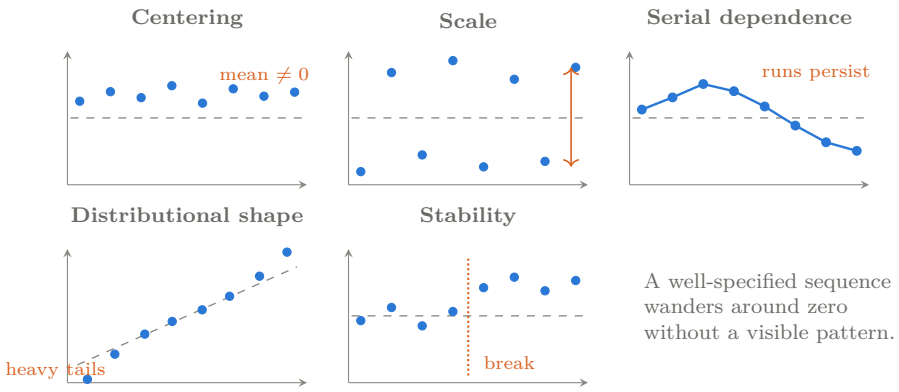
[46] A standardized residual rescales a forecast error by its model-implied standard deviation and, in several dimensions, removes contemporaneous covariance. It is measured in forecast-standard-deviation units.

[47] A matrix square root of a positive-definite matrix F is a matrix L satisfying $F = LL'$. It generalizes the scalar identity $\sigma^2 = \sigma\sigma$.

[48] The Cholesky factorization writes a symmetric positive-definite matrix as $F = LL'$, where L is lower triangular with positive diagonal entries. It provides a stable way to solve systems and whiten correlated residuals.

Figure 7 gives a stylized visual vocabulary for these checks. The panels are not test statistics; they show the kinds of patterns a statistic or plot is meant to detect.

Figure 7. What innovation failures look like. A centered, correctly scaled innovation cloud should fluctuate without pattern around zero. Each panel isolates one departure: a shifted center, excessive spread, persistent runs, unusually frequent extremes, or a relationship that changes partway through the sample.



With a ragged panel, residual dimension varies by period. Diagnostics must use the observations that were actually present, and early diffuse periods should be treated separately. Rejecting a diagnostic does not by itself prescribe the fix; it tells us what kind of model feature to investigate.

8 Filtered and Smoothed State Estimates

Filtering deliberately refuses to use the future. That is exactly right for a real-time nowcast, but it is not always the question an analyst wants to answer. After the full vintage is available, later observations may clarify whether an earlier movement was a genuine change in the state or temporary measurement noise. Smoothing passes that later information backward while keeping the vintage itself fixed.

8.1 Two Information Sets, Two Estimates

A **filtered estimate** respects the information available at each reference period within a vintage. A **smoothed estimate** asks a different question: using the entire vintage Y_n, what do we infer about an earlier state? At the last period the two coincide, because there are no later observations to add. Earlier in the sample they can differ in both their means and their uncertainty.

8.2 Deriving the Backward Recursion

Starting from the last filtered state, the **Rauch–Tung–Striebel smoother** runs backward. Conditional on Y_τ, the state equation implies

$$\begin{aligned}\operatorname{Cov}(\alpha_\tau, \alpha_{\tau+1} \mid Y_\tau) &= P_{\tau|\tau} T', \\ \operatorname{Var}(\alpha_{\tau+1} \mid Y_\tau) &= P_{\tau+1|\tau}.\end{aligned}$$

Why is the backward weight covariance times inverse variance? Consider a linear prediction of the current state error from the next-state prediction error:

$$(\alpha_\tau - a_{\tau|\tau}) \approx B(\alpha_{\tau+1} - a_{\tau+1|\tau}).$$

The least-squares normal equation sets the residual uncorrelated with the predictor,

$$B \operatorname{Var}(\alpha_{\tau+1} \mid Y_\tau) = \operatorname{Cov}(\alpha_\tau, \alpha_{\tau+1} \mid Y_\tau).$$

Solving for B gives covariance times inverse variance. Substituting the two blocks above defines the smoother gain:

$$J_\tau = P_{\tau|\tau} T' P_{\tau+1|\tau}^{-1}.$$

This has the same covariance-over-variance form as the Kalman gain because both are linear projections. The filter projects a state on a contemporaneous data surprise; the smoother projects an earlier state on what we later learned about the next state. Applying Lemma 4.1 to α_τ and $\alpha_{\tau+1}$, then averaging over what the full vintage Y_n tells us about $\alpha_{\tau+1}$, gives the backward recursions

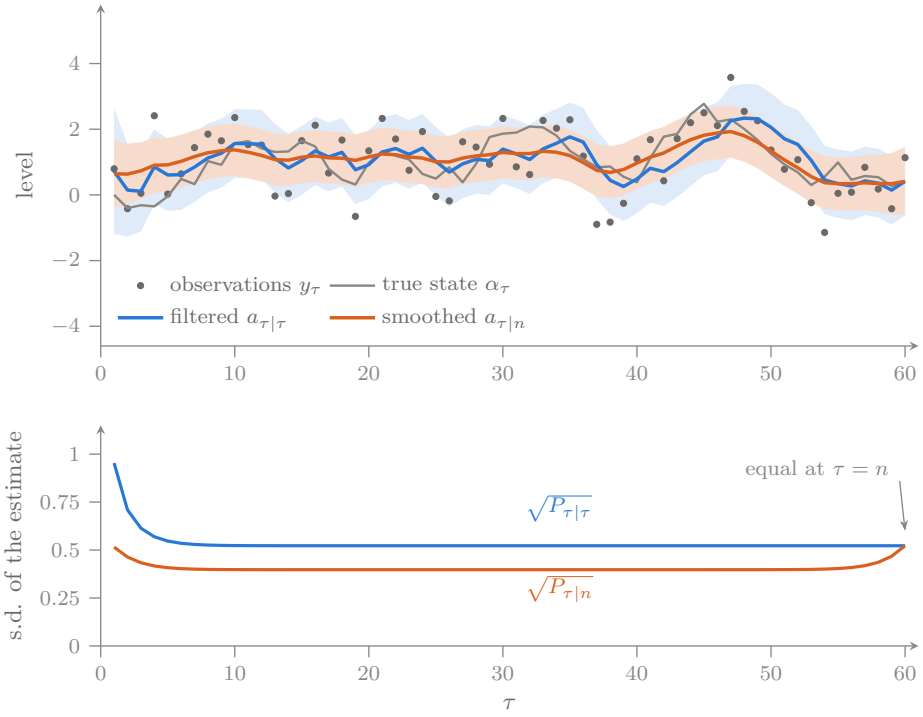
$$\begin{aligned}a_{\tau|n} &= a_{\tau|\tau} + J_\tau(a_{\tau+1|n} - a_{\tau+1|\tau}), \\ P_{\tau|n} &= P_{\tau|\tau} + J_\tau(P_{\tau+1|n} - P_{\tau+1|\tau})J_\tau'.\end{aligned} \tag{24}$$

Read the mean line as "filtered estimate plus backward revision." The difference $a_{\tau+1|n} - a_{\tau+1|\tau}$ is what later observations taught the model about the next state. The smoother gain J_τ maps the predictable part of that news back to period τ. In the covariance line, $P_{\tau+1|n} - P_{\tau+1|\tau}$ is negative semidefinite because later data cannot add uncertainty about the next state; premultiplying and postmultiplying by J_τ transfers that

reduction backward.

Figure 8 compares the filtered and smoothed paths and their conditional standard deviations.

Figure 8. Filtered and smoothed estimates. The filtered path uses observations only through each date. The smoothed path uses the full simulated sample. Where later states are informative about earlier ones, the smoothed band is narrower.



Filtered estimates are the appropriate objects for reconstructing what could have been known in real time. Smoothed estimates are useful for historical description and for parameter-estimation procedures such as EM. Smoothing cannot increase conditional variance, although it may leave it unchanged when future observations contain no information about a particular earlier state.

That distinction matters directly for evaluation. A historical smoothed path may be the best retrospective account of the economy, but using it as though it were available at an earlier forecast origin would reintroduce look-ahead bias.

9 Real-Time Backtesting with GDP Vintages

The purpose of evaluation is not merely to rank two RMSEs. It is to learn when a model adds information, how that conclusion depends on the target vintage and horizon, and how uncertain the apparent advantage is. A credible design must recreate the decision problem the forecaster actually faced.

9.1 Replaying the Information Set

A credible **backtest** replays the information flow. Choose a sequence of forecast origins r. At each origin:

1. build the vintage containing only values available by r;
2. apply transformations using information available by r;
3. estimate or fix parameters according to a rule chosen in advance;
4. produce and store the nowcast and its uncertainty; and
5. record the horizon relative to the target's release date.

This produces an accuracy profile by horizon, not just one error statistic. A model may be useful early in the quarter and add little near the GDP release, or the reverse.

The parameter rule in step 3 is part of the experiment. Expanding-window estimation mimics a model re-estimated using all history available at each origin; fixed-parameter evaluation isolates filtering performance but gives the historical forecaster parameter knowledge it may not have had. Either can be useful if it is labeled. Silently mixing the two makes the comparison irreproducible.

9.2 Targets, Losses, and Benchmarks

Three choices should be documented.

Evaluation target. First-release GDP measures accuracy for predicting the number initially published. A fixed later vintage measures accuracy relative to a more mature estimate of activity. Reporting both makes the distinction visible.

Loss function. The conditional mean minimizes expected squared error, which motivates root mean squared error. Mean absolute error is less sensitive to a small number of extreme quarters. If rankings change across losses, report which episodes drive the difference.

Benchmark. Useful baselines include the historical mean, a small autoregression in quarterly GDP, survey

expectations such as the Survey of Professional Forecasters, and model-based products such as GDPNow. Survey consensus and model benchmarks are different comparison classes and should be labeled accordingly.

A useful reporting table crosses these decisions rather than hiding them in a single average: target vintage by forecast horizon by loss, with the candidate model and each benchmark shown on the same origins.

9.3 Comparing Forecast Accuracy

Forecast comparisons are uncertain because the evaluation sample often contains only a few dozen quarters. A Diebold–Mariano test begins with the loss differential $d_r = L(e_{1r}) - L(e_{2r})$ and asks whether its mean is zero. A negative mean favors model 1 under the chosen loss; the test statistic divides the sample mean by an estimate of its sampling standard error.

That standard error is the difficult part. Multi-step forecast errors overlap, so adjacent d_r values can be serially correlated even when each model is well specified. The long-run variance estimate must allow that dependence. When one model nests the other, parameter estimation can shift the comparison's finite-sample distribution; a plain symmetric test may be poorly calibrated. With a small number of quarters, both the mean loss differential and its long-run variance are noisy, and one crisis can reverse the ranking.

The practical conclusion is to report the estimated difference together with an uncertainty interval or an appropriate test, inspect results by horizon and episode, and use a block bootstrap or small-sample correction when its assumptions fit the design. An estimated difference in RMSE is evidence, not a permanent league table.

9.4 The 2008Q4 Target: First Release Versus Later Vintage

The revisions to 2008Q4 GDP illustrate why the scoring target must be named. BEA reported a 3.8 percent decline in January 2009 and a 6.3 percent decline in March; the current BEA series reports an 8.5 percent decline (Federal Reserve Bank of St. Louis, 2026; U.S. Bureau of Economic Analysis, 2009a, 2009b).

Table 6 records the three estimates used in the comparison.

Table 6. Published estimates of the same quarter changed as source data and national-account information accumulated.

Vintage	U.S. real GDP growth, 2008Q4 (SAAR)
Advance estimate, January 30, 2009	−3.8%
Final estimate, March 26, 2009	−6.3%
Current estimate, 2026 vintage	−8.5%

A real-time model that predicted −4.0 percent in January 2009 used only the information then available. Its error is 0.2 percentage points if the target is the advance release and 4.5 percentage points if the target is the later vintage. Neither calculation changes the model's January information set. They measure performance against different definitions of the object to be predicted.

This episode also shows why a backtest should retain the individual errors. An average alone would not reveal whether a model missed the onset of the recession, predicted the first release well but not later revisions, or simply used a different target definition.

10 From the Kalman Filter to Dynamic Factor Models

The session began with a release that moved a GDP estimate. We can now trace that movement from the vintage ledger, through prediction and updating, into a target nowcast and an attributed revision. It is useful to close by separating what this machinery has solved from the modeling problem that remains.

10.1 What the Filter Now Produces

The release-morning calculation now has a general form:

$$\text{revision} = \text{gain} \times \text{innovation}.$$

The innovation measures what was new. The gain translates that news into a revision of the latent state. A measurement row then translates the state into the target, such as quarterly GDP. Missing observations

remove rows from the update, while a vintage ledger determines which rows were genuinely available.

For a fixed model and vintage, the complete output is more than a number: a point nowcast, a conditional uncertainty band, stored innovations and diagnostics, and a release-level news decomposition. A real-time backtest then asks how that whole procedure would have performed at past release dates.

10.2 What the Filter Does Not Choose

The main limitation is now visible. A useful nowcast rarely relies on one indicator, and indicators are correlated with one another. The filter will optimally update *given* T, Z, Q, H, but it does not decide which common economic forces should enter the state, which indicators measure them, or how a panel of hundreds of series can be estimated parsimoniously.

10.3 The Question for Session 2

Session 2 develops the dynamic factor model, which extracts a small number of common movements from a larger panel. The question changes from "how should one noisy release update a state?" to "how can many noisy, correlated releases help identify the state?" The Kalman recursions will not change. What changes is the content of the state, the structure of the measurement matrix, and the way the parameters are estimated. That is the natural next step: the updating machine is in place; we now have to build the multivariate economic model it will run.

Exercises

Exercises marked **[core]** prepare material used later in the course.

Exercise 1: Build the two-clock ledger

Problem. [core] For one monthly macroeconomic series, record the reference period, first-release date, first value, next revision date, and revised value. Explain which cells would appear in vintages dated one day before each release and one day after it. Why would storing only a single date create either misalignment or look-ahead bias?

Exercise 2: Open the release

Problem. [core, pencil] Before a release, $\alpha \sim \mathcal{N}(1.5, 0.36)$. The indicator satisfies $y = \alpha + \varepsilon$, with $H = 0.64$, and the released value is 0.5. Compute v, F, K, a^+, and P^+. Explain the result in words. Repeat for a release equal to 1.5 and distinguish the effect on the posterior mean from the effect on its variance.

Exercise 3: Reproduce the unit test

Problem. [core, computational] Implement the local level filter and reproduce the two-period numerical unit test from the notes without hard-coding any intermediate values. Add assertions for both filtered means and variances. Then replace y_2 by a missing value and verify that the second update is skipped.

Exercise 4: Signal and noise

Problem. [core, pencil] Derive the steady-state predicted variance and gain for the local level model. Compute $\bar{K}$ for $q \in \{0.01, 0.1, 1, 10\}$. For each value, state whether variation in the observed series is being attributed mainly to state movement or measurement noise.

Exercise 5: A ragged update

Problem. [core, pencil] Let $\alpha \mid \Omega_{r^-} \sim \mathcal{N}(0, 1)$, $Z = (1, 2)'$, and $H = \mathrm{diag}(1, 4)$. At vintage r, only the first observation is released and its value is 1. At the next vintage, the second observation is released with value -1. Perform the two scalar updates in release order. Identify the reference-period rows and the release dates separately.

Exercise 6: Why time alone does not narrow a fixed-target band

Problem. [core, pencil] Consider the local level model and a target h periods ahead. Show that its variance is $P_{\tau|\tau} + hQ$. Advance one period with no observation and show that the variance for the same fixed target is unchanged.

Exercise 7: Derive the quarterly weights

Problem. [core, pencil] Let a quarterly flow be the sum of three monthly levels. Use a first-order log approximation to derive the five monthly-growth weights used to map a monthly latent process into quarter-

on-quarter growth. Explain why an end-of-quarter stock variable would require a different aggregation rule.

Exercise 8: Decompose the news

Problem. [core] Using the numbers in Exercise 2, compute the revision as Kv. Then suppose a parameter is re-estimated after the release and moves the reported nowcast by another 0.1 percentage point. Write a small revision table that separates data news from parameter re-estimation.

Exercise 9: Evaluate 2008Q4

Problem. [core, data] Retrieve the GDP vintages corresponding to the advance, final, and current estimates of 2008Q4 real GDP growth. Document the retrieval dates and units. For a January 2009 nowcast of -4.0 percent, compute errors against the first release and the later vintage. State which target answers which economic question.

Exercise 10: Further state-space practice

Problem. [extra] Put an ARMA(1,1) process in state-space form and verify the dimensions of T, R, Z, Q, H. Then derive the Joseph covariance form from the shorter covariance update when K is the optimal gain.

Further Reading

- Durbin and Koopman (2012), chs. 2 and 4–7, gives a systematic treatment of filtering, **diffuse initialization**, smoothing, and likelihood evaluation.
- Harvey (1989) develops structural time series models and their economic applications.
- Giannone et al. (2008) introduces the modern mixed-frequency nowcasting framework used throughout this course.
- Bańbura et al. (2011) discusses ragged-edge data and news decompositions.
- Mariano and Murasawa (2003) derives the monthly-to-quarterly aggregation described in Section 5.2.
- Lewis et al. (2022) is the reading for Session 2 and the Weekly Economic Index practicum.

Session Glossary

The glossary collects the bold technical terms introduced in this session.

Every bold glossary term in the notes is linked to its definition below. After following a link, use the PDF viewer's Back command to return to the passage.

Backcast An estimate of a reference period that has ended but whose official target has not yet been released. A backcast uses the same machinery as a forecast or nowcast; only the target's position on the release calendar differs.

Backtest or backtesting A historical evaluation that reruns a forecasting procedure at past origins using only the information that was genuinely available then. A real-time backtest therefore needs archived vintages as well as archived outcomes.

Benchmark A comparison forecast used to judge whether a more elaborate method adds value. Useful benchmarks are credible alternatives a forecaster could actually have used, not deliberately weak straw models.

Centering The location of a residual distribution relative to zero. Innovations centered away from zero reveal a systematic tendency to predict too high or too low.

Diffuse Describes prior information that is deliberately weak or uninformative for an initial state component. A diffuse treatment is appropriate when a nonstationary level has no finite unconditional distribution.

Diffuse initialization An initialization that represents little or no reliable prior information about one or more state components. Exact diffuse methods take the limiting case analytically; a large finite covariance is an approximation.

Distributional shape Features of a distribution beyond its center and scale, including skewness, tail thickness, and outliers. Innovation shape matters because a Gaussian model assigns specific probabilities to extreme forecast errors.

Dynamic factor model A model in which many observed series load on a small number of evolving common factors. It turns a large, correlated indicator panel into a low-dimensional state-space system.

Filtered estimate An estimate of a state using observations available only through that state's date within a vintage. It is the appropriate object for reconstructing a real-time information set.

Forecast An estimate of a target whose reference period has not yet begun or ended, depending on the convention being used. In these notes, the term contrasts with a nowcast of the period under way

and a backcast of a completed but unpublished period.

Initialization The choice of the state distribution before the first filtering update. It supplies the initial mean and covariance from which every later prediction and update is recursively computed.

Innovation The observed release minus its conditional expectation just before it arrives. It is the portion of the observation that is new relative to the model's information set.

Innovation variance The model-implied conditional variance of an innovation. It combines uncertainty about the predicted state with the measurement error in the arriving observation.

Kalman gain The coefficient or matrix that maps an innovation into a revision of the estimated state. It rises when the prior state is uncertain relative to the observation and falls when the observation is noisy.

Latent Not directly observed. A latent state is inferred from its dynamics and from observed variables that measure it imperfectly.

Local level model A state-space model in which an unobserved level follows a random walk and the observed series equals that level plus measurement error. It is the simplest model that separates genuine state movement from observation noise.

Log-likelihood The logarithm of the likelihood, viewed as a function of model parameters for the observed data. Taking logs converts a product of conditional densities into a sum and is numerically easier to optimize.

Loss function A rule that assigns a cost to a forecast error. Squared loss emphasizes large misses; absolute loss is less sensitive to a small number of extreme errors.

Measurement error The component of an observation that does not reflect the state the model is trying to measure. It can include sampling error, reporting noise, and series-specific movements.

Nonstationary Describes a process whose unconditional distribution is not stable over time. A random walk is nonstationary because its variance grows with the horizon and it has no finite unconditional covariance.

Nowcast An estimate of a target for the period currently under way, made before the official target is observed. Every nowcast must be paired with an information or release date.

Nowcast commentary A plain-language account of why a nowcast changed, normally identifying each release, its model surprise, and its contribution. It is the interpretive layer of a production nowcasting system.

Prediction step The part of the Kalman recursion that carries a filtered state distribution forward through the transition equation before the next measurement is used. It propagates the mean and adds uncertainty from new state shocks.

Production report The reproducible record accompanying an operational nowcast. It should identify the vintage, target, parameters or model version, uncertainty interval, release contributions, and any re-estimation effect.

Ragged edge The staircase of missing values at the end of a real-time data panel caused by series being released on different schedules. The pattern reflects publication timing, not arbitrary deletion.

Rauch–Tung–Striebel smoother A backward recursion that revises earlier filtered state estimates using later observations in the same vintage. It is commonly abbreviated RTS smoother.

Reference period The economic week, month, or quarter described by an observation. It is distinct from the later date on which that observation is released.

Release date The calendar date and time at which an observation or revision becomes available to the forecaster. It determines whether the value may enter a particular real-time information set.

Scale The dispersion of a residual distribution. Standardized innovations with variance far from one indicate that the model's forecast uncertainty is too large or too small.

Serial dependence Predictable association between values of a series at different dates. Serial dependence in innovations means that some time-series structure remains unmodeled.

Shock A new, unpredictable disturbance that moves a state between periods. Its covariance determines how much state uncertainty accumulates during prediction.

Signal-to-noise ratio A comparison of variation attributed to the evolving state with variation attributed to measurement noise. In the local level model it is Q/H and governs how responsive the filter is to new observations.

Smoothed estimate An estimate of an earlier state that uses later observations from the full vintage. It is a hindsight estimate and therefore must not be presented as what was known in real time.

Stability Constancy of a model's predictive relationships and error behavior over time. A model can be well centered and scaled on average yet fail during a particular regime.

State The smallest collection of variables needed to summarize the model's current condition and generate the distribution of future observations. A state can contain latent factors, lags, trends, seasonal components, or time-varying coefficients.

Stationary Describes a process with a time-invariant unconditional distribution. For a stable linear state equation, stationarity permits initialization at the process's unconditional mean and covariance.

Surprise The realized value of a release minus what the model expected immediately beforehand. In a Kalman update, surprise is another name for the innovation.

Vintage A snapshot of all data values available as of a particular release date, including the missing cells and revisions known then. Two vintages may contain different values for the same reference period.

PART II

Session 2 · Dynamic Factor Models & Weekly Economic Trackers

Suppose it is Thursday morning and weekly initial unemployment claims have just been released above the model's expectation. Continued claims, released in the same report, came in close to expectation. Earlier in the week, bank credit and consumer loans moved sideways. Session 1 tells us exactly how any *one* of these releases should update an estimate of current activity. It does not tell us what to do when four of them speak at once and do not fully agree, let alone the ten weekly series behind the published Weekly Economic Index or the roughly 130 monthly series of a research panel.

The temptation is to treat the panel's correlation as a nuisance: pick the best indicator, or average everything. This session develops the opposite view. The correlation across series *is* the signal. A small number of common forces move employment, production, sales, and credit together, and a model built around that fact, the dynamic factor model, turns a wide, ragged, correlated panel into exactly the kind of state-space system whose filtering, uncertainty, and news accounting Session 1 already solved.

By the end of the session, you should be able to:

1. write a static factor model, state its assumptions, and split any series' variance into a common and an idiosyncratic share;
2. explain what a factor model does and does not identify, and choose a normalization deliberately;
3. estimate factors by principal components, explain when averaging over many series works and when correlated noise defeats it, and choose the number of factors;
4. cast a dynamic factor model in state space so that the Kalman filter delivers ragged-edge factor estimates, uncertainty bands, a likelihood, and a release-level news decomposition, and select among the three standard estimation strategies; and
5. build a weekly economic tracker from public data and criticize its design: transformation, calibration window, scaling, and failure modes.

The session assumes Session 1 in full: the state-space form, the Kalman filter and smoother, missing-observation handling, the aggregation weights, and the Gaussian conditioning lemma. It also assumes basic facts about covariance matrices. Eigenvalues appear in one specific role, and the meaning needed for that role is explained where it arises. The Practicum 1 build is used as a recurring example; its numbers are quoted as computed there.

1 From One Indicator to a Panel

1.1 Three Ways to Fail with Many Series

Session 1's filter needs a model before it can weigh evidence: given T, Z, Q, and H, the gain follows. With one indicator those matrices could be written down by hand. A production nowcast does not have one indicator. The Federal Reserve staff track hundreds of series; the FRED-MD research panel[49] curates about 130 monthly indicators for exactly this kind of work (McCracken and Ng, 2016). Before building the right model for such a panel, it is worth seeing precisely how the obvious shortcuts fail.

Use only the best series. Choosing a single indicator discards every other series' information. Session 1's machinery then applies, but the estimate inherits all of that indicator's idiosyncrasies: every strike, definitional change, or sampling problem in the chosen series moves the activity estimate one for one.

Average everything. An equally weighted average at least uses the panel, but it ignores what Session 1 taught: series differ in how much of their movement is information and how much is noise. Averaging also has no answer for the ragged edge[50]: an average over whichever series happen to be available changes its meaning from week to week.

Regress GDP on the panel. Ordinary regression asks the data to estimate one coefficient per series. With quarterly GDP as the dependent variable there are perhaps a hundred usable observations and, in a wide panel, more regressors than data points. Worse, the regressors are strongly correlated with one another, so the coefficients that fit best are erratic combinations with no stable interpretation. The regression treats the panel's correlation as collinearity: a difficulty to be fought.

All three failures point at the same missing ingredient: a model of *why* the series are correlated.

1.2 Comovement Is the Signal

The empirical fact that rescues the wide panel is **comovement**: when aggregate activity turns, employment, production, income, sales, freight, and credit turn together, not identically and not simultaneously to the week, but visibly and persistently. Burns and Mitchell (1946) built the NBER's business-cycle chronology on that observation long before it had a statistical form. Sargent and Sims (1977) gave it one, showing that a small number of common "index" processes account for the bulk of the covariation among U.S.

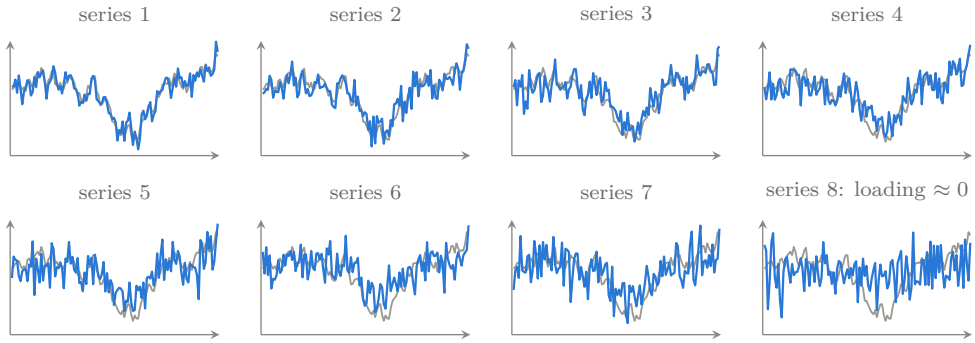
[49] FRED-MD is a public, monthly-updated panel of roughly 130 monthly U.S. macroeconomic series, distributed with standard transformation codes so that factor-model results can be replicated and compared across papers.

[50] The ragged edge, from Session 1, is the staircase of missing values at the end of a real-time panel caused by series being released on different schedules.

aggregates. The modern statement is blunt: the business cycle is low-dimensional. A handful of common forces, often a single one, drive the shared movement of everything we can measure.

Figure 9 shows the claim in stylized form. Eight standardized series share one underlying activity factor; each adds its own noise. No single series is a clean reading, yet the common swing is unmistakable across the panel, except in the last series, which barely participates.

Figure 9. Comovement in a stylized panel. Eight standardized series generated from one common factor plus independent noise. The shared cycle is visible in every panel but the last, whose loading is near zero: that series is almost entirely its own story.



If comovement is real, then the correct response to "too many series" is not selection but aggregation with a model: treat every series as a noisy witness to a small common state, and let the model decide how much each witness' testimony is worth. You have already seen this succeed in miniature. The Practicum 1 build combined four public weekly series into one factor and reproduced the official ten-series Weekly Economic Index with correlation 0.90. The rest of this session builds the machinery behind that result, and explains the episodes where the four-series version failed.

The first step is the smallest model in which cross-series correlation is signal rather than nuisance.

2 The Static Factor Model

2.1 One Equation, Two Kinds of Variation

Let x_τ be an $N \times 1$ vector of observed series for reference period τ, each already transformed to be standardized[51] and stationary. The **static factor model** writes each series as a linear combination of

[51] A standardized series has had its sample mean subtracted and been divided by its sample standard deviation, so it is measured in standard-deviation units with mean zero. Which sample supplies the mean and standard deviation turns out to

$k \ll N$ common factors plus a series-specific remainder:

$$x_\tau = \Lambda f_\tau + e_\tau. \tag{25}$$

Read Equation 25 one object at a time. The $k \times 1$ vector f_τ collects the common factors: latent forces, in the sense of Session 1's states, that many series feel at once. The $N \times k$ matrix Λ holds the **factor loadings**; its ith row λ_i' says how strongly, and with what sign, series i responds to each factor. The remainder e_τ is the **idiosyncratic component**: sampling error, sector-specific shocks, definitional quirks: everything in series i that is not broadly shared. "Static" refers to the equation's form, not to the data: the factors may be highly persistent, but each period's observation depends only on that period's factor.

The model's assumptions parallel Session 1's, and it is worth stating them with the same care.

B1 (Orthogonality). Factors and idiosyncratic components are uncorrelated with each other at all leads and lags. This is what makes the split in Equation 25 meaningful: it lets every covariance between two different series be carried by the factors, and it is used each time we decompose a variance or compute a factor-extraction weight. If a "series-specific" shock in fact moves the factor, the accounting between common and idiosyncratic variation breaks down.

B2 (Weak idiosyncratic dependence). Idiosyncratic components may be correlated across series, but only weakly: no group of series shares enough leftover comovement to imitate a factor. The special case with fully uncorrelated idiosyncratics is the *exact* factor model; the realistic case, which tolerates limited cross-correlation, is the **approximate factor model** of Bai and Ng (2002). The assumption is used by every argument that averages noise away across series. It is genuinely restrictive: initial and continued unemployment claims share labor-market reporting noise that no factor should be forced to absorb, and Section 4 quantifies what such shared noise costs.

B3 (Pervasiveness). Each factor loads on a non-vanishing fraction of the series: as the panel grows, $\Lambda'\Lambda/N$ stays bounded away from zero. This is what makes a factor recoverable from averages: a "factor" visible in only three series of a hundred cannot be separated from their idiosyncratic noise. It is used in the consistency results of Section 5.

Together, B1–B3 say what the words "common" and "idiosyncratic" mean, and they are diagnostic claims as much as conveniences: loadings and residual correlations estimated later will tell us when they are inadequate.

matter; Section 9 returns to that choice.

2.2 The Common Share

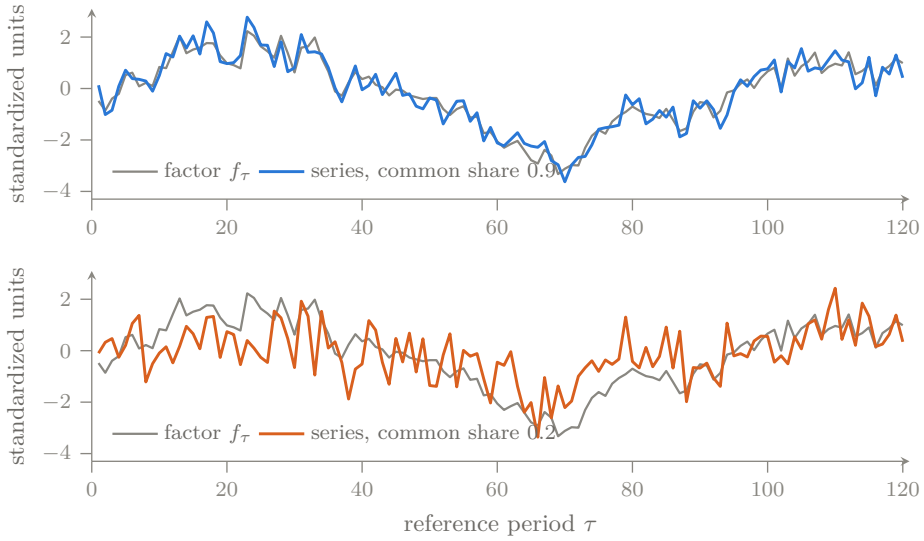
B1 makes variances add. For a standardized series i with $\mathrm{Var}(f_\tau) = I_k$,

$$\underbrace{1}_{\text{total}} = \underbrace{\|\lambda_i\|^2}_{\text{common}} + \underbrace{\sigma^2_{e,i}}_{\text{idiosyncratic}} . \tag{26}$$

The quantity $\kappa_i = \|\lambda_i\|^2$ is series i's **common share**: the fraction of its variance explained by the common factors. It is the single most useful summary of an indicator's value to a factor model. A series can be individually fascinating (much discussed, promptly released, economically interpretable) and still be nearly useless here, because most of its movement is its own.

Figure 10 makes the number visible. Two series track the same factor; one has a common share of 0.9, the other 0.2. Knowing only the paths, a reader would call the first "informative" and the second "noisy," which is exactly what the decomposition formalizes.

Figure 10. What the common share looks like. Both series load on the same factor (gray). With a common share of 0.9, the upper series is nearly a clean reading. With 0.2, the lower series contains the factor, but most of its week-to-week movement is idiosyncratic.



2.3 What the Model Says About Data

A latent-variable equation cannot be checked directly, so it is worth asking what Equation 25 implies about observable quantities. Take the covariance matrix of the panel under B1:

$$\Sigma = \mathrm{Var}(x_\tau) = \Lambda\Lambda' + \Psi, \qquad \Psi = \mathrm{Var}(e_\tau). \tag{27}$$

Read Equation 27 as the model's entire observable content. Every covariance between two *different* series is carried by the low-rank piece $\Lambda\Lambda'$: for an exact factor model with one factor, $\mathrm{Cov}(x_{i\tau}, x_{j\tau}) = \lambda_i\lambda_j$ for $i \neq j$. The factor model is therefore a restriction on a covariance matrix: the claim that $N(N-1)/2$ distinct cross-series covariances can be reproduced by Nk loadings. That is a strong compression for large N, and it is testable: residual covariances that the factors cannot explain are evidence against B2, in the same way that autocorrelated innovations were evidence against Session 1's assumptions.

Table 7 collects the notation used in this session. As in Session 1, each symbol is introduced again where it does its work.

Table 7. Notation for Session 2. The state-space symbols and the vintage superscript keep their Session 1 meanings; the number of factors is written k throughout to avoid colliding with the release date r.

Symbol	Meaning
x_τ	$N \times 1$ vector of standardized observed series for reference period τ
$N,\ n$	Number of series (cross-section) and number of time periods
f_τ	$k \times 1$ vector of common factors; k is the number of factors
$\Lambda,\ \lambda_i'$	$N \times k$ loading matrix and its ith row
$e_\tau,\ \Psi$	Idiosyncratic components and their covariance matrix
κ_i	Common share of series i; $\kappa_i = \lVert\lambda_i\rVert^2$ for standardized data
$A_1, \ldots, A_p,\ u_\tau,\ \Sigma_u$	Factor VAR coefficients, factor shocks, and their covariance
$\rho_i,\ \varepsilon_{i\tau},\ \sigma_i^2$	Idiosyncratic AR(1) coefficient, shock, and shock variance
w	Cross-sectional weights applied to the panel to extract a factor

Symbol	Meaning
$\hat{\mu}_1 \geq \hat{\mu}_2 \geq \cdots$	Eigenvalues of the sample covariance matrix of the panel
$T, R, Z, Q, H,\ \alpha_\tau$	The state-space system and state vector, exactly as defined in Session 1
r	Release date defining a data vintage, exactly as in Session 1

The model now defines what a factor is. It does not yet say *which* factor: the same data admit many loading–factor pairs, and the next section separates what is identified from what is chosen.

3 Rotation: What the Data Cannot Decide

3.1 Scale and Sign First

Start with one factor, because the whole problem is already visible there. If (λ_i, f_τ) reproduces the data, so does $(c\lambda_i,\ f_\tau/c)$ for any $c \neq 0$: double every loading and halve the factor, and every product $\lambda_i f_\tau$ is unchanged, and with it every fitted value and every covariance. Taking $c = -1$ flips the factor's sign and every loading's sign at once. Nothing observable distinguishes these representations. The factor's scale and sign are not properties of the data; they are conventions the modeler must supply.

Figure 11 displays three of these representations. The fitted series is identical in all three panels. The "factor" is not.

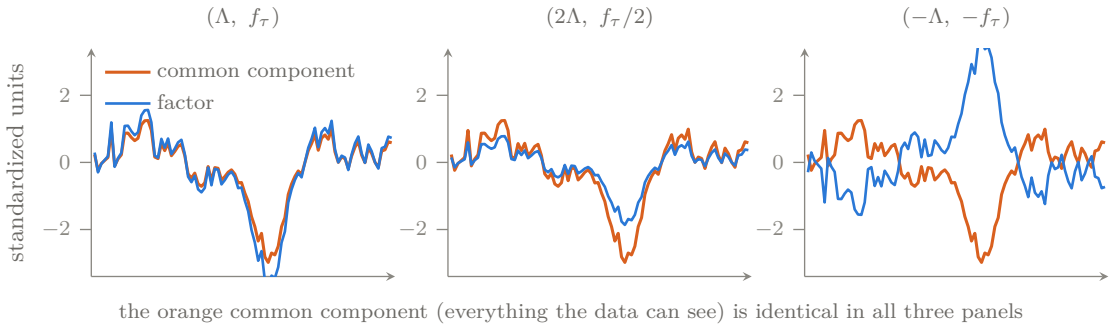
3.2 The General Statement

With k factors the ambiguity is larger. For any invertible $k \times k$ matrix M,

$$\Lambda f_\tau = (\Lambda M^{-1})(M f_\tau),$$

so the pair $(\Lambda M^{-1},\ M f_\tau)$ fits identically. The factors are identified only up to an invertible transformation: in the factor-analysis tradition, a **rotation**, though M need not be a rotation in the geometric sense. What

Figure 11. Identical fit, different factors. The pairs (Λ, f), $(2\Lambda, f/2)$, and $(-\Lambda, -f)$ generate exactly the same observed panel. The data cannot choose among them; a normalization must.



is identified is the factor space: the set of linear combinations of the series that the factors span, and therefore the common component Λf_τ of each series.

This is not a defect to engineer away. It is a fact to manage with a **normalization**: a set of restrictions that selects one representative from each equivalence class. The standard choice, the one principal components imposes automatically, is $\mathrm{Var}(f_\tau) = I_k$ with $\Lambda'\Lambda$ diagonal, plus a sign convention for each factor. The Practicum 1 build's sign rule, "the factor correlates positively with activity," is exactly such a convention.

Two consequences deserve to be stated plainly, because both change how results are reported:

1. **Factors have no intrinsic units.** "The factor fell by 0.3" means nothing until the factor is scaled against something interpretable. This is why the P1 build regressed GDP growth on its factor before publishing a number, and why Section 9 treats that scaling step as part of the model.
2. **Interpretation is earned, not derived.** Calling the first factor "real activity" is justified by inspecting its loadings and its behavior over known episodes, not by the algebra, which would be equally satisfied by the factor's negative.

With a normalization fixed, the model is finally specific enough to use. The natural first question is the Session 1 question: given the model's parameters, how should the data revise our estimate of the latent factor?

4 The Factor as a Weighted Testimony

4.1 Two Witnesses, One Update

Take the smallest interesting case: one factor, two standardized series, and loadings treated as known. Let

$$f \sim \mathcal{N}(0,1), \qquad x_i = \lambda_i f + e_i, \quad \lambda_1 = \lambda_2 = 0.8, \quad \sigma_{e,1}^2 = \sigma_{e,2}^2 = 0.36,$$

with independent idiosyncratic noise, so each series has common share $\kappa = 0.64$ by Equation 26. Suppose the two series are observed at $x_1 = -1.5$ and $x_2 = -0.5$: one witness reports conditions well below normal, the other only mildly below.

This is a job for Session 1's Gaussian conditioning lemma, with f in the role of the unobserved block. The joint distribution implied by the model is

$$\begin{pmatrix} f \\ x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.64 \\ 0.8 & 0.64 & 1 \end{pmatrix} \right).$$

The blocks come straight from Equation 27: each series has unit variance, their covariance is $\lambda_1\lambda_2 = 0.64$, and each covaries with the factor by its loading. Applying the lemma, the weight vector is $w' = \Sigma_{fx}\Sigma_{xx}^{-1}$, and the arithmetic gives

$$w_1 = w_2 = \tfrac{20}{41} \approx 0.488, \qquad \mathbb{E}[f \mid x_1, x_2] = \tfrac{20}{41}(-1.5) + \tfrac{20}{41}(-0.5) = -\tfrac{40}{41} \approx -0.976, \tag{28}$$

with conditional variance

$$\mathrm{Var}(f \mid x_1, x_2) = 1 - w'\Sigma_{xf} = \tfrac{9}{41} \approx 0.220.$$

Compare the one-witness case. Observing only $x_1 = -1.5$ gives gain $\lambda_1/(\lambda_1^2 + \sigma_{e,1}^2) = 0.8$, hence $\mathbb{E}[f \mid x_1] = -1.2$ and conditional variance 0.36. The second witness did two things. It pulled the estimate up, from -1.2 to -0.976, because its milder reading is evidence that part of the first series' plunge was idiosyncratic. And it cut the conditional variance from 0.36 to 0.220: two noisy witnesses, properly weighted, are more precise than one.

4.2 The Precision View

The same update can be computed a second way, and the second way is the one that scales. With independent idiosyncratic noise the posterior precision[52] adds one term per witness:

$$\mathrm{Var}(f \mid x_1, \dots, x_N)^{-1} = 1 + \sum_{i=1}^{N} \frac{\lambda_i^2}{\sigma_{e,i}^2}, \qquad \mathbb{E}[f \mid x] = \mathrm{Var}(f \mid x) \sum_{i=1}^{N} \frac{\lambda_i}{\sigma_{e,i}^2}\, x_i. \tag{29}$$

Read Equation 29 as a courtroom rule. Each witness contributes information $\lambda_i^2/\sigma_{e,i}^2$: testimony counts for more when the witness is more exposed to the truth (large loading) and less erratic (small idiosyncratic variance). Each observed value enters the estimate weighted by $\lambda_i/\sigma_{e,i}^2$, so a series with a negative loading correctly enters with a negative weight, and a series with a near-zero loading is correctly ignored no matter how dramatic its movements. For the numbers above, the posterior precision is $1 + 2(0.64/0.36) = 41/9$, reproducing $\mathrm{Var} = 9/41$ exactly.

Extraction, in other words, is a cross-sectional regression: at each date, the panel is regressed on the loadings, with each series weighted by its precision. The time dimension plays no role yet; one cross-section is one sample.

4.3 More Witnesses, and the Blessing of Dimensionality

Now let the panel grow. With N identical independent witnesses, Equation 29 gives

$$\mathrm{Var}(f \mid x_1, \dots, x_N) = \frac{1}{1 + N\lambda^2/\sigma_e^2} \;\longrightarrow\; 0 \quad \text{as } N \to \infty.$$

Every additional series is another vote, its idiosyncratic error independent of the others, and the noise averages away. In a wide enough panel the factor stops being latent in any practical sense: the cross-section alone pins it down. This is the **blessing of dimensionality**: the reversal by which the panel's width, the very thing that broke regression in Section 1, becomes the source of statistical power. It is also the intuition behind the consistency results for principal components in Section 5, which make the argument without assuming the loadings are known.

The blessing has a condition attached, and it is B2. Suppose the witnesses are not independent: each pair of idiosyncratic components shares correlation ρ_e[53]. Averaging then removes only the independent part of

[52] Precision is the reciprocal of a variance. Precisions of independent sources of information add, which is why it is the natural bookkeeping unit when evidence accumulates.

[53] Equicorrelation assigns the same correlation to every pair. It is the cleanest way to model a sector-wide disturbance such

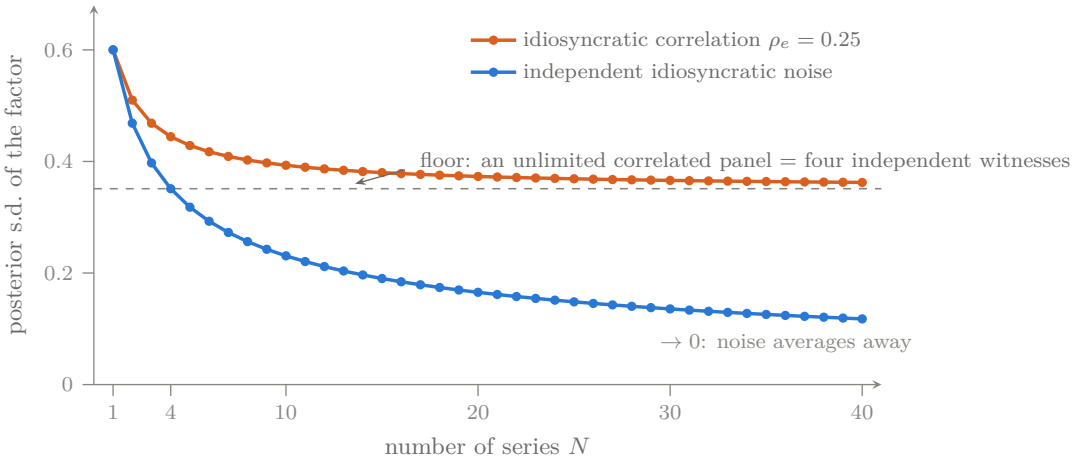
the noise. The average of N such witnesses has idiosyncratic variance $\sigma_e^2\,[(1-\rho_e)/N+\rho_e] \to \sigma_e^2\rho_e$, so the posterior variance no longer goes to zero:

$$\mathrm{Var}(f \mid x_1, \ldots, x_N) \;\longrightarrow\; \frac{1}{1+\lambda^2/(\sigma_e^2\rho_e)} \quad \text{as } N \to \infty. \tag{30}$$

The limit is worth restating without symbols: **an unlimited supply of witnesses from one sector is worth exactly $1/\rho_e$ independent witnesses.** With $\rho_e = 0.25$, an infinite panel of correlated series carries the same information as four clean ones; for the anchor values $\lambda = 0.8$, $\sigma_e^2 = 0.36$, both give posterior variance $9/73$. This is the formal content of a lesson the P1 journal learned empirically: a panel of ten series from three sectors is not a ten-witness panel, and adding a fifteenth series from an already-crowded sector adds almost nothing. Boivin and Ng (2006) document the practical force of this point: larger panels assembled without regard to idiosyncratic correlation can *worsen* factor estimates.

Figure 12 traces both curves.

Figure 12. Why breadth beats depth. Posterior standard deviation of the factor as series are added, for the anchor values $\lambda = 0.8$, $\sigma_e^2 = 0.36$. Independent witnesses drive uncertainty toward zero. Witnesses whose idiosyncratic noise is equicorrelated at $\rho_e = 0.25$ hit a floor equal to four independent witnesses, no matter how many are added.



Everything in this section conditioned on known loadings. In practice Λ must be estimated from the same panel, and the natural estimator turns out to implement precisely this averaging logic.

as reporting noise or a sectoral shock that touches every series in the group equally.

5 Principal Components

5.1 The Least-Squares Problem

Stack the standardized panel as an $n \times N$ matrix X, one row per period, one column per series. If both the loadings and the factor path are unknown, the least-squares approach chooses them to fit the panel as closely as possible:

$$\min_{\{\lambda_i\},\{f_\tau\}} \sum_{i=1}^{N} \sum_{\tau=1}^{n} \left(x_{i\tau} - \lambda_i' f_\tau\right)^2 , \tag{31}$$

subject to the normalization of Section 3, without which the problem has no unique solution. This is **principal components** in its regression form. For one factor the solution has a clean characterization: the optimal weights w form the first eigenvector[54] of the sample covariance matrix of the panel, and the estimated factor is the weighted cross-sectional average

$$\hat{f}_\tau = w' x_\tau . \tag{32}$$

The derivation is one concentration step: for a fixed factor path, the best loading for each series is an ordinary regression coefficient; substituting it back turns Equation 31 into the problem of choosing the single direction in series-space that captures the most panel variance, which is what the first eigenvector is. In computation the whole object comes at once from the singular value decomposition[55] of X: the first right singular vector is w, and the share of panel variance the factor explains is $\hat{\mu}_1 / \sum_j \hat{\mu}_j$. Further factors, if wanted, are the subsequent eigenvectors, orthogonal by construction: the normalization $\Lambda'\Lambda$ diagonal made real.

Notice what Equation 32 is. It is a weighted average of witnesses, with the structure of Equation 29, and with weights estimated from the panel's own covariances rather than supplied by hand. A series that comoves strongly with the rest of the panel earns a large weight; a series that moves alone earns a weight near zero. The estimator discovers the courtroom rule.

[54] An eigenvector of a matrix is a direction the matrix stretches without turning; the corresponding eigenvalue is the stretch factor. For a covariance matrix, the first eigenvector is the direction of maximum variance, and its eigenvalue is the variance in that direction.

[55] The singular value decomposition factors any matrix as $X = USV'$ with orthonormal U and V and nonnegative diagonal S. The columns of V are the eigenvectors of $X'X$, so the first column supplies the principal-component weights, and the squared singular values are proportional to the eigenvalues $\hat{\mu}_j$.

5.2 The P1 Build as a Worked Instance

The Practicum 1 instructor build runs exactly this computation on four public weekly series (each transformed and standardized; the transformation is the subject of Section 9). Its estimated weights, computed on the pre-2020 calibration sample, are worth reading like regression coefficients:

Table 8. Principal-component weights from the P1 instructor build, estimated on the pre-2020 sample. The first component explains 48 percent of the calibration panel's variance.

Series	Weight	Reading
Initial claims (sign-flipped)	0.69	High common share; carries the factor
Continued claims (sign-flipped)	0.69	High common share; carries the factor
Consumer loans	0.20	Weakly cyclical in year-over-year terms
Bank credit	−0.08	Nearly acyclical; correctly ignored

The near-zero weight on bank credit is not a failure. Banks' balance sheets *expand* in panics as firms draw committed credit lines, so the series is nearly acyclical in year-over-year terms, and the estimator, seeing no comovement with the rest of the panel, correctly declines to use it. A "wrong" loading is information about the data; listen to it before deleting the series. The concentration of weight in the two claims series is also informative, and less comfortable: it says the four-series factor is mostly a labor-market factor. That observation returns with force in Section 9.

5.3 Why Averaging Works Without Knowing the Loadings

Two classical results turn principal components from a heuristic into an estimator with guarantees. First, **consistency in both dimensions**: as N and n grow, the estimated factor space converges to the true one, and the convergence survives the weak idiosyncratic correlation of B2 (Bai and Ng, 2002; Stock and Watson, 2002). The mechanism is the blessing of dimensionality from Section 4: each series is another noisy vote, and pervasiveness (B3) guarantees the votes keep coming. Second, **a rule for choosing k**: the

information criteria[56] of Bai and Ng (2002) penalize additional factors at a rate calibrated to the panel's two dimensions, so minimizing them estimates the number of factors consistently.

Practice adds two amendments. The criteria are the referee, but parsimony is the prior: in U.S. aggregate data, one or two factors carry the business cycle, and a marginal factor tends to pick up sectoral detail: a second factor often loads on housing or finance. And no criterion substitutes for inspecting the scree plot[57] and the loadings: a factor whose loadings have no economic pattern is usually a symptom of B2 failing, not a discovery.

5.4 What Principal Components Cannot Do

Principal components is static and complete-data, and both words are load bearing. It has no notion of factor *dynamics*: yesterday's factor does not inform today's estimate, there are no forecasts, and persistence in the data is wasted. It has no native treatment of *missing observations*: the SVD needs a balanced panel[58], so the ragged edge, the defining feature of real-time data, forces either deleting incomplete rows or imputing them. And it produces no *uncertainty bands*: the factor estimate arrives without a conditional variance, so there is nothing to widen when data are thin and nothing from which to compute news weights.

All three absences point in the same direction. Dynamics, missing data, and uncertainty are exactly what Session 1's state-space machinery provides. The factor model must be put in state space.

6 The Dynamic Factor Model

6.1 Dynamics for the Factors, and for the Noise

A **dynamic factor model** keeps the measurement idea of Equation 25 and adds time-series structure to both unobserved pieces:

[56] An information criterion scores a model by its fit minus a penalty for its size. The Bai and Ng (2002) criteria calibrate that penalty to the double limit $N, n \to \infty$, which ordinary AIC- or BIC-style penalties do not handle.

[57] A scree plot displays the ordered eigenvalues $\hat{\mu}_1 \geq \hat{\mu}_2 \geq \cdots$. A sharp drop after the kth eigenvalue suggests k factors; the plot is suggestive rather than decisive, which is why the formal criteria exist.

[58] A balanced panel has an observation for every series in every period. Real-time panels are unbalanced by construction because of the ragged edge.

$$
\begin{aligned}
x_\tau &= \Lambda f_\tau + e_\tau, \\
f_\tau &= A_1 f_{\tau-1} + \cdots + A_p f_{\tau-p} + u_\tau, \quad u_\tau \sim \mathcal{N}(0, \Sigma_u), \\
e_{i\tau} &= \rho_i \, e_{i,\tau-1} + \varepsilon_{i\tau}, \qquad\qquad \varepsilon_{i\tau} \sim \mathcal{N}(0, \sigma_i^2),
\end{aligned}
\tag{33}
$$

with the $\varepsilon_{i\tau}$ independent across series. The factor equation is a vector autoregression[59] capturing the business cycle's persistence: this month's activity predicts next month's. The idiosyncratic equations give each series' private component its own persistence: a strike or a definitional break does not vanish in one week. Setting $k = 1$, $p = 1$, and $\rho_i \neq 0$ already produces a serious nowcasting model.

Why must the idiosyncratic dynamics be modeled at all? Because Session 1's assumptions demand it. If $e_{i\tau}$ is serially correlated and left in the measurement error, assumption A2 (serially independent measurement noise) fails, and the filter's innovations remain forecastable. Session 1's discussion of that assumption already named the repair: persistence in the errors "should be modeled as additional states." The dynamic factor model takes its own advice.

6.2 The State-Space Form

Stack the factor and every idiosyncratic component into the state. For the one-factor, first-order case,

$$
\alpha_\tau = \begin{pmatrix} f_\tau \\ e_{1\tau} \\ \vdots \\ e_{N\tau} \end{pmatrix}, \qquad T = \begin{pmatrix} a & & & \\ & \rho_1 & & \\ & & \ddots & \\ & & & \rho_N \end{pmatrix}, \qquad Z = \begin{pmatrix} \Lambda & I_N \end{pmatrix}, \qquad H = 0,
\tag{34}
$$

with $R = I_{N+1}$ and $Q = \operatorname{diag}(\sigma_u^2, \sigma_1^2, \ldots, \sigma_N^2)$. Check the dimensions before reading further: the state has $N + 1$ elements, T is $(N + 1) \times (N + 1)$ and diagonal, and Z is $N \times (N + 1)$: each measurement row picks up the factor through its loading and its own idiosyncratic state through the identity block. A factor VAR(p) replaces the single a with a $kp \times kp$ companion block, exactly as in Session 1's AR(p) construction.

The striking entry is $H = 0$: the measurement equation is *exact*. Nothing has been swept under the rug: the measurement noise has been promoted into the state, where the filter can model its persistence instead of assuming it away. The innovations remain perfectly well defined, because uncertainty about the state still flows into every predicted observation through Z.

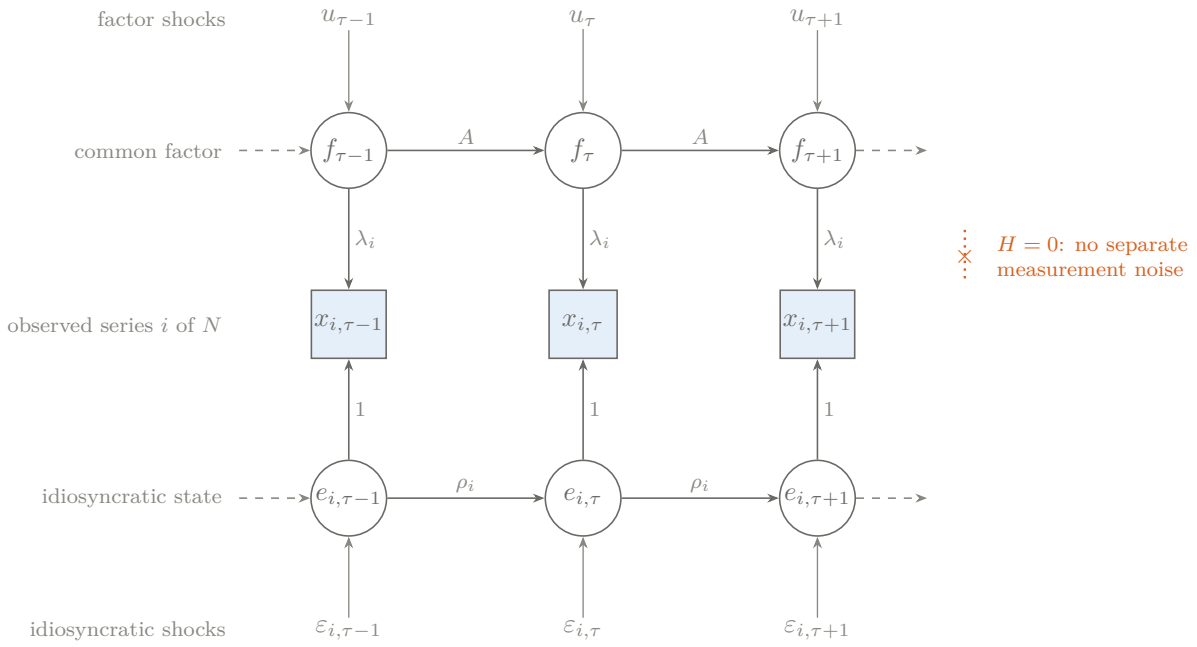
[59] A vector autoregression (VAR) predicts a vector of variables from p of its own lags. It is the multivariate analogue of the AR(p) process put in companion form in Session 1.

Once Equation 33 is written as Equation 34, everything Session 1 built applies without modification, and it is worth an explicit inventory. One forward pass of the Kalman filter now delivers:

- the factor estimate $\mathbb{E}[f_\tau \mid Y_\tau]$ *with a conditional variance*, the uncertainty band principal components could not produce;
- forecasts of the factor, and hence of every series, through the factor VAR;
- the likelihood, evaluated from the innovations exactly as in Session 1;
- native ragged-edge handling: missing observations delete measurement rows through the selection matrix $W_\tau^{(r)}$, with no imputation and no deleted history; and
- mixed frequency: append monthly latent GDP growth driven by the factor, give quarterly GDP the five-weight aggregation row that Session 1 derived, and the same filter produces a quarterly GDP nowcast from a monthly panel. That combination of factor structure, the aggregation row, and ragged-edge filtering is, in structure, the nowcasting model of Giannone et al. (2008).

Figure 13 draws the dependence structure. Compare it with Session 1's state-space graph: the new feature is that what was a single measurement-error arrow is now a persistent chain of its own inside the state.

Figure 13. The dynamic factor model as a state-space system. The common factor (top chain) and each idiosyncratic component (lower chains) evolve as persistent states. Each observed series combines the factor, through its loading, with its own idiosyncratic state. The measurement equation is exact: $H = 0$ because all noise now lives in the state.



One practical warning belongs next to the elegance. The state has $N + kp$ elements and the filter's cost

grows with the cube of the state dimension, so carrying every idiosyncratic component of a 130-series panel through Equation 34 is expensive. Production systems exploit the structure (the diagonal blocks in T, or sequential univariate processing of the measurement rows) or keep only the persistent idiosyncratics in the state and leave approximately white ones in H. The modeling logic is unchanged; only the linear algebra is arranged more carefully.

The system matrices T, Z, Q now contain unknown loadings, VAR coefficients, and variances. Session 1 assumed such parameters known; a factor model for a real panel cannot.

7 Estimating the Model on a Ragged Panel

Three estimation strategies are standard. They differ in how much of the model they use, and the practical craft is knowing which is sufficient for the task at hand.

Principal components alone. Estimate $\hat{\Lambda}$ and $\hat{f}_\tau$ by the SVD of Section 5 and stop. Fast, transparent, and reproducible in a dozen lines, but static, balanced-panel only, and silent about uncertainty. It is the right tool for a first look, for historical analysis on complete data, and for a tracker whose weekly panel is nearly balanced. This is the Practicum 1 build.

Two-step. The **two-step estimator** of Doz et al. (2011) runs principal components on the balanced part of the panel to obtain $\hat{\Lambda}$ and the factor path, fits the factor VAR and idiosyncratic autoregressions to those estimates by least squares, then *fixes all parameters* and runs the Kalman filter and smoother of Session 1 over the full ragged panel. One pass, no iteration. The first step supplies the parameters; the second supplies everything principal components lacked: bands, forecasts, ragged-edge estimates, news weights. In realistic designs the efficiency loss from not iterating is small, which is why the two-step is the workhorse of daily production systems: fast, stable, and easy to audit.

Maximum likelihood by EM. The **EM estimator** treats the factor path as missing data and iterates two steps until the likelihood converges: an E-step that runs the Kalman smoother to compute expected states and their second moments given the current parameters, and an M-step that re-estimates every parameter by regressions on those smoothed moments (Bańbura and Modugno, 2014; Doz et al., 2012). Each iteration weakly increases the likelihood. The estimator handles *arbitrary* missingness (mixed frequencies, ragged edges, series that begin late or end early) inside estimation itself, not only in filtering, and it delivers the full likelihood machinery of Session 1. The price is implementation care: convergence is monotone but can

be slow, and the likelihood surface contains the rotation ridge of Section 3, so the normalization must be imposed before any parameter is interpreted. Because the model is estimated by Gaussian methods without the belief that every series is exactly Gaussian, the approach is described as quasi-maximum likelihood[60]. This is the standard behind serious central-bank nowcasting systems.

Table 9. Three estimators for the dynamic factor model. Each row uses strictly more of the model than the row above and costs correspondingly more.

Strategy	Uses dynamics	Handles ragged edge	Uncertainty bands	Cost	Right tool when
Principal components	no	no	no	seconds	balanced panel, first look, simple tracker
Two-step	yes	in filtering	yes	one filter pass	daily production; parameters revised rarely
EM (quasi-ML)	yes	in estimation	yes	many smoother passes	arbitrary missingness; full-system estimation

The course follows the table downward. Practicum 1 used principal components; this session's exercises implement the two-step on a ragged panel; the capstone uses the two-step or EM as ambition allows.

8 News, Now Operational

Session 1 derived the news decomposition in general form: a nowcast revision equals $\sum_i w_i v_i$, surprise times weight, release by release. In the dynamic factor model that formula stops being abstract and becomes a daily production report, because the model finally says where the weights come from. The weight on release i is built from the Kalman gain of the estimated system, and the structure of Equation 33 makes

[60] A quasi-maximum likelihood estimator maximizes a likelihood written under convenient distributional assumptions that need not hold exactly. Under conditions given by Doz et al. (2012), the factor estimates remain consistent even when the Gaussian assumption is only an approximation.

its determinants legible: a release earns a large weight when its common share is high (its surprise says a lot about the factor), when it arrives early relative to other information about the same period (there is still factor uncertainty left to resolve), and when its idiosyncratic variance is small (the surprise is unlikely to be private noise).

That reading turns the decomposition into a design tool as well as a reporting tool. If the weekly claims release dominates every news report, that is not an accident of the week; it is the model reporting that claims combine high common share with first-in-the-week timing. And the caution from Session 1 carries over with more force: the model's expectation is not the survey consensus. A release can beat the consensus and still fall short of the model's forecast, in which case a correct news report says the release *lowered* the nowcast. A production commentary must say which expectation defines its surprises.

With estimation and news in place, the machinery is complete. What remains is the published object this session has been circling: the weekly tracker.

9 Anatomy of a Weekly Tracker

Lewis et al. (2022) is the cleanest published specimen of a **weekly economic tracker**: the Weekly Economic Index (WEI), updated every Thursday and now published by the Federal Reserve Bank of Dallas. It is the course's replication target because every design choice in it is visible, small enough to rebuild, and instructive when it fails. Four choices define it.

9.1 Choice 1: Ten Series, Five Sectors

Table 10. The ten weekly inputs of the Weekly Economic Index, grouped by sector (Lewis et al., 2022).

Sector	Series
Labor	Initial claims; continued claims; staffing index
Consumption	Redbook same-store retail sales; consumer confidence
Production	Raw steel production; electricity output
Shipping and energy	Rail carloads; fuel sales
Taxes	Federal tax withholding

Breadth across sectors is the design principle, and Section 4 says why: series within a sector share idiosyncratic noise, so a second sector buys more precision than a fifth series from the first. The P1 build is the controlled demonstration. Its four series span only labor and credit, and its factor is, by Table 8, essentially a claims factor. On the common sample it still tracks the WEI at correlation 0.90; the blessing of dimensionality is generous. But its single largest divergence, 11.4 percentage points below the WEI in the week of March 28, 2020, is exactly the week when a claims catastrophe was deeper than the economy-wide catastrophe. The official index's retail, fuel, and electricity series pulled its reading up; the four-series panel had no other sector to consult. Breadth is variance insurance, and the premium comes due in the weeks that matter most.

9.2 Choice 2: 52-Week Differences

Weekly data carry ferocious seasonality: holidays, school calendars, weather, and week-ending conventions, none of it aligned across series. The WEI removes it by transformation: every input enters as its 52-week difference (in logs where appropriate),

$$\tilde{x}_{i\tau} = \ln x_{i\tau} - \ln x_{i,\tau-52}, \tag{35}$$

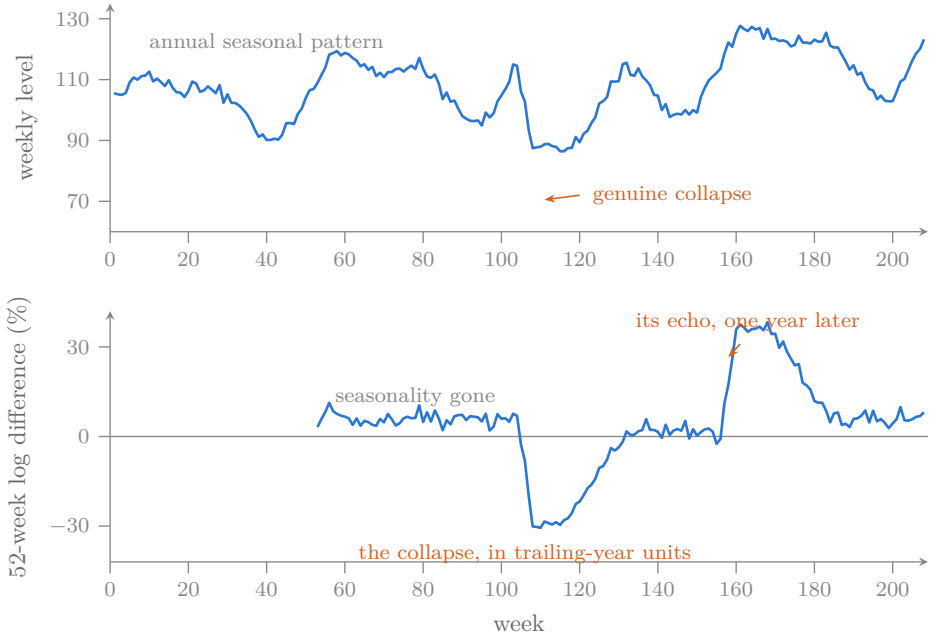
the growth of each week over the same week one year earlier. Any seasonal pattern that repeats annually cancels in the subtraction, with no seasonal model estimated and no adjustment procedure to maintain.

The price has two parts, and both are visible in Figure 14. First, the units change: the transformed series measures *trailing-year* growth, so the index is smoother and slower than the underlying weekly pulse: a genuine one-week collapse enters gradually and lingers. Second, every extreme week returns. Because week $\tau - 52$ sits in the denominator of this year's comparison, the collapse of April 2020 reappears, sign-flipped, as a spurious surge in April 2021. This is the **base effect**: a movement in a year-over-year series driven by the comparison week a year earlier rather than by anything current. A tracker's user must know the difference between news and an anniversary.

The transformation, in other words, is a seasonal adjustment in disguise: one that works only for patterns that repeat on a fixed 52-week cycle. Floating holidays drift against week numbers, and every few years a 53rd week[61] breaks the alignment entirely. What the transformation assumes implicitly, Session 3 will model explicitly.

[61] Because 52 weeks are 364 days, calendar years contain 52 weeks and a fraction; week-numbering conventions insert a 53rd week roughly every five to six years, misaligning "the same week last year" for every series at once.

Figure 14. What the 52-week difference removes, and what it costs. Top: a raw weekly series with a repeating annual pattern and a genuine mid-sample collapse. Bottom: the 52-week log difference. The seasonal pattern cancels without any model, but the collapse now enters in trailing-year units, and it echoes with the opposite sign exactly one year later, when the collapsed weeks rotate into the comparison base.



9.3 Choice 3: One Factor, Scaled to GDP

The WEI extracts a single principal component from the ten transformed series and then confronts the rotation problem of Section 3 head-on: a factor has no units. The solution is a scaling regression: regress four-quarter GDP growth on the quarterly-averaged factor and report the fitted value, so that a WEI reading of 2 can be read as "consistent with GDP growth of about 2 percent over the trailing year." The P1 build does the same and estimates, on its calibration sample, tracker $= 2.05 + 0.90 \times$ factor with a quarterly R^2 of 0.63. The regression is not decoration; it is the final, essential normalization, and it inherits every property of the sample it is fit on.

Which sample that should be is the sharpest lesson of the practicum. A **coincident index** standardizes its inputs and calibrates its scaling over some reference sample, and that sample choice is a modeling decision. The P1 journal's second entry records what happens otherwise: standardizing over a sample that includes April 2020, when transformed claims sat roughly eight standard deviations from normal, inflates every standard deviation so badly that the 2008–09 recession shrinks to about one standard deviation and nearly vanishes from the fitted tracker. One episode redefined the yardstick for every other episode. The repair is a deliberate **calibration window**: estimate means, variances, factor weights, and the GDP scaling

on a sample of normal times (the build uses pre-2020 data), then apply those fixed parameters to the full history, so that extreme episodes are *measured* on a normal-times yardstick rather than allowed to compress it. Every standardization has a reference population; choosing it is modeling, not housekeeping.

9.4 Choice 4: Simplicity, Deliberately

The published WEI is principal components, not a full dynamic factor model, and its maintainers know both options. The choice is instructive. A transparent index that anyone can recompute has operational value, and on a nearly balanced weekly panel the static estimator gives up little. The full state-space treatment earns its complexity when the ragged edge, uncertainty bands, or news attribution are required, which is why the same authors' production systems, and this course's capstone, use the machinery of Section 6. The 2020 episode forced even the simple index to adapt: the published methodology was adjusted when the pandemic's outliers distorted the historical relationships: their production version of the standardization trap. Exercise 10 reads the published record of what changed.

Figure 15 shows the state-space payoff at the point where it matters operationally: the newest weeks of the panel, where some series have reported and others have not.

Figure 15. What the state-space form buys at the edge. Principal components requires complete rows, so its factor path ends at the last week in which every series has reported. The filtered estimate continues through the ragged edge, using whichever series have arrived, with a band that widens as the information thins.

The WEI has cousins, and knowing them frames the design space. The Chicago Fed's CFNAI extracts one factor from 85 monthly series by principal components; it is the direct descendant of the Stock and Watson (1989) coincident index, which was the first to state the factor-plus-filter program explicitly. The ADS index of Aruoba et al. (2009) runs an exact mixed-frequency dynamic factor model on daily through

quarterly data, and is closest in spirit to what this course builds. Baumeister et al. (2024) push the same design to weekly indices for each U.S. state. Between the transparent PCA tracker and the full state-space system lies most of applied nowcasting, and knowing when the simple end of the spectrum suffices is part of the craft.

10 From Comovement to a Production System

The session began with a Thursday panel that disagreed with itself. The resolution was not a better single indicator but a model of why indicators agree at all: a small number of pervasive factors, read through loadings, with everything series-specific modeled as such. It is worth an inventory of what that model now produces, because each item traces to a specific piece of machinery.

A factor estimate with a defensible weighting comes from the precision logic of Section 4, automated by principal components. An honest account of what is identified comes from Section 3, enforced by a stated normalization and a scaling regression into GDP units. Ragged-edge estimates with uncertainty bands, forecasts, a likelihood, and release-by-release news attribution come from the state-space form of Section 6 and whichever estimator in Table 9 the application justifies. And a public, reproducible weekly tracker comes from four design choices whose costs are now explicit.

What the session did not solve, it postponed deliberately, and one postponement now stands in the way. Every weekly series in this session entered through the 52-week difference, and Section 9 showed what that transformation really is: a seasonal adjustment in disguise, valid only for patterns that repeat on a fixed annual cycle, paid for in trailing-year units and anniversary echoes. High-frequency data deserve better: a model in which the seasonal pattern is estimated, can evolve, and can be separated from the signal at the data's own frequency. What such a model looks like for weekly and even hourly data, and what "seasonally adjusted" precisely means once you must produce it yourself, is Session 3.

Exercises

Exercises marked [**core**] prepare material used later in the course.

Exercise 1: The covariance a factor model implies

Problem. [core, pencil] A one-factor model has three standardized series with loadings $\lambda = (0.9,\ 0.6,\ 0.3)'$, factor variance 1, and mutually uncorrelated idiosyncratic components. Write the full 3×3 covariance matrix of the observed series. Compute each series' common share and each pairwise correlation. Which series is the most valuable witness to the factor, and what, if anything, does series 3 contribute?

Exercise 2: Two witnesses

Problem. [core, pencil] One factor with prior $f \sim \mathcal{N}(0, 1)$ is measured by two standardized series, each with loading $\lambda = 0.8$ and idiosyncratic variance $\sigma_e^2 = 0.36$, idiosyncratic components independent. The observed values are $x_1 = -1.5$ and $x_2 = -0.5$. Compute the posterior mean and variance of f, twice: once by Gaussian conditioning on the joint distribution and once by precision weighting. Compare with the posterior after observing only x_1, and explain in words why the second witness pulled the estimate in the direction it did.

Exercise 3: The witness floor

Problem. [core, pencil] N standardized series each measure a factor $f \sim \mathcal{N}(0, 1)$ with loading λ and idiosyncratic variance σ_e^2. Derive the posterior variance of f when the idiosyncratic components are (a) independent and (b) equicorrelated with correlation ρ_e. Show that in case (b) the posterior variance has a strictly positive limit as $N \to \infty$, and prove the statement: an unlimited supply of equicorrelated witnesses is worth exactly $1/\rho_e$ independent ones. Evaluate everything for $\lambda = 0.8$, $\sigma_e^2 = 0.36$, $\rho_e = 0.25$.

Exercise 4: Rotation, concretely

Problem. [core, pencil] (a) For a one-factor model, exhibit two loading–factor pairs besides (Λ, f_τ) that generate identical data, and identify what each violates in the normalization $\mathrm{Var}(f_\tau) = I$, $\Lambda'\Lambda$ diagonal, plus a sign convention. (b) For $k = 2$, show that any pair $(\Lambda M^{-1},\ M f_\tau)$ with M invertible fits identically, and determine which matrices M survive the full normalization.

Exercise 5: Replicate the P1 extraction

Problem. [core, computational, data] Reproduce the Practicum 1 factor extraction from scratch: pull the four weekly FRED series, transform each to a 52-week log difference with the claims series sign-flipped, standardize on the pre-2020 calibration window, and compute the first principal component of the calibration panel. Verify your weights against the instructor build. Then compute each series' common share with respect to your estimated factor and rank the four series by usefulness. Does the ranking match the weights? Should it?

Exercise 6: State-space fluency

Problem. [core, pencil] Write the full system matrices (T, R, Q, Z, H) for a dynamic factor model with $k = 2$ factors following a VAR(2), AR(1) idiosyncratic components, and $N = 5$ monthly series. Give every matrix's dimensions and the state dimension. Then, for the one-factor version of the model, explain where the five-weight quarterly aggregation row for GDP enters and exactly which additional states it requires.

Exercise 7: The two-step upgrade

Problem. [core, computational] Implement the two-step estimator on a simulated ragged panel: generate $n = 200$ periods of a one-factor panel with $N = 4$ series, loadings $(0.9, 0.8, 0.6, 0.3)$, factor AR(1) coefficient 0.9, and a ragged edge (the last three observations of series 1 and the last observation of series 2 missing). Step 1: principal components and regressions on the balanced rows. Step 2: Kalman filter and smoother over the full panel with parameters fixed. Add assertions. Where does the smoothed factor differ most from the PCA factor, and which estimate tracks the true factor better?

Exercise 8: Who gets credit for the news?

Problem. [core, pencil] A one-factor nowcast has prior $f \sim \mathcal{N}(0,\, 0.5)$ at the start of a week. Two releases arrive: claims, with loading 0.8 and idiosyncratic variance 0.36, surprises at $v_1 = -1.0$; credit, with loading 0.3 and idiosyncratic variance 0.91, surprises at $v_2 = +0.5$ (both measured against the prior). Compute the factor revision and its release-by-release attribution three ways: jointly, sequentially claims-first, and sequentially credit-first. Verify all three give the same final estimate, and explain why the middle numbers differ.

Exercise 9: Choosing the number of factors

Problem. [data] Implement the Bai and Ng (2002) criterion IC_{p2} and validate it on a simulated panel with a known number of factors before trusting it on real data. Then download the FRED-MD monthly panel (McCracken and Ng, 2016), apply the published transformation codes, standardize, and report the chosen number of factors, the variance shares of the leading components, and an interpretation of factor 2's largest loadings. [extra] Repeat the criterion on the pre-2020 subsample and compare.

Exercise 10: WEI archaeology

Problem. [data] The published Weekly Economic Index confronted its own version of the standardization trap in 2020. Using Lewis et al. (2022), Lewis et al. (2021), and the Dallas Fed's WEI documentation, identify what the pandemic did to the index's estimated relationships and what its maintainers changed in response. Relate each change to the calibration-window logic of the P1 build. Then propose, in one page, the v2 design for the four-series P1 tracker that best imports the lessons.

Further Reading

- Stock and Watson (2002) and Bai and Ng (2002) are the two pillars of large-panel factor econometrics. Read them for the theorems' *conditions*, weak dependence and pervasiveness, which is where the practice lives.
- Boivin and Ng (2006) documents when more series make factor estimates worse; it is the empirical companion to the correlated-witness floor of Section 4.
- Doz et al. (2011) develops the two-step estimator implemented in Exercise 7; Doz et al. (2012) supplies the quasi-maximum-likelihood theory behind it.
- Bańbura and Modugno (2014) is the EM reference for arbitrary missingness patterns and the standard citation behind production nowcasting systems.
- Lewis et al. (2022) is the practicum paper; read it with the P1 build open. Wegmüller and Glocker (2023) replicates and extends it, and is a model of what a careful replication looks like.
- Giannone et al. (2008) deserves a rereading now: its model is the factor-plus-aggregation state-space system of Section 6, and it will look familiar in a way it could not have after Session 1.

Session Glossary

The glossary collects the bold technical terms introduced in this session.

Every bold glossary term in the notes is linked to its definition below. After following a link, use the PDF viewer's Back command to return to the passage.

Approximate factor model A factor model that tolerates limited correlation among idiosyncratic components rather than requiring them to be fully uncorrelated. It is the realistic setting for macroeconomic panels, and the principal-components consistency results hold under it.

Base effect A movement in a year-over-year growth series caused by the comparison period a year earlier rather than by anything current. After an extreme episode, its anniversary produces an echo of the opposite sign that must not be read as news.

Blessing of dimensionality The property that adding series to a factor panel sharpens the factor estimate, because independent idiosyncratic noise averages away across the cross-section. It holds only while idiosyncratic components are weakly correlated and factors load broadly.

Calibration window The sample deliberately chosen for estimating standardization constants, factor weights, and scaling regressions, with the fitted parameters then applied to the full history. Fixing a normal-times window prevents one extreme episode from redefining the yardstick used to measure every other episode.

Coincident index A single series constructed to summarize the current state of aggregate activity, as distinct from a forecast of a specific future target. Its usefulness depends on stated units, a stated reference sample, and a stated release calendar.

Common share The fraction of a series' variance explained by the common factors rather than its idiosyncratic component. Indicators with high common shares carry the most information about aggregate activity, regardless of how interesting the series is on its own.

Comovement The tendency of many economic series to rise and fall together over the business cycle. Factor models treat that shared variation as the signal to be extracted rather than as collinearity to be fought.

Dynamic factor model A model in which many observed series load on a few common factors that evolve over time, with persistent idiosyncratic components modeled as additional states. In state-space form it inherits the Kalman filter's treatment of the ragged edge, uncertainty, likelihood, and news.

EM estimator An iterative estimator that alternates between a Kalman-smoother pass computing expected states given parameters and regressions on those smoothed moments re-estimating the

parameters. Each iteration weakly raises the likelihood, and arbitrary missing-data patterns are handled inside estimation itself.

Factor loading A coefficient connecting an observed series to a common factor. Its sign and magnitude describe how the series moves with the factor under the chosen normalization, and inspecting loadings is how a factor earns an economic interpretation.

Idiosyncratic component The part of an observed series not explained by the common factors: sampling error, sector-specific shocks, and definitional quirks. Its persistence and its correlation across series are modeling decisions with real consequences.

Normalization A restriction on factor scale, sign, or rotation that selects one representation from the many that fit the data identically. It makes reported factors and loadings interpretable without changing anything observable.

Principal components The least-squares estimator of a factor model, computed from the leading eigenvectors of the panel's sample covariance matrix. It is fast and transparent on balanced panels but supplies no dynamics, no missing-data handling, and no uncertainty bands.

Rotation An invertible transformation of the factors paired with the offsetting transformation of the loadings. Rotated factors fit the observed data exactly as well, which is why a normalization must be imposed before factors are interpreted.

Static factor model A representation in which each observed series equals its loadings times the current common factors plus an idiosyncratic component. It describes comovement within a period but does not specify how the factors evolve.

Two-step estimator A procedure that estimates loadings and dynamics by principal components and regression on the balanced part of a panel, then fixes those parameters and runs the Kalman filter and smoother over the full ragged panel. It trades a small efficiency loss for speed, stability, and auditability in production.

Weekly economic tracker A high-frequency index summarizing current aggregate activity from weekly indicators, usually one factor scaled into GDP units. Its transformation window, calibration sample, and release calendar must be stated before its readings can be interpreted as growth.

References

ARUOBA, S. B., DIEBOLD, F. X., & SCOTTI, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427.

BAI, J., & NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

BAŃBURA, M., GIANNONE, D., & REICHLIN, L. (2011). Nowcasting. In M. P. Clements & D. F. Hendry (eds.) *The Oxford handbook of economic forecasting* (pp. 193–224). Oxford University Press.

BAŃBURA, M., & MODUGNO, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, *29*(1), 133–160.

BAUMEISTER, C., LEIVA-LEÓN, D., & SIMS, E. (2024). Tracking weekly state-level economic conditions. *The Review of Economics and Statistics*, *106*(2), 483–504.

BOIVIN, J., & NG, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132*(1), 169–194.

BURNS, A. F., & MITCHELL, W. C. (1946). *Measuring business cycles*. National Bureau of Economic Research.

DOZ, C., GIANNONE, D., & REICHLIN, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, *164*(1), 188–205.

DOZ, C., GIANNONE, D., & REICHLIN, L. (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics*, *94*(4), 1014–1024.

DURBIN, J., & KOOPMAN, S. J. (2012). *Time series analysis by state space methods* (2nd ed.). Oxford University Press.

federal reserve bank of st. louis. (2026). *Real gross domestic product [A191RL1Q225SBEA].*

giannone, D., reichlin, L., & small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.

harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.

lewis, D. J., mertens, K., stock, J. H., & trivedi, M. (2021). High-frequency data and a weekly economic index during the pandemic. *AEA Papers and Proceedings*, *111*, 326–330.

lewis, D. J., mertens, K., stock, J. H., & trivedi, M. (2022). Measuring real activity using a weekly economic index. *Journal of Applied Econometrics*, *37*(4), 667–687.

mariano, R. S., & murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443.

mccracken, M. W., & ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

sargent, T. J., & sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In *New methods in business cycle research* (pp. 45–109). Federal Reserve Bank of Minneapolis.

stock, J. H., & watson, M. W. (1989). New indexes of coincident and leading economic indicators. In O. J. Blanchard & S. Fischer (eds.) *NBER macroeconomics annual 1989* (pp. 351–394). MIT Press.

stock, J. H., & watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

u.s. bureau of economic analysis. (2009a). *Gross domestic product, fourth quarter 2008 (advance).*

u.s. bureau of economic analysis. (2009b). *Gross domestic product, fourth quarter 2008 (final) and*

corporate profits.

wegmüller, P., & glocker, C. (2023). US weekly economic index: Replication and extension. *Journal of Applied Econometrics*, *38*(6), 977–985.

Course Glossary

Session 1 terms

Backcast An estimate of a reference period that has ended but whose official target has not yet been released. A backcast uses the same machinery as a forecast or nowcast; only the target's position on the release calendar differs.

Backtest or backtesting A historical evaluation that reruns a forecasting procedure at past origins using only the information that was genuinely available then. A real-time backtest therefore needs archived vintages as well as archived outcomes.

Benchmark A comparison forecast used to judge whether a more elaborate method adds value. Useful benchmarks are credible alternatives a forecaster could actually have used, not deliberately weak straw models.

Centering The location of a residual distribution relative to zero. Innovations centered away from zero reveal a systematic tendency to predict too high or too low.

Diffuse Describes prior information that is deliberately weak or uninformative for an initial state component. A diffuse treatment is appropriate when a nonstationary level has no finite unconditional distribution.

Diffuse initialization An initialization that represents little or no reliable prior information about one or more state components. Exact diffuse methods take the limiting case analytically; a large finite covariance is an approximation.

Distributional shape Features of a distribution beyond its center and scale, including skewness, tail thickness, and outliers. Innovation shape matters because a Gaussian model assigns specific probabilities to extreme forecast errors.

Dynamic factor model A model in which many observed series load on a small number of evolving common factors. It turns a large, correlated indicator panel into a low-dimensional state-space system.

Filtered estimate An estimate of a state using observations available only through that state's date within a vintage. It is the appropriate object for reconstructing a real-time information set.

Forecast An estimate of a target whose reference period has not yet begun or ended, depending on the convention being used. In these notes, the term contrasts with a nowcast of the period under way and a backcast of a completed but unpublished period.

Initialization The choice of the state distribution before the first filtering update. It supplies the initial

mean and covariance from which every later prediction and update is recursively computed.

Innovation The observed release minus its conditional expectation just before it arrives. It is the portion of the observation that is new relative to the model's information set.

Innovation variance The model-implied conditional variance of an innovation. It combines uncertainty about the predicted state with the measurement error in the arriving observation.

Kalman gain The coefficient or matrix that maps an innovation into a revision of the estimated state. It rises when the prior state is uncertain relative to the observation and falls when the observation is noisy.

Latent Not directly observed. A latent state is inferred from its dynamics and from observed variables that measure it imperfectly.

Local level model A state-space model in which an unobserved level follows a random walk and the observed series equals that level plus measurement error. It is the simplest model that separates genuine state movement from observation noise.

Log-likelihood The logarithm of the likelihood, viewed as a function of model parameters for the observed data. Taking logs converts a product of conditional densities into a sum and is numerically easier to optimize.

Loss function A rule that assigns a cost to a forecast error. Squared loss emphasizes large misses; absolute loss is less sensitive to a small number of extreme errors.

Measurement error The component of an observation that does not reflect the state the model is trying to measure. It can include sampling error, reporting noise, and series-specific movements.

Nonstationary Describes a process whose unconditional distribution is not stable over time. A random walk is nonstationary because its variance grows with the horizon and it has no finite unconditional covariance.

Nowcast An estimate of a target for the period currently under way, made before the official target is observed. Every nowcast must be paired with an information or release date.

Nowcast commentary A plain-language account of why a nowcast changed, normally identifying each release, its model surprise, and its contribution. It is the interpretive layer of a production nowcasting system.

Prediction step The part of the Kalman recursion that carries a filtered state distribution forward through the transition equation before the next measurement is used. It propagates the mean and adds uncertainty from new state shocks.

Production report The reproducible record accompanying an operational nowcast. It should identify the vintage, target, parameters or model version, uncertainty interval, release contributions, and any

re-estimation effect.

Ragged edge The staircase of missing values at the end of a real-time data panel caused by series being released on different schedules. The pattern reflects publication timing, not arbitrary deletion.

Rauch–Tung–Striebel smoother A backward recursion that revises earlier filtered state estimates using later observations in the same vintage. It is commonly abbreviated RTS smoother.

Reference period The economic week, month, or quarter described by an observation. It is distinct from the later date on which that observation is released.

Release date The calendar date and time at which an observation or revision becomes available to the forecaster. It determines whether the value may enter a particular real-time information set.

Scale The dispersion of a residual distribution. Standardized innovations with variance far from one indicate that the model's forecast uncertainty is too large or too small.

Serial dependence Predictable association between values of a series at different dates. Serial dependence in innovations means that some time-series structure remains unmodeled.

Shock A new, unpredictable disturbance that moves a state between periods. Its covariance determines how much state uncertainty accumulates during prediction.

Signal-to-noise ratio A comparison of variation attributed to the evolving state with variation attributed to measurement noise. In the local level model it is Q/H and governs how responsive the filter is to new observations.

Smoothed estimate An estimate of an earlier state that uses later observations from the full vintage. It is a hindsight estimate and therefore must not be presented as what was known in real time.

Stability Constancy of a model's predictive relationships and error behavior over time. A model can be well centered and scaled on average yet fail during a particular regime.

State The smallest collection of variables needed to summarize the model's current condition and generate the distribution of future observations. A state can contain latent factors, lags, trends, seasonal components, or time-varying coefficients.

Stationary Describes a process with a time-invariant unconditional distribution. For a stable linear state equation, stationarity permits initialization at the process's unconditional mean and covariance.

Surprise The realized value of a release minus what the model expected immediately beforehand. In a Kalman update, surprise is another name for the innovation.

Vintage A snapshot of all data values available as of a particular release date, including the missing cells and revisions known then. Two vintages may contain different values for the same reference period.

Terms introduced in Session 2

Approximate factor model A factor model that tolerates limited correlation among idiosyncratic components rather than requiring them to be fully uncorrelated. It is the realistic setting for macroeconomic panels, and the principal-components consistency results hold under it.

Base effect A movement in a year-over-year growth series caused by the comparison period a year earlier rather than by anything current. After an extreme episode, its anniversary produces an echo of the opposite sign that must not be read as news.

Blessing of dimensionality The property that adding series to a factor panel sharpens the factor estimate, because independent idiosyncratic noise averages away across the cross-section. It holds only while idiosyncratic components are weakly correlated and factors load broadly.

Calibration window The sample deliberately chosen for estimating standardization constants, factor weights, and scaling regressions, with the fitted parameters then applied to the full history. Fixing a normal-times window prevents one extreme episode from redefining the yardstick used to measure every other episode.

Coincident index A single series constructed to summarize the current state of aggregate activity, as distinct from a forecast of a specific future target. Its usefulness depends on stated units, a stated reference sample, and a stated release calendar.

Common share The fraction of a series' variance explained by the common factors rather than its idiosyncratic component. Indicators with high common shares carry the most information about aggregate activity, regardless of how interesting the series is on its own.

Comovement The tendency of many economic series to rise and fall together over the business cycle. Factor models treat that shared variation as the signal to be extracted rather than as collinearity to be fought.

EM estimator An iterative estimator that alternates between a Kalman-smoother pass computing expected states given parameters and regressions on those smoothed moments re-estimating the parameters. Each iteration weakly raises the likelihood, and arbitrary missing-data patterns are handled inside estimation itself.

Factor loading A coefficient connecting an observed series to a common factor. Its sign and magnitude describe how the series moves with the factor under the chosen normalization, and inspecting loadings is how a factor earns an economic interpretation.

Idiosyncratic component The part of an observed series not explained by the common factors: sampling error, sector-specific shocks, and definitional quirks. Its persistence and its correlation across series

are modeling decisions with real consequences.

Normalization A restriction on factor scale, sign, or rotation that selects one representation from the many that fit the data identically. It makes reported factors and loadings interpretable without changing anything observable.

Principal components The least-squares estimator of a factor model, computed from the leading eigenvectors of the panel's sample covariance matrix. It is fast and transparent on balanced panels but supplies no dynamics, no missing-data handling, and no uncertainty bands.

Rotation An invertible transformation of the factors paired with the offsetting transformation of the loadings. Rotated factors fit the observed data exactly as well, which is why a normalization must be imposed before factors are interpreted.

Static factor model A representation in which each observed series equals its loadings times the current common factors plus an idiosyncratic component. It describes comovement within a period but does not specify how the factors evolve.

Two-step estimator A procedure that estimates loadings and dynamics by principal components and regression on the balanced part of a panel, then fixes those parameters and runs the Kalman filter and smoother over the full ragged panel. It trades a small efficiency loss for speed, stability, and auditability in production.

Weekly economic tracker A high-frequency index summarizing current aggregate activity from weekly indicators, usually one factor scaled into GDP units. Its transformation window, calibration sample, and release calendar must be stated before its readings can be interpreted as growth.

Index