# The Nowcasting Problem, State Space Models & the Kalman Filter

Session 1 · Lecture notes

Tyler Sotomayor

2026-07-30

## 1 The nowcasting problem

### 1.1 Why the present is uncertain

The state of the economy *right now* is not observable. This sounds like a philosophical remark; it is in fact an institutional one. The headline measure of aggregate activity — real GDP — is compiled quarterly and published with a lag: the U.S. *advance* estimate arrives roughly four weeks after the quarter closes, which means that in the middle of any given quarter, the most recent official reading of activity refers to a period that ended up to four and a half months earlier. Policy, portfolios, and firms cannot wait that long, and in a crisis the lag is fatal to good decisions.

Worse, the number that eventually arrives is provisional. GDP is revised in the second and third estimates, at annual revisions, and again at comprehensive benchmark revisions years later. The revisions are not noise around zero; they are often large and systematically informative. The canonical example is 2008Q4, at the depth of the financial crisis:

Table 1: The same quarter, told three ways.

| Vintage | U.S. real GDP growth, 2008Q4 (SAAR) |
|---|---|
| Advance estimate (Jan 2009) | −3.8% |
| Third estimate (Mar 2009) | −6.3% |
| Comprehensive revision (2011+) | −8.5% |

The advance estimate understated the collapse by nearly five percentage points. Anyone evaluating a forecasting model in 2009 against the −3.8% print — or worse, backtesting years later against the −8.5% figure as if it had been knowable in real time — is doing something methodologically incoherent. This single table motivates a discipline we will maintain all course: **models are estimated and evaluated on data vintages**, the snapshots of the data as they actually stood on a given date, which for U.S. macro series are archived in the St. Louis Fed's ALFRED database.

Meanwhile, between GDP releases, an enormous volume of higher-frequency information arrives continuously: weekly initial claims for unemployment insurance, monthly industrial production and retail sales, daily asset prices and card-spending aggregates, hourly electricity load, continuous news text and search behavior. None of these *is* GDP, but all of them are correlated with the same underlying object — aggregate real activity. The problem of **nowcasting** (Giannone et al. 2008) is to exploit that correlation systematically.

### 1.2 The information structure

Formally, let $y_Q$ denote the low-frequency target (say, quarterly GDP growth for the current quarter $Q$) and let $\Omega_t$ denote the information set available at date $t$: every release of every series, as published, up to and including date $t$. The nowcast is the conditional expectation

$$\hat{y}_{Q|t} \;=\; \mathbb{E}\big[\, y_Q \mid \Omega_t \,\big], \tag{1}$$

and a full nowcasting system also reports the conditional distribution around it — a point estimate without a band is a rhetorical device, not a statistic.

Three structural features of $\Omega_t$ make Equation 1 hard to compute, and together they dictate the machinery of this course:

1. **Mixed frequency.** The target is quarterly; the indicators are monthly, weekly, daily. Any model that requires a balanced panel at a single frequency has already thrown away most of the information set.
2. **The ragged edge.** Series are released asynchronously with heterogeneous publication lags: on any given day, industrial production might be available through May, retail sales through June, claims through last Thursday, and financial data through this morning. The bottom-right corner of the data matrix is systematically, *predictably* incomplete — a staircase, not a rectangle. The jargon term is the *ragged edge* (Bańbura et al. 2011).
3. **Vintages and revisions.** $\Omega_t$ contains *first releases and interim revisions as of* $t$, not the final numbers. A model evaluated on revised data enjoys information its real-time competitor never had. The literature calls honest evaluation *pseudo real-time* (replaying history vintage by vintage) and we will do it constantly.

One further feature deserves emphasis because it is the source of most of the economic interest in nowcasting: information arrives as **discrete release events**. The nowcast does not drift continuously; it jumps when a release surprises, and it should jump by an amount that reflects both the size of the surprise and the informativeness of the series. A good nowcasting model is therefore also an *attribution* device: it tells you which release moved the estimate and by how much — the "news decomposition" we derive in Section 5.2.

### 1.3 What a solution must do

Collect the requirements. We need a modeling framework that (i) links many noisy indicators of different frequencies to a small number of latent objects, (ii) digests arbitrary patterns of missing data without ad hoc patching, (iii) updates beliefs release by release, with a well-defined measure of surprise, and (iv) delivers a full conditional distribution, not just a point. There is exactly one classical framework that does all four natively: the **linear Gaussian state space model**, estimated and filtered with the **Kalman filter**. That is what the rest of this session builds, carefully.

## 2 State space models

### 2.1 The general form

A linear Gaussian state space model consists of two equations. The **state equation** (or transition equation) describes the evolution of a latent state vector $\alpha_t \in \mathbb{R}^m$:

$$\alpha_{t+1} \;=\; T\,\alpha_t + R\,\eta_t, \qquad \eta_t \sim \mathcal{N}(0,\, Q), \tag{2}$$

and the **measurement equation** (or observation equation) links the observation vector $y_t \in \mathbb{R}^p$ to the state:

$$y_t \;=\; Z\,\alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,\, H), \tag{3}$$

with $\{\eta_t\}$ and $\{\varepsilon_t\}$ mutually independent white noise, independent of the initial condition $\alpha_1 \sim \mathcal{N}(a_1, P_1)$. The system matrices $T$ (dynamics), $R$ (noise loading), $Q$ (state noise covariance), $Z$ (measurement loadings), and $H$ (measurement noise covariance) may in general depend on $t$; we suppress the subscript except where time-variation is the point (as with missing data, where it is).

Two remarks on interpretation. First, the state $\alpha_t$ is whatever the modeler needs it to be — "the underlying level of activity," a vector of factors and their lags, a time-varying regression coefficient. The framework is agnostic; the economics enters through the choice of matrices. Second, the measurement error $\varepsilon_t$ absorbs everything about an indicator that is *not* the common object: sampling error in a survey, idiosyncratic sectoral shocks, definitional mismatch. The decomposition of each observable into "signal via $Z\alpha_t$" plus "idiosyncratic noise" is the central modeling act in nowcasting.

## 2.2 A zoo of special cases

The reason this form is worth mastering is that nearly every model in this course is a state space model with particular matrices. Working through the examples below is the fastest way to internalize the notation — and several are exercises at the end of the chapter.

**AR(1).** Take $m = p = 1$, $Z = 1$, $H = 0$, $T = \phi$, $R = 1$, $Q = \sigma^2$. Then $y_t = \alpha_t$ follows $y_{t+1} = \phi y_t + \eta_t$. Measurement error is zero: the state is observed. The filter will simply reproduce the data — a useful degenerate case to check your code against.

**AR(p) in companion form.** Stack $\alpha_t = (y_t, y_{t-1}, \dots, y_{t-p+1})'$ and let

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

The same trick — stacking lags into the state — is how *any* finite-order dynamics enters the framework, and you will use it constantly. A VAR(1) in $k$ variables is the case where the blocks above are $k \times k$ matrices.

**Local level.** The simplest genuinely latent model, and the workhorse of Section 6:

$$\alpha_{t+1} = \alpha_t + \eta_t, \qquad y_t = \alpha_t + \varepsilon_t.$$

A random-walk level observed in noise: $T = R = Z = 1$, $Q = \sigma_\eta^2$, $H = \sigma_\varepsilon^2$.

**Local linear trend.** Add a slope $\beta_t$ that itself drifts:

$$\begin{aligned} \mu_{t+1} &= \mu_t + \beta_t + \eta_{1t}, \\ \beta_{t+1} &= \beta_t + \eta_{2t}, \end{aligned} \qquad y_t = \mu_t + \varepsilon_t,$$

i.e. $\alpha_t = (\mu_t, \beta_t)'$, $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $Z = (1, 0)$. Setting various variances to zero recovers deterministic trends, smooth trends, and — with a seasonal block appended — the structural time series models of Harvey (1989) that underlie seasonal adjustment (Session 3).

**Dynamic factor model (preview of Session 2).** Let $f_t$ be a scalar (or small vector) common factor with AR dynamics, and let each of $p$ indicators load on it:

$$y_{it} = \lambda_i f_t + \varepsilon_{it}, \qquad f_{t+1} = \phi f_t + \eta_t.$$

Here $Z = (\lambda_1, \dots, \lambda_p)'$ and $H$ is diagonal: the indicators are conditionally independent given the factor. This is the nowcasting model of Giannone et al. (2008), and with $p$ large it is the industry standard. Everything you learn today about filtering transfers wholesale.

**Time-varying-parameter regression.** In $y_t = x_t'\beta_t + \varepsilon_t$ with $\beta_{t+1} = \beta_t + \eta_t$, set $\alpha_t = \beta_t$, $Z_t = x_t'$ (time-varying!), $T = I$. The filter then delivers recursive least squares with drifting coefficients — a tool we will meet again with Markov switching in Session 6.

## 2.3 Mixed frequency: the Mariano–Murasawa construction

The mixed-frequency problem has an elegant state space resolution: **model everything at the highest frequency and treat the low-frequency series as a partially observed aggregate**. The classic construction is due to Mariano and Murasawa (2003).

Work at monthly frequency and let $g_t$ denote unobserved month-on-month GDP growth (in logs). Quarterly GDP is the sum of its three monthly levels; taking three-month log differences of the quarterly level and expanding in monthly log differences yields the (approximate but extremely accurate) aggregation identity

$$y_t^Q \;=\; \tfrac{1}{3}g_t + \tfrac{2}{3}g_{t-1} + g_{t-2} + \tfrac{2}{3}g_{t-3} + \tfrac{1}{3}g_{t-4}, \tag{4}$$

where $y_t^Q$ is quarterly growth assigned to the third month of each quarter. The tent-shaped weights $(\frac{1}{3}, \frac{2}{3}, 1, \frac{2}{3}, \frac{1}{3})$ arise because a quarterly level compares an average of three months to the average of the three months before — months in the middle of that window are counted more often.

To implement Equation 4, stack $g_t$ and four of its lags in the state, give the quarterly series a measurement row whose $Z$ entries are the tent weights, and declare that row **observed only every third month and missing otherwise**. That last move — treating a frequency mismatch as a missing-data pattern — is the conceptual pivot of the whole approach, and it costs nothing, as Section 2.4 shows.

## 2.4 Missing data as measurement-equation surgery

Suppose at time $t$ only a subset of the $p$ series is observed. Let $W_t$ be the selection matrix that keeps the observed rows (e.g. if series 1 and 3 of three are observed, $W_t = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$). Then the measurement equation *for that period* is simply

$$y_t^{obs} = (W_t Z)\,\alpha_t + W_t\varepsilon_t, \qquad W_t\varepsilon_t \sim \mathcal{N}(0,\, W_t H W_t'),$$

a state space model with time-varying $Z_t = W_t Z$ and $H_t = W_t H W_t'$ of reduced row dimension. Nothing else changes. If *nothing* is observed at $t$, the measurement equation is empty and the period contributes only a prediction step. The ragged edge, quarterly-in-monthly aggregation, publication delays, even outlier removal (delete the row) — all are instances of this one operation. This is why state space methods dominate production nowcasting: the hard institutional facts of the data become bookkeeping, not obstacles.

# 3 The Kalman filter

## 3.1 What we are computing

Fix the notation for conditional moments given data through time $s$, $Y_s = \{y_1, \dots, y_s\}$:

$$a_{t|s} = \mathbb{E}[\alpha_t \mid Y_s], \qquad P_{t|s} = \mathrm{Var}(\alpha_t \mid Y_s).$$

Because the model is linear with Gaussian shocks and a Gaussian initial state, *every* conditional distribution in sight is Gaussian; in particular

$$\alpha_t \mid Y_s \;\sim\; \mathcal{N}\big(a_{t|s},\, P_{t|s}\big)\,.$$

So the pair $(a_{t|s}, P_{t|s})$ is not a summary of our beliefs — it *is* our beliefs, exactly. The Kalman filter (Kalman 1960) is the recursion that computes $(a_{t|t}, P_{t|t})$ and $(a_{t+1|t}, P_{t+1|t})$ for $t = 1, 2, \dots$ in a single forward pass, at cost linear in the sample size.

## 3.2 The one lemma you need

Everything rests on the conditioning formula for jointly Gaussian vectors.

> **!** Lemma (Gaussian conditioning / 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 $\Sigma_{yy}$ invertible, then
>
> $$x \mid y \;\sim\; \mathcal{N}\Big(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y),\;\; \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\Big).$$
>
> The mean is the *linear projection* of $x$ on $y$; the variance is the projection residual variance, and — a fact worth staring at — it does not depend on the realized value of $y$.

Under Gaussianity this is exact; without Gaussianity the same formulas still deliver the *best linear* predictor, which is why the Kalman filter retains a minimum-MSE-among-linear-rules interpretation even when normality fails.

## 3.3 Derivation of the recursions

Suppose at time $t$ we hold $\alpha_t \mid Y_t \sim \mathcal{N}(a_{t|t}, P_{t|t})$. Two steps advance the clock.

**Prediction.** Apply the state equation Equation 2 and take conditional moments:

$$a_{t+1|t} = T\, a_{t|t}, \qquad P_{t+1|t} = T\, P_{t|t}\, T' + R\, Q\, R'. \tag{5}$$

The mean moves with the dynamics; the variance is pushed through $T$ and then *inflated* by the fresh state noise. While no data arrives, repeated application of Equation 5 is all there is: uncertainty compounds period by period. (You have seen this: it is the cone fanning out on the course homepage.)

**Update.** When $y_{t+1}$ arrives, write the joint distribution of state and observation given $Y_t$, both of which are linear in Gaussian objects:

$$\begin{pmatrix} \alpha_{t+1} \\ y_{t+1} \end{pmatrix} \Big| Y_t \;\sim\; \mathcal{N}\left( \begin{pmatrix} a_{t+1|t} \\ Z a_{t+1|t} \end{pmatrix}, \begin{pmatrix} P_{t+1|t} & P_{t+1|t} Z' \\ Z P_{t+1|t} & Z P_{t+1|t} Z' + H \end{pmatrix} \right).$$

Now apply the lemma with $x = \alpha_{t+1}$, $y = y_{t+1}$. Define the **innovation** and its variance,

$$v_{t+1} = y_{t+1} - Z\, a_{t+1|t}, \qquad F_{t+1} = Z\, P_{t+1|t}\, Z' + H, \tag{6}$$

and the **Kalman gain** $K_{t+1} = P_{t+1|t}\, Z'\, F_{t+1}^{-1}$. The lemma delivers, exactly,

$$a_{t+1|t+1} = a_{t+1|t} + K_{t+1}\, v_{t+1}, \qquad P_{t+1|t+1} = (I - K_{t+1} Z)\, P_{t+1|t}. \tag{7}$$

That is the whole filter. It is worth pausing on how little machinery was needed: one lemma, applied once per period.

> The gain is the whole story
>
> $K_{t+1}$ converts a unit of surprise in the data into units of revision of the state estimate. Inspect its anatomy: it is *increasing* in the prior state uncertainty $P_{t+1|t}$ and *decreasing* in the measurement noise $H$. When we know little and the data is clean, the filter chases the data; when we know much or the data is noisy, it barely moves. A nowcasting model is, at heart, a machine for computing **how much attention each data release deserves** — and the gain is that attention weight, derived rather than assumed. When practitioners say "the market ignored the claims number," a Kalman filter would say "$K$ was small in that direction this week."

The innovation $v_{t+1}$ has three properties you should verify (Exercise 3): it has mean zero, variance $F_{t+1}$, and is *independent of* $Y_t$ — the innovations are the "new information" in a precise, orthogonal sense. The sequence $\{v_t\}$ is therefore serially uncorrelated if the model is correctly specified, which is the basis of every diagnostic check we will run on fitted nowcasting models.

## 3.4 The algorithm, assembled

> Algorithm — Kalman filter
>
> Given $a_{1|0} = a_1$, $P_{1|0} = P_1$, for $t = 1, \dots, n$:
>
> 1. If $y_t$ is (partially) observed, form $Z_t = W_t Z$, $H_t = W_t H W_t'$ and compute $v_t = y_t^{obs} - Z_t a_{t|t-1}$, $F_t = Z_t P_{t|t-1} Z_t' + H_t$, $K_t = P_{t|t-1} Z_t' F_t^{-1}$, and update $a_{t|t} = a_{t|t-1} + K_t v_t$, $P_{t|t} = (I - K_t Z_t) P_{t|t-1}$.
> 2. If $y_t$ is entirely missing, set $a_{t|t} = a_{t|t-1}$, $P_{t|t} = P_{t|t-1}$.
> 3. Predict: $a_{t+1|t} = T a_{t|t}$, $P_{t+1|t} = T P_{t|t} T' + R Q R'$.
>
> Store $\{a_{t|t-1}, P_{t|t-1}, v_t, F_t, K_t\}$; they are reused by the smoother and the likelihood.

## 3.5 Initialization

The recursion needs $a_1, P_1$. Three cases arise in practice.

**Known initial conditions** occur in engineered systems (you know where the rocket started) and essentially never in economics.

**Stationary states.** If the state process is stationary (all eigenvalues of $T$ strictly inside the unit circle), initialize at the unconditional distribution: $a_1 = 0$ (after demeaning) and $P_1$ solving the discrete Lyapunov equation $P_1 = T P_1 T' + R Q R'$, i.e. $\mathrm{vec}(P_1) = (I - T \otimes T)^{-1} \mathrm{vec}(R Q R')$. This is the right choice for stationary factors and AR blocks.

**Nonstationary states** (random-walk levels, trends — including our local level model) have no unconditional distribution. The practical device is the **diffuse prior**: $P_1 = \kappa I$ with $\kappa$ huge, expressing near-total initial ignorance; the first few observations then essentially *set* the level, after which the filter behaves as if it had always known it. Rigorous treatments take $\kappa \to \infty$ analytically ("exact diffuse initialization," Durbin and Koopman (2012), ch. 5); numerically, $\kappa$ around $10^6$ times the data variance works and is what most libraries do by default. The transient matters: diffuse periods contribute differently to the likelihood, and estimates during the first few periods should not be over-interpreted.

### 3.6 Numerical practice

Textbook formulas are not always what you should compute. Three habits separate working filters from fragile ones:

- **Symmetrize.** Round-off drifts $P$ away from symmetry; enforce $P \leftarrow (P + P')/2$ each step.
- **Use the Joseph form** of the covariance update,

$$P_{t|t} = (I - K_t Z_t)\, P_{t|t-1}\, (I - K_t Z_t)' + K_t H_t K_t',$$

  which is algebraically identical to Equation 7 but preserves positive semi-definiteness under any (even wrong) gain — at the cost of a few more multiplications. Production systems use it, or go further to square-root filters that propagate a Cholesky factor of $P$.
- **Never invert $F_t$ explicitly.** Solve the linear systems; better, if $H$ is diagonal, process the elements of $y_t$ *one at a time* (univariate filtering, Durbin and Koopman (2012) ch. 6) — a sequence of scalar updates that is both faster and more stable, and which meshes beautifully with data arriving release by release.

## 4 Smoothing

The filter answers "what do we know *now*?" History-writing asks a different question: "given the *whole* sample $Y_n$, what was the state at time $t$?" The answer, $a_{t|n} = \mathbb{E}[\alpha_t \mid Y_n]$ with variance $P_{t|n}$, is computed by a *backward* pass through the stored filter output — the Rauch–Tung–Striebel smoother:

$$J_t = P_{t|t}\, T'\, P_{t+1|t}^{-1}, \qquad a_{t|n} = a_{t|t} + J_t \left(a_{t+1|n} - a_{t+1|t}\right), \qquad P_{t|n} = P_{t|t} + J_t \left(P_{t+1|n} - P_{t+1|t}\right) J_t',$$

initialized at $t = n$ where smoothed and filtered coincide. Each smoothed estimate blends the filtered estimate with the (discounted) surprise contained in the future.

The division of labor in nowcasting is clean: **filtered** estimates are the real-time product — they respect the information set and are what you publish today; **smoothed** estimates are for historical description, parameter estimation (via EM, Section 5), and understanding what the model now thinks was happening in, say, March 2020. Comparing the two paths for the same episode is one of the most instructive plots you can make, and you will make it in Practicum 1. Smoothed uncertainty satisfies $P_{t|n} \preceq P_{t|t}$ — hindsight is genuinely sharper — with equality only at $t = n$.

## 5 The likelihood, estimation, and the news

### 5.1 The likelihood is a by-product

The system matrices contain unknown parameters $\theta$ (variances, AR coefficients, factor loadings). By the **prediction error decomposition**, the log-likelihood of the sample factors through the innovations, all of which the filter has already computed:

$$\log L(\theta) \;=\; -\tfrac{1}{2}\sum_{t=1}^{n}\left(p_t \log 2\pi + \log|F_t| + v_t' F_t^{-1} v_t\right), \tag{8}$$

where $p_t$ is the number of series actually observed at $t$ (missing data simply contributes smaller blocks). One forward pass yields the exact likelihood; a numerical optimizer wrapped around it yields maximum likelihood. Two practical notes. First, parameterize so constraints hold automatically (optimize $\log\sigma$, not $\sigma$). Second, for high-dimensional models the **EM algorithm** — which alternates the smoother (E-step) with regressions on smoothed moments (M-step) — is far more stable than direct optimization and is the standard approach for nowcasting-scale factor models (Bańbura and Modugno 2014). We implement both, small-scale, in Practicum 1.

Identification deserves one honest sentence: not every parameterization is identified (a factor's scale can be traded against its loadings, for instance), and every nowcasting paper quietly fixes a normalization. When we build factor models in Session 2 you will see exactly which.

### 5.2 The news decomposition

Between two data vintages — yesterday's information set $\Omega_{old}$ and today's $\Omega_{new}$, which adds one or more releases — the nowcast of any target $y^*$ changes by an amount the framework attributes *exactly*. Because updates are linear projections on innovations, one can show (Bańbura et al. 2011):

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

where $v$ is the vector of innovations in the newly released figures — the **news**, i.e. release minus what the model expected the release to be. The right-hand side splits release by release into additive contributions: *"initial claims came in 12K below expectation and moved the GDP nowcast by +0.08pp; the retail sales surprise added +0.05pp."* This is precisely the attribution language of the Atlanta Fed's GDPNow commentary and of every professional nowcast tracker, and it falls out of the Kalman filter with no additional modeling. Note what it implies: a release that arrives *exactly at expectation* moves the nowcast not at all, no matter how big the number — only surprise moves beliefs.

## 6 The local level model, in full

We close by working the simplest latent model to the bone, because every qualitative property of grand nowcasting systems is already visible in it — and because you can manipulate it live in the playground.

### 6.1 The filter reduces to adaptive smoothing

With $\alpha_{t+1} = \alpha_t + \eta_t$, $y_t = \alpha_t + \varepsilon_t$, all matrices scalar, write $P_t \equiv P_{t|t-1}$. The recursions Equation 5, Equation 6, Equation 7 collapse to

$$a_{t|t} = a_{t|t-1} + K_t\left(y_t - a_{t|t-1}\right), \qquad K_t = \frac{P_t}{P_t + \sigma_\varepsilon^2},$$

with $P_{t+1} = P_t(1 - K_t) + \sigma_\eta^2$. Rearranged,

$$a_{t|t} \;=\; (1 - K_t)\, a_{t-1|t-1} + K_t\, y_t :$$

an **exponentially weighted moving average whose weight is chosen optimally each period**. Simple exponential smoothing — the oldest trick in applied forecasting — is thus the exact optimal

filter for a random walk observed in noise, once its smoothing constant is set to the steady-state gain derived next. Ad hoc practice and optimal theory meet.

### 6.2 The steady state, derived

Define the signal-to-noise ratio $q = \sigma_\eta^2/\sigma_\varepsilon^2$ and the scaled predicted variance $\pi_t = P_t/\sigma_\varepsilon^2$. The variance recursion becomes

$$\pi_{t+1} = \frac{\pi_t}{\pi_t + 1} + q.$$

This scalar Riccati recursion is monotone and contracts to a unique fixed point $\bar{\pi}$ solving $\bar{\pi}^2 - q\,\bar{\pi} - q = 0$:

$$\bar{\pi} = \frac{q + \sqrt{q^2 + 4q}}{2}, \qquad \bar{K} = \frac{\bar{\pi}}{\bar{\pi} + 1}. \tag{9}$$

Comparative statics tell the whole economics. As $q \to \infty$ (turbulent state, pristine data), $\bar{K} \to 1$: yesterday's estimate is nearly worthless, chase the data. As $q \to 0$ (placid state, noisy data), $\bar{K} \to 0$ like $\sqrt{q}$: average hard, trust the model. The convergence of $K_t$ to $\bar{K}$ from its diffuse start (where $K_1 \approx 1$ — the first observation essentially *is* the estimate) is plotted live in the playground's second panel.

### 6.3 The ragged edge in one frame

Switch on the playground's *data outage*. During the gap the update step is skipped, so $P$ grows by $\sigma_\eta^2$ each period — the band fans out linearly in variance — and the point estimate freezes at its last value: the model coasts. At the first post-outage observation the accumulated $P$ makes the gain *temporarily much larger than* $\bar{K}$: the filter knows it has been flying blind and lunges at the new data point, after which the gain relaxes back to steady state. That single frame — coast, fan out, lunge, relax — is the entire ragged-edge logic of a production nowcasting system, and you now own every equation behind it.

## 7 Historical note

The filter is named for Rudolf E. Kálmán, whose 1960 paper (Kalman 1960) recast estimation as recursive state updating — a formulation the aerospace community adopted almost immediately (the Apollo guidance computer flew a descendant of it to the Moon). Economics arrived at the same mathematics through a different door: the structural time series tradition of Harvey (1989), the real-time coincident indicator of Aruoba et al. (2009), and finally the nowcasting synthesis of Giannone et al. (2008), which recognized that the filter's native handling of missing, mixed-frequency data was exactly what the ragged edge required. It is one of the cleanest cases in economics of an imported tool fitting the institutional facts better than anything home-grown.

## Exercises

**1. Warm-up with matrices.** Write the local linear trend model with a stochastic seasonal component (period 4) in state space form: give $\alpha_t$, $T$, $R$, $Z$ explicitly and state the dimension of each.

**2. Companion-form fluency.** Cast the ARMA(1,1) process $y_t = \phi y_{t-1} + \theta\epsilon_{t-1} + \epsilon_t$ in state space form. (Hint: one clean solution uses a 2-dimensional state; measurement error is zero.)

**3. Innovations are news.** Using the lemma, prove that $\mathbb{E}[v_{t+1} \mid Y_t] = 0$ and that $\mathrm{Cov}(v_{t+1}, v_s) = 0$ for $s \le t$. Conclude that the innovation sequence of a correctly specified model is white noise, and propose a diagnostic test based on this fact.

**4. Joseph form.** Show algebraically that the Joseph-form covariance update equals $(I - KZ)P_{t+1|t}$ when $K$ is the optimal gain, and that it remains positive semi-definite for *any* $K$. Why is the second property valuable on a computer?

**5. Steady state.** Derive Equation 9 from the variance recursion. Then show the recursion is a contraction toward $\bar{\pi}$: verify that $\pi_{t+1} - \bar{\pi}$ has the same sign as $\pi_t - \bar{\pi}$ and smaller magnitude. Finally compute $\bar{K}$ for $q \in \{0.01, 0.1, 1, 10\}$ and check your numbers against the playground.

**6. Missing-data filter.** Implement the local level filter in Python with an arbitrary missingness pattern (30 lines suffice). Reproduce the playground's outage experiment: verify the variance grows linearly during the gap and that the first post-gap gain exceeds $\bar{K}$.

**7. The 2008Q4 exercise (bridge to the course theme).** Download from ALFRED the vintages of 2008Q4 real GDP growth available on 2009-01-30, 2009-03-26, and today. Compute the revision sequence. If a nowcasting model had predicted $-4.0\%$ in January 2009, discuss — precisely — against which number its accuracy should be judged, and why.

# Further reading

- Durbin and Koopman (2012), chs. 2, 4–7 — the canonical modern reference; our notation follows it.
- Harvey (1989) — structural time series models; the book that brought this toolkit into economics.
- Giannone et al. (2008) — the founding nowcasting paper; read §2 with these notes beside it.
- Bańbura et al. (2011) — the ragged edge and news decomposition in production; the cleanest survey.
- Bańbura and Modugno (2014) — EM estimation with arbitrary missing data, the method behind serious nowcasting systems.
- Lewis et al. (2022) — **read before Session 2**; we will replicate it.


Aruoba, S. Borağan, Francis X. Diebold, and Chiara Scotti. 2009. "Real-Time Measurement of Business Conditions." *Journal of Business & Economic Statistics* 27 (4): 417–27.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin. 2011. "Nowcasting." In *The Oxford Handbook of Economic Forecasting*, edited by Michael P. Clements and David F. Hendry. Oxford University Press.

Bańbura, Marta, and Michele Modugno. 2014. "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data." *Journal of Applied Econometrics* 29 (1): 133–60.

Durbin, James, and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods*. 2nd ed. Oxford University Press.

Giannone, Domenico, Lucrezia Reichlin, and David Small. 2008. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics* 55 (4): 665–76.

Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Kalman, Rudolf E. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering* 82 (1): 35–45.

Lewis, Daniel J., Karel Mertens, James H. Stock, and Mihir Trivedi. 2022. "Measuring Real Activity Using a Weekly Economic Index." *Journal of Applied Econometrics* 37 (4): 667–87.

Mariano, Roberto S., and Yasutomo Murasawa. 2003. "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series." *Journal of Applied Econometrics* 18 (4): 427–43.