

Time Series Analysis

Chapter 10: State space models, Kalman filter and Markov switching

Daniel Traian PELE

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

Acronyms used in this material

- **ACF**: Autocorrelation Function
- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **AR**: AutoRegressive (model)
- **ARCH**: AutoRegressive Conditional Heteroskedasticity
- **ARIMA**: AutoRegressive Integrated Moving Average (model)
- **ARMA**: AutoRegressive Moving Average (model)
- **BET**: Bucharest Exchange Trading (index)
- **BIC**: Bayesian Information Criterion
- **EM**: Expectation–Maximisation (algorithm)
- **EODHD**: EOD Historical Data (market-data provider)
- **ETS**: Error, Trend, Seasonality (exponential smoothing state space models)
- **EU**: European Union
- **FRED**: Federal Reserve Economic Data
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GDP**: Gross Domestic Product
- **HP**: Hodrick–Prescott (filter)
- **MA**: Moving Average
- **ML**: Machine Learning
- **MS-AR**: Markov-Switching AutoRegressive (model)
- **NASA**: National Aeronautics and Space Administration
- **NBER**: National Bureau of Economic Research
- **OLS**: Ordinary Least Squares
- **RTS**: Rauch–Tung–Striebel (smoother)
- **SD**: Standard Deviation
- **SES**: Simple Exponential Smoothing
- **SWARCH**: Switching ARCH
- **TVP**: Time-Varying Parameter (regression)
- **UC**: Unobserved Components (model)
- **US**: United States
- **VAR**: Vector AutoRegression

Today's question and route

- **Question**: how do we measure what we cannot observe directly: the true level of a noisy series, the output gap, a changing beta, a recession?
  - the answer: write the unobserved quantity as a **state** and let the data update our estimate of it, period by period
- **Route** of the chapter
  - the state space form: measurement and transition equations; local level, local linear trend, ARMA, time-varying regression
  - the Kalman filter by hand; smoothing; likelihood and estimation; missing values
  - trend and cycle: Romanian GDP, the HP filter and its critique; dynamic factors and nowcasting
  - Markov switching: recessions, Romanian growth regimes, volatility regimes
- We build on Chapter 0 (exponential smoothing), Chapter 2 (ARMA), Chapter 5 (GARCH) and Chapter 8 (long memory)

Learning outcomes

- Write a model in state space form: local level, local linear trend, AR(2), a regression with a time-varying coefficient
- Run the Kalman filter by hand and explain the Kalman gain
- Show that simple exponential smoothing is the steady state of the local level filter
- Estimate a state space model by maximum likelihood, smooth it, and handle missing values and forecasts
- Decompose GDP into trend and cycle and judge the HP filter against the critique of Hamilton (2018)
- Estimate a Markov-switching model and read its regime probabilities, durations and pitfalls

Reading and tools

- Textbook: Huang and Petukhina (2022), the chapter on state space models and the Kalman filter
  - state space: Durbin and Koopman (2012) (the Nile example of this chapter comes from their book); Harvey (1989); Shumway and Stoffer (2017), Ch. 6
  - regime switching: Hamilton (1989); Hamilton (1994), Ch. 13 (Kalman filter) and Ch. 22 (regime changes); Kim and Nelson (1999)
- Python Quantlets of this chapter: Quantlets/Ch_10
  - a Kalman filter written out in `numpy`; `statsmodels`: `UnobservedComponents`, `SARIMAX`, `DynamicFactor`, `MarkovRegression`, `MarkovAutoregression`
- Lecture notebook: open in Google Colab
- Video course: Kalman Filter

Rudolf Kálmán and the Apollo navigation



Rudolf E. Kálmán (1930–2016)

Photo: ETH-Bibliothek Zürich, Bildarchiv (2007); CC BY-SA 4.0; Wikimedia Commons



Apollo 8 (1968): Jim Lovell at the guidance and navigation station

Photo: NASA (1968); public domain; Wikimedia Commons

- Kalman (1960): a recursive estimate of a hidden state from noisy measurements
  - at each step: predict the state, then correct the prediction with the new measurement
  - the correction weight, the **Kalman gain**, balances the two sources of uncertainty
- The filter was adapted at NASA Ames for the navigation of the Apollo spacecraft
  - state: position and velocity; measurements: star sightings and radar
- In economics the state is a trend, a cycle, a coefficient or a regime; the measurements are the published data

Hidden states in economics and finance

Observed series	Hidden state	Model in this chapter
Nile flow, monthly inflation	the underlying level	local level
real GDP	trend and output gap	unobserved components
stock index returns	a time-varying beta	time-varying regression
several activity indicators	one common factor	dynamic factor model
GDP growth, returns	recession or expansion; calm or turbulent	Markov switching

- The first four have a **continuous** state and are filtered by the Kalman filter
- The last has a **discrete** state (a regime) and is filtered by the Hamilton filter
- Both filters do the same thing: predict the state, then update the prediction with y_t

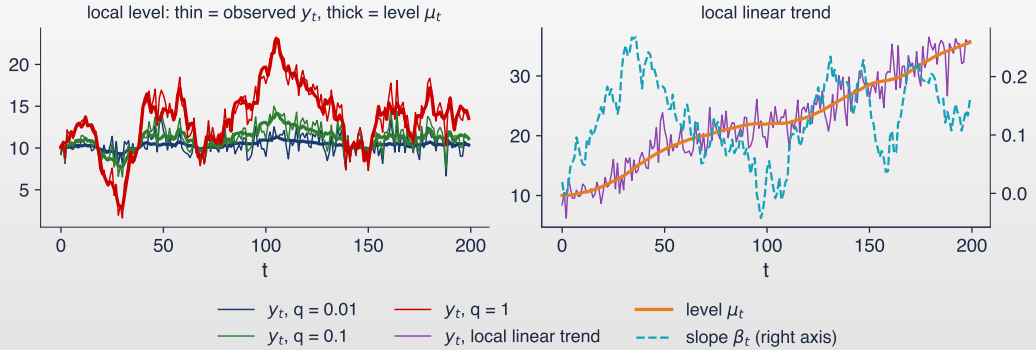
The state space form

- **Measurement equation**: $y_t = Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H)$
  - y_t: the observation (here a scalar); α_t: the **state vector** ($m \times 1$), not observed; Z_t: a $1 \times m$ row of known numbers
- **Transition equation**: $\alpha_{t+1} = T \alpha_t + R \eta_t, \quad \eta_t \sim N(0, Q)$
  - the state evolves as a first-order vector autoregression (VAR(1), Chapter 6); T is $m \times m$
  - R: maps the shocks η_t into the state; H, Q: the variances of the measurement and of the state shocks
  - initial state: $\alpha_1 \sim N(a_1, P_1)$, a_1, P_1: its mean and variance; ε_t, η_s and α_1 mutually independent
- The **system matrices** Z_t, H, T, R, Q contain the parameters θ; they are estimated by maximum likelihood
  - the same form covers ARMA, exponential smoothing, trend–cycle models, regressions with moving coefficients and factor models Durbin and Koopman (2012)

The local level model

- **Local level** (random walk plus noise): $y_t = \mu_t + \varepsilon_t, \quad \mu_{t+1} = \mu_t + \eta_t$
  - state $\alpha_t = \mu_t$, the level; $Z = T = R = 1$, $H = \sigma_\varepsilon^2$, $Q = \sigma_\eta^2$
  - ε_t: measurement noise that disappears next period; η_t: a permanent shift of the level
- **Signal-to-noise ratio** $q = \sigma_\eta^2/\sigma_\varepsilon^2$: how much of the movement is permanent
  - $q = 0$: a constant level (white noise around a mean); $q \to \infty$: a pure random walk
- Reduced form: $\Delta y_t = \eta_{t-1} + \varepsilon_t - \varepsilon_{t-1}$ is an MA(1) with $\rho(1) = -1/(q+2)$
  - so the local level model is an ARIMA(0,1,1) (Chapter 3) with a negative MA coefficient
  - $\gamma(0) = \sigma_\eta^2 + 2\sigma_\varepsilon^2$ and $\gamma(1) = -\sigma_\varepsilon^2$ give $\rho(1)$

Local level and local linear trend paths



- Left: the same shocks $\varepsilon_t, \eta_t \sim N(0,1)$ with $\sigma_\varepsilon^2 = 1$ and $q = 0.01;\ 0.1;\ 1$ (thin: y_t, thick: μ_t). Right: a local linear trend, whose slope β_t is itself a random walk

TSA_ch10_state_space_examples

Interpreting the simulated paths

- Small q: the level hardly moves and y_t looks like noise around a mean; large q: the level wanders and y_t follows it
  - first-difference autocorrelation: -0.51 for $q = 0.01$ (theory $-1/2.01$) and -0.28 for $q = 1$ (theory $-1/3$)
- The local linear trend has a slope that changes slowly: periods of fast and slow growth
  - this is the trend model we will use for GDP
- The estimation problem: from y_t alone, how much is level and how much is noise? The Kalman filter answers it for given variances; the likelihood chooses the variances

The local linear trend

- **Local linear trend**: $y_t = \mu_t + \varepsilon_t, \quad \mu_{t+1} = \mu_t + \beta_t + \eta_t, \quad \beta_{t+1} = \beta_t + \zeta_t$
  - state $\alpha_t = (\mu_t, \beta_t)'$: the level and the slope (the growth rate)
  - $Z = (1,\ 0), \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad Q = \text{diag}(\sigma_\eta^2, \sigma_\zeta^2)$
- Special cases
  - $\sigma_\eta^2 = \sigma_\zeta^2 = 0$: a deterministic linear trend $\mu_1 + \beta_1 t$ (Chapter 3)
  - $\sigma_\zeta^2 = 0$: a random walk with drift β
  - $\sigma_\eta^2 = 0$: the **smooth trend** (integrated random walk), the model behind the HP filter (Section 5)
- Forecasts: $\hat{y}_{T+h} = \hat{\mu}_T + h\hat{\beta}_T$, the state space version of Holt's method (Chapter 0)

ARMA models in state space form

- **AR(2)**: $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$; state $\alpha_t = (y_t, y_{t-1})'$
  - $y_t = (1,\ 0)\,\alpha_t$ $(H = 0)$, $\quad \alpha_{t+1} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_{t+1}$
  - T is the companion matrix of Chapter 2; its eigenvalues are the inverse AR roots
- **ARMA(1,1)**: $y_t = \phi y_{t-1} + u_t + \theta u_{t-1}$ with $\alpha_t = (y_t, \theta u_t)'$
  - $T = \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix}$, $R = (1,\ \theta)'$; any ARMA(p, q) fits with $m = \max(p, q+1)$
- Why it matters: the Kalman filter gives the **exact** Gaussian likelihood of an ARMA model, with missing values allowed
  - the stationary start: P_1 solves $P = TPT' + RQR'$, i.e. $\mathrm{vec}(P_1) = (I - T \otimes T)^{-1}\mathrm{vec}(RQR')$
  - this is how `statsmodels` `SARIMAX` estimates the ARIMA models of Chapters 2–4

Worked example: AR(2) for US GDP growth

- Data: US real GDP growth, quarter on quarter, 1947–2019 ($T = 291$, FRED GDPC1)
  - AR(2) with a constant by `SARIMAX`: $\hat{\phi}_1 = 0.319$, $\hat{\phi}_2 = 0.111$, $\hat{\sigma}^2 = 0.739$, mean $\hat{\mu} = 0.774\%$ per quarter
- Our own filter, with the state space matrices of the previous slide and the stationary P_1
  - log-likelihood: Kalman -368.9157; `SARIMAX` -368.9157: the same number
- The prediction errors v_t of the filter are the one-step forecast errors of the AR(2); the likelihood is built from them (Section 4)

Regression with time-varying parameters

- **Time-varying parameter (TVP) regression**: $y_t = \alpha + \beta_t x_t + \varepsilon_t, \quad \beta_{t+1} = \beta_t + \eta_t$
  - state $(\alpha, \beta_t)'$; the row $Z_t = (1,\ x_t)$ **changes over time**; $Q = \mathrm{diag}(0, \sigma_\eta^2)$
  - $\sigma_\eta^2 = 0$: ordinary least squares (OLS), computed recursively
- Examples
  - a market beta that rises in crises; the pass-through of the exchange rate to inflation; the persistence of inflation
- Better than a rolling window: no window length to choose, every observation is used, and β_t comes with a standard error

Exponential smoothing as a state space model

- Chapter 0: simple exponential smoothing (SES) updates a level $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$ and forecasts $\hat{y}_{t+1} = \ell_t$
  - equivalently $\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1})$: old level plus a share α of the surprise
- Muth (1960): SES gives the optimal forecasts when the data follow the local level model
  - Section 3 shows it: α is the steady-state Kalman gain
- ETS models (error, trend, seasonal) are state space models with a single source of error Hyndman et al. (2008); the models here have separate shocks for each component

Recap: The state space form

- $y_t = Z_t\alpha_t + \varepsilon_t$, $\alpha_{t+1} = T\alpha_t + R\eta_t$: a hidden VAR(1) observed with noise
- Local level: $q = \sigma_\eta^2/\sigma_\varepsilon^2$; reduced form ARIMA(0,1,1)
- Local linear trend, ARMA and TVP regression are other choices of Z_t, T, Q
- The Kalman likelihood of an AR(2) equals the exact ARMA likelihood

The idea of the filter

- Notation: $Y_t = (y_1, \ldots, y_t)$; we track two numbers (vectors) for the state
  - prediction: $a_t = E(\alpha_t \mid Y_{t-1})$ with variance P_t
  - filtered estimate: $a_{t|t} = E(\alpha_t \mid Y_t)$ with variance $P_{t|t}$
- Each period has two steps
  - **update**: y_t arrives; correct a_t by a share of the surprise $v_t = y_t - Z_t a_t$
  - **prediction**: push the filtered state through the transition equation to get a_{t+1}
- With Gaussian errors, $a_{t|t}$ is the exact conditional mean; without normality it is still the best **linear** estimate Durbin and Koopman (2012)
- Start: a_1, P_1; for a non-stationary state we use a **diffuse** start (P_1 very large: we know nothing)

The two steps in formulas

- **Update** (when y_t is observed)
  - prediction error $v_t = y_t - Z_t a_t$, with variance $F_t = Z_t P_t Z_t' + H$
  - **Kalman gain** $K_t = P_t Z_t' F_t^{-1}$
  - $a_{t|t} = a_t + K_t v_t, \quad P_{t|t} = P_t - K_t Z_t P_t$
- **Prediction**
  - $a_{t+1} = T a_{t|t}, \quad P_{t+1} = T P_{t|t} T' + R Q R'$
- **Local level**: $F_t = P_t + \sigma_\varepsilon^2, \quad K_t = P_t/(P_t + \sigma_\varepsilon^2)$
  - $a_{t+1} = a_t + K_t(y_t - a_t), \quad P_{t+1} = P_t(1 - K_t) + \sigma_\eta^2$
- Missing y_t: skip the update ($a_{t|t} = a_t$, $P_{t|t} = P_t$) and predict

The Kalman gain

- Local level: $K_t = \dfrac{P_t}{P_t + \sigma_\varepsilon^2}$ is the weight of the new observation, between 0 and 1
  - $a_{t|t} = (1 - K_t)\, a_t + K_t\, y_t$: a weighted average of the prediction and the observation
- Two uncertainties compete
  - noisy measurements (large σ_ε^2): small K_t, trust the model
  - uncertain state (large P_t, e.g. at the start or after missing data): large K_t, trust the data
- The variance always falls with an observation: $P_{t|t} = (1 - K_t)P_t \leq P_t$
  - and grows by σ_η^2 in the prediction step: the two forces settle at a steady state
- The same formula as the weighting of two independent estimates by their precisions $1/P_t$ and $1/\sigma_\varepsilon^2$

Worked example: three periods by hand (1/2)

- Local level with $\sigma_\varepsilon^2 = 4$, $\sigma_\eta^2 = 1$ ($q = 0.25$); prior $a_1 = 10$, $P_1 = 4$; data $y = (12, 11, 14)$
  - $t = 1$: $F_1 = 4 + 4 = 8$, $K_1 = 4/8 = 0.5$, $v_1 = 12 - 10 = 2$
  - $a_{1|1} = 10 + 0.5 \cdot 2 = 11$, $P_{1|1} = 4(1 - 0.5) = 2$; prediction $a_2 = 11$, $P_2 = 2 + 1 = 3$
- $t = 2$: $F_2 = 3 + 4 = 7$, $K_2 = 3/7 = 0.429$, $v_2 = 11 - 11 = 0$
  - no surprise: the level stays at $a_{2|2} = 11$, but the uncertainty falls to $P_{2|2} = 3 \cdot 4/7 = 1.71$
  - prediction $a_3 = 11$, $P_3 = 1.71 + 1 = 2.71$
- The gain falls from 0.5 to 0.429: the filter learns about the level and relies less on each new observation

Worked example: three periods by hand (2/2)

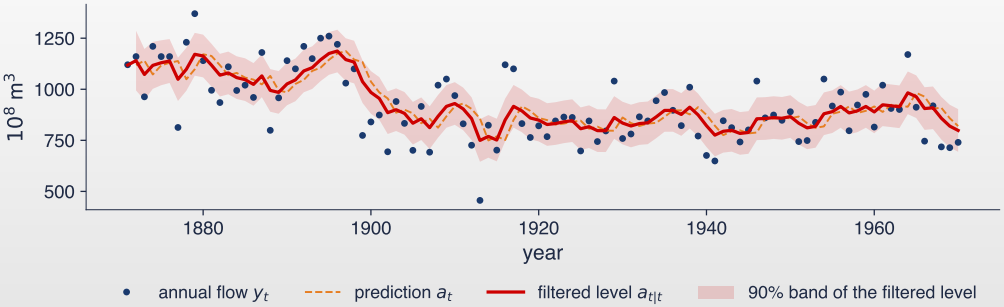
t	y_t	a_t	P_t	F_t	K_t	v_t	$a_{t\mid t}$	$P_{t\mid t}$
1	12	10.00	4.00	8.00	0.500	2	11.00	2.00
2	11	11.00	3.00	7.00	0.429	0	11.00	1.71
3	14	11.00	2.71	6.71	0.404	3	12.21	1.62

- $t = 3$: $K_3 = 2.71/6.71 = 0.404$, $v_3 = 14 - 11 = 3$, $a_{3\mid 3} = 11 + 0.404 \cdot 3 = 12.21$
  - forecast for $t = 4$: $a_4 = 12.21$ with $P_4 = 2.62$; the forecast of y_4 has variance $P_4 + \sigma_\varepsilon^2$
- The gain keeps falling towards the steady state $\bar{K} = 0.390$ ($\bar{P} = 2.56$, next slide)
  - the surprise of 3 moves the level by only $0.404 \cdot 3$: part of it is treated as noise

Steady state and exponential smoothing

- For the local level model, P_t converges to the solution of $\bar{P} = \bar{P}(1 - \bar{K}) + \sigma_\eta^2$ with $\bar{K} = \bar{P}/(\bar{P} + \sigma_\varepsilon^2)$
  - solution: $\bar{P} = \sigma_\varepsilon^2 \dfrac{q + \sqrt{q^2 + 4q}}{2}, \quad \bar{K} = \dfrac{\bar{P}}{\bar{P} + \sigma_\varepsilon^2}$
  - example: $q = 0.25$ gives $\bar{P}/\sigma_\varepsilon^2 = (0.25 + \sqrt{1.0625})/2 = 2.56/4$ and $\bar{K} = 0.390$
- In the steady state: $a_{t+1} = a_t + \bar{K}(y_t - a_t) = \bar{K} y_t + (1 - \bar{K}) a_t$
  - this is SES with $\alpha = \bar{K}$: exponential smoothing is the Kalman filter of a local level model Muth (1960)
- Inverse map: an SES weight α corresponds to $q = \alpha^2/(1 - \alpha)$
  - the state space version adds what SES lacks: a likelihood for q, forecast intervals, missing values

The Nile: the Kalman filter of the local level model



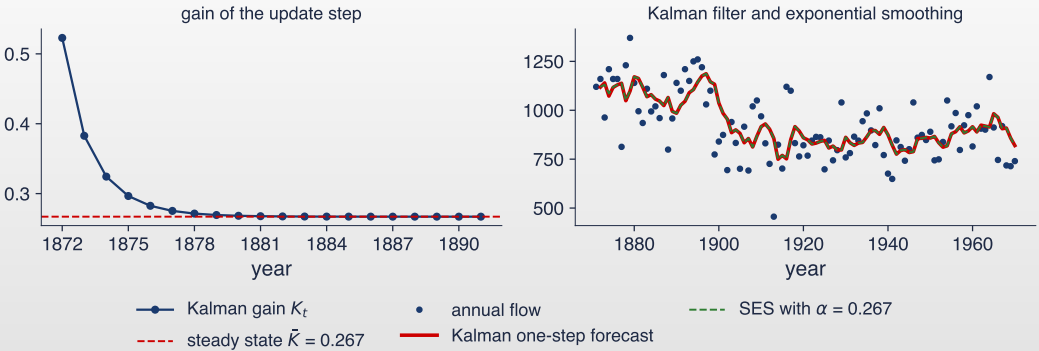
- Annual flow at Aswan, 1871–1970 (10^8 m^3, statsmodels); variances by maximum likelihood with a diffuse start: $\hat{\sigma}^2_\varepsilon = 15100$, $\hat{\sigma}^2_\eta = 1468$, $\hat{q} = 0.097$, the estimates of Durbin and Koopman (2012)

TSA_ch10_kalman_filter

Interpreting the Nile filter

- Most of the year-to-year movement is noise: $\hat{q} = 0.097$, so each flow moves the level by about a quarter of the surprise ($\bar{K} = 0.267$)
  - `statsmodels` `UnobservedComponents('llevel')` gives 15078 and 1479: the same model, a slightly different start
- 1899: the flow is 774, the predicted level 1133; the surprise $v = -359$ lowers the level only to 1037
  - the filter needs several low years to accept the new level (834 in 1905): the first Aswan dam was completed in 1902
- In 1970 the level is 798 with standard error 63, well below the sample mean 919

Convergence of the Kalman gain and equivalence with SES



⊡ Left: K_t of the Nile filter; right: the Kalman one-step forecasts a_t and SES with $\alpha = \bar{K}$, started at the first observation

TSA_ch10_kalman_filter

Interpreting the gain

- K_t starts at 0.523 (after the diffuse start), 0.297 in year 5, and is within 0.001 of $\bar{K} = 0.267$ from year 11 on
  - the forecasts differ by 0.24 in year 5 and by less than 0.02 after 20 years
- SES fitted directly by least squares (Chapter 0) gives $\hat{\alpha} = 0.246$, i.e. $q = \alpha^2/(1-\alpha) = 0.080$
  - close to the likelihood answer: two criteria, one model
- The state space view explains SES: α is large when the level moves a lot relative to the noise

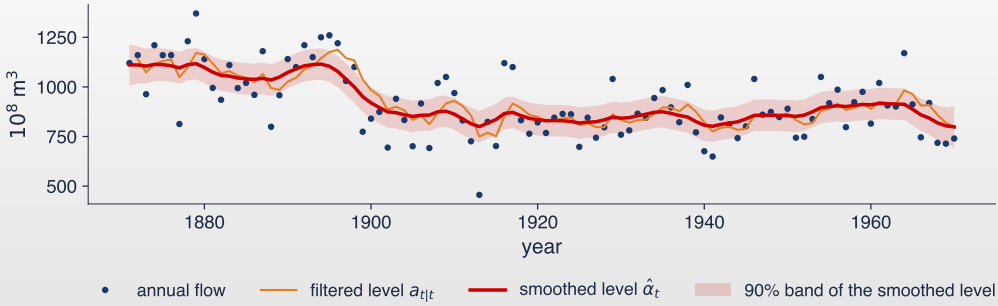
Recap: The Kalman filter

- Update: $v_t = y_t - Z_t a_t$, $F_t = Z_t P_t Z_t' + H$, $K_t = P_t Z_t' F_t^{-1}$, $a_{t|t} = a_t + K_t v_t$
- Prediction: $a_{t+1} = T a_{t|t}$, $P_{t+1} = T P_{t|t} T' + RQR'$
- The gain weighs the data against the model; it settles at a steady state
- Local level in steady state = SES with $\alpha = \bar{K}$; Nile: $\bar{K} = 0.267$

Smoothing: using the whole sample

- **Filtering** uses Y_t (the past and the present): the real-time estimate
  - **smoothing** uses Y_n (the whole sample): $\hat{\alpha}_t = E(\alpha_t \mid Y_n)$, with variance $V_t \leq P_{t|t}$
- **Rauch–Tung–Striebel** smoother Rauch, Tung and Striebel (1965): one backward pass after the filter, from $t = n-1$ to 1
  - n: the sample size; J_t: the smoother gain, the weight of the revision coming from the future
  - $J_t = P_{t|t} T' P_{t+1}^{-1}, \quad \hat{\alpha}_t = a_{t|t} + J_t(\hat{\alpha}_{t+1} - a_{t+1})$
  - $V_t = P_{t|t} + J_t(V_{t+1} - P_{t+1})J_t'$; at $t = n$ smoothed and filtered coincide
- Use: the filtered state for decisions in real time and forecasting; the smoothed state for history (dating breaks, the past output gap)

The Nile: filtered against smoothed level



- Filtered level $a_{t|t}$ and smoothed level $\hat{\alpha}_t$ with its 90% band; the same maximum-likelihood variances

TSA_ch10_smoothing_likelihood

Interpreting the smoothed level

- The smoothed level reacts to the 1898–1899 drop **before** it happens in the data: in 1897 it is already 1038, against 1145 filtered
  - it knows the future: useful for history, impossible in real time
- In 1900: smoothed 920, filtered 985; the smoothed level adapts faster because it sees the low years after 1900
  - standard error in 1920: 48 smoothed against 63 filtered
- The smoothed path shows a step near 1899 more clearly than any moving average of Chapter 0

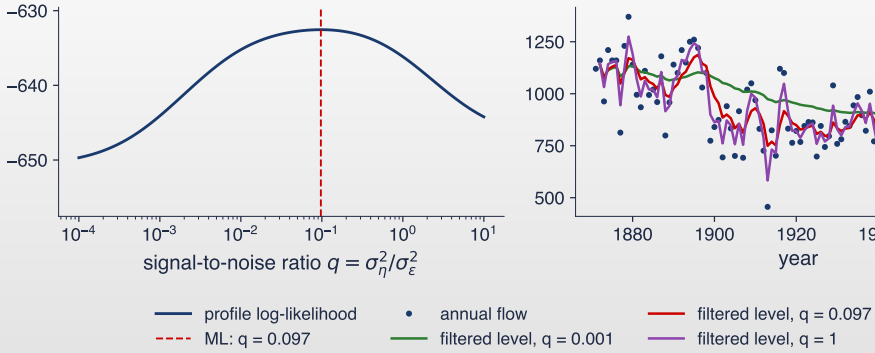
The likelihood: prediction-error decomposition

- The joint density factorises into one-step conditional densities: $p(y_1, \ldots, y_n) = \prod_t p(y_t \mid Y_{t-1})$
  - with Gaussian errors, $y_t \mid Y_{t-1} \sim N(Z_t a_t, F_t)$: the filter delivers the mean and the variance
- **Prediction-error decomposition**
  - $\log L(\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{n}\left(\log F_t + \frac{v_t^2}{F_t}\right)$
  - θ: the parameters (the variances of the model); v_t, F_t: the prediction errors and their variances, from the filter
  - one run of the filter gives $\log L$ for one value of θ; a numerical optimiser searches over θ
- Diffuse start: the first d terms (here $d = 1$ for the level) are left out, since F_t is huge there Durbin and Koopman (2012)
- The same idea as the exact ARMA likelihood of Chapter 2, now for any state space model

Maximum likelihood in practice

- Optimise over $\log \sigma^2$ (variances stay positive); several starting values, since the surface can be flat
  - a variance estimated at zero is common and meaningful: that component is deterministic
- **Concentrating**: for the local level, write $\sigma_\eta^2 = q\sigma_\varepsilon^2$, for fixed q, $\hat{\sigma}_\varepsilon^2 = \frac{1}{n-1}\sum_t v_t^2/F_t^*$ in closed form, with v_t, F_t^* from the filter run with $\sigma_\varepsilon^2 = 1$, $\sigma_\eta^2 = q$
  - the profile log-likelihood is then a function of q alone (next chart)
- Standard errors from the numerical Hessian; model choice by AIC and BIC; tests on the standardised prediction errors $e_t = v_t/\sqrt{F_t}$
  - `statsmodels`: `UnobservedComponents`, `SARIMAX`, `DynamicFactor` all use this machinery

Profile likelihood and the choice of q



- Left: profile log-likelihood of the Nile local level model against q (log scale); right: filtered levels for $q = 0.001$, the ML value and $q = 1$

TSA_ch10_smoothing_likelihood

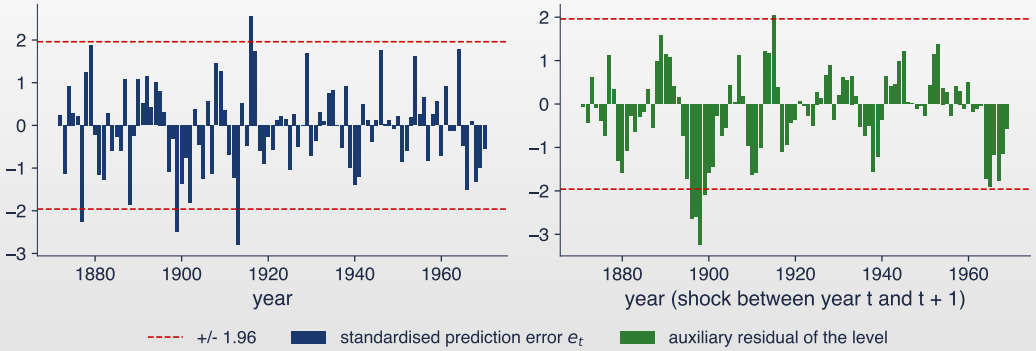
Interpreting the profile likelihood

- A clear maximum at $\hat{q} = 0.097$, with $\hat{\sigma}^2_\varepsilon = 15100$ from the closed form
  - likelihood-ratio statistic against $q = 0.001$ (an almost constant level): 23.1; against $q = 1$: 7.2
- Small q: the level is almost flat and misses the drop after 1899; large q: the level chases every flood
  - the likelihood picks the compromise that predicts best one step ahead
- Choosing q is the same as choosing the SES weight α in Chapter 0, now with a statistical criterion

Diagnostics

- If the model is right, the standardised prediction errors $e_t = v_t/\sqrt{F_t}$ are i.i.d. $N(0,1)$
  - the checks of Chapters 1–2: the ACF and the Ljung–Box test (independence), Jarque–Bera (normality), a plot over time (outliers)
- **Auxiliary residuals** Durbin and Koopman (2012): the smoothed disturbances $\hat{\varepsilon}_t$ and $\hat{\eta}_t$, standardised
  - a large $\hat{\varepsilon}_t$: an **outlier** in one observation
  - a large $\hat{\eta}_t$: a **break** in the level, a permanent shift
- Computed by a backward recursion (the disturbance smoother) after the filter

Diagnostics of the Nile model



- Left: standardised one-step prediction errors e_t; right: standardised auxiliary residual of the level, $\hat{\eta}_t$ (a shock between year t and $t+1$); dashed: ± 1.96

TSA_ch10_smoothing_likelihood

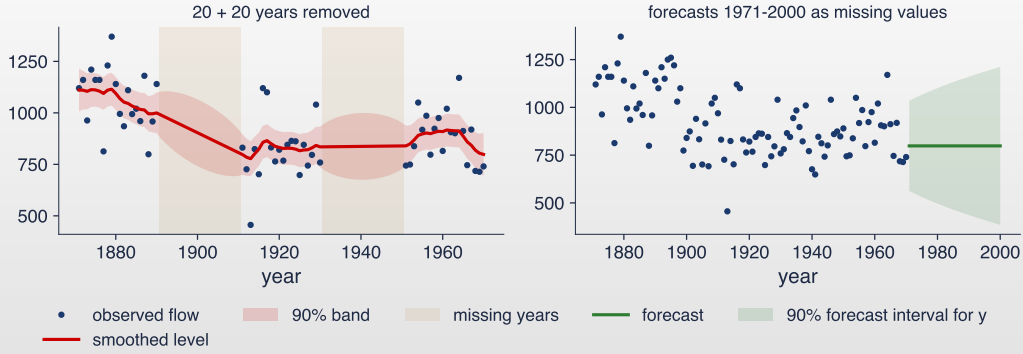
Interpreting the diagnostics

- Prediction errors: Ljung–Box $Q(10) = 13.2$ ($p = 0.21$), Jarque–Bera 0.05 ($p = 0.98$); 4 of 99 outside ± 1.96
  - no evidence against the model in the one-step errors
- The level residual has its minimum in 1898 (-3.23): the shift of the level between 1898 and 1899
  - the observation residual has its minimum in 1913 (-3.04): a single very dry year, an outlier
- The same break as in Chapter 8 (spurious long memory of the Nile), now located by the model itself

Missing values and forecasting

- A missing y_t: no update; $a_{t+1} = Ta_t$, $P_{t+1} = TP_tT' + RQR'$
  - the likelihood simply skips that term; nothing is interpolated by hand
- Forecasting h steps ahead = treating $y_{n+1}, \ldots, y_{n+h}$ as missing
  - local level: $\hat{y}_{n+h} = a_{n+1}$ for every h, with variance $P_{n+1} + (h-1)\sigma_\eta^2 + \sigma_\varepsilon^2$
- Uses: irregular data (holidays), series that start at different dates, mixed frequencies, the ragged edge of a nowcast
  - the smoother fills the gaps with $\hat{\alpha}_t$ and the corresponding standard error

Missing years and forecasts of the Nile



- Left: 1891–1910 and 1931–1950 removed (the experiment of Durbin and Koopman (2012)), smoothed level with 90% band; right: forecasts for 1971–2000 with 90% intervals for y

TSA_ch10_missing_data

Interpreting the missing data

- Inside a gap the smoothed level is a straight line between the two edges; its standard error grows to 99 in 1900, against 48 with all data
  - the band widens towards the middle of the gap and narrows near observed years
- Forecasts are flat at 798: a random-walk level has no predictable direction
  - standard error of y: 144 one year ahead, 251 after 30 years
- The same code handles missing data, forecasts and estimation: one of the main practical strengths of state space models

Recap: Smoothing, likelihood and missing data

- Filtered: real time; smoothed (RTS): the whole sample, smaller variance
- $\log L = -\frac{1}{2}\sum(\log 2\pi + \log F_t + v_t^2/F_t)$, maximised numerically
- Diagnostics on e_t; auxiliary residuals find outliers and breaks (Nile 1898–1899)
- Missing data and forecasts: skip the update step

Unobserved components models

- **Structural time series** or **unobserved components** (UC) model Harvey (1989): $y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t$
  - trend μ_t (local level or local linear trend), cycle ψ_t, seasonal γ_t, irregular ε_t
  - the decomposition of Chapter 0, but each component is a stochastic process with its own variance
- The **cycle**: a stationary AR(2) $\psi_t = \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2} + \kappa_t$, or a damped stochastic cycle with period $2\pi/\lambda_c$
  - ϕ_1, ϕ_2: the AR coefficients of the cycle; κ_t: the cycle shock; λ_c: the frequency of the cycle, in radians
  - for log GDP, ψ_t is the **output gap**: the percentage deviation of output from its trend (potential)
- In Python: `UnobservedComponents(y, level='smooth trend', autoregressive=2)`; the whole state is estimated by the Kalman filter and smoother

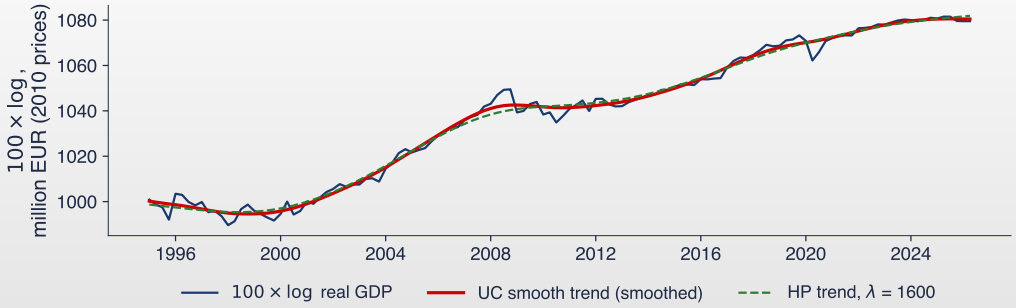
The HP filter is a state space smoother

- Hodrick and Prescott (1997): the trend τ_t minimises $\sum_t (y_t - \tau_t)^2 + \lambda \sum_t (\Delta^2 \tau_t)^2$; $\lambda = 1600$ for quarterly data
  - $\Delta^2 \tau_t = \tau_t - 2\tau_{t-1} + \tau_{t-2}$: the change in the slope of the trend; $\lambda \geq 0$: the smoothing parameter
  - large λ: a straight-line trend; $\lambda = 0$: the trend is the series itself
- This is exactly the **smoothed** trend of a UC model: smooth trend ($\Delta^2 \mu_{t+1} = \zeta_{t-1}$) plus white noise, with $\lambda = \sigma_\varepsilon^2 / \sigma_\zeta^2$ Harvey and Jaeger (1993)
  - the HP filter imposes the signal-to-noise ratio and a white-noise gap; the UC model estimates both
- Being a smoother, HP uses future data: its last values are revised as new quarters arrive

Hamilton (2018): why you should never use the HP filter

- Hamilton (2018) lists three problems
  - the HP gap of a random walk has cycles that are not in the data: spurious dynamics
  - the end of the sample is treated differently from the middle: large revisions, misleading real-time gaps
  - $\lambda = 1600$ is a convention, far from the value a likelihood would choose
- His alternative, the **regression filter**: OLS of y_{t+h} on $1, y_t, y_{t-1}, y_{t-2}, y_{t-3}$ with $h = 8$ quarters
  - the cycle is the residual: what could not be predicted two years earlier; one-sided by construction
- No method is the truth: we compare the UC model, HP and the Hamilton filter on Romanian GDP

Romanian GDP: trend



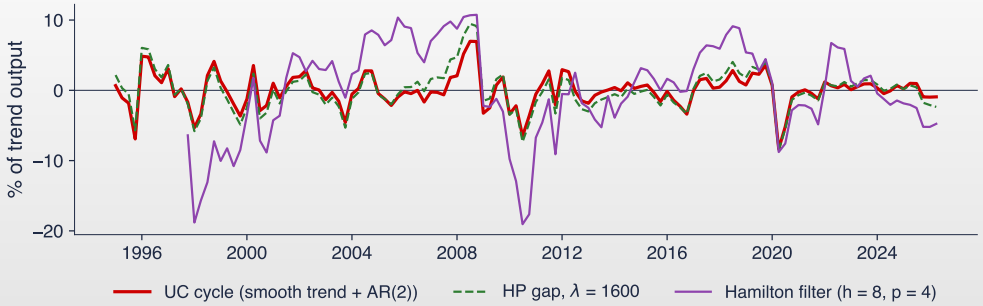
- 100 × log of real GDP, quarterly, seasonally and calendar adjusted, chain-linked volumes (Eurostat namq_10_gdp), 1995Q1–2026Q2 ($T = 126$); UC model: smooth trend plus AR(2) cycle; HP with $\lambda = 1600$

TSA_ch10_trend_cycle

Interpreting the trend

- Average growth 2.5% per year, with a slow start (the 1997–1999 recession), a boom until 2008, a stagnation and a recovery after 2013
  - the two trends are almost identical: the difference between methods lies in the gap, not in the trend
- UC estimates: $\hat{\sigma}^2_\zeta = 0.078$ (trend slope), cycle AR(2) with $\hat{\phi}_1 = 0.61$, $\hat{\phi}_2 = -0.13$, $\hat{\sigma}^2_\kappa = 5.27$; the irregular variance is estimated at zero
  - the ratio of the shock variances, cycle to trend slope, $\hat{\sigma}^2_\kappa / \hat{\sigma}^2_\zeta = 68$, is far below the 1600 ($\sigma^2_\varepsilon / \sigma^2_\zeta$) that HP imposes: a more flexible trend
- A flexible trend absorbs part of the boom: the UC gap will be smaller than the HP gap

The Romanian output gap: three methods



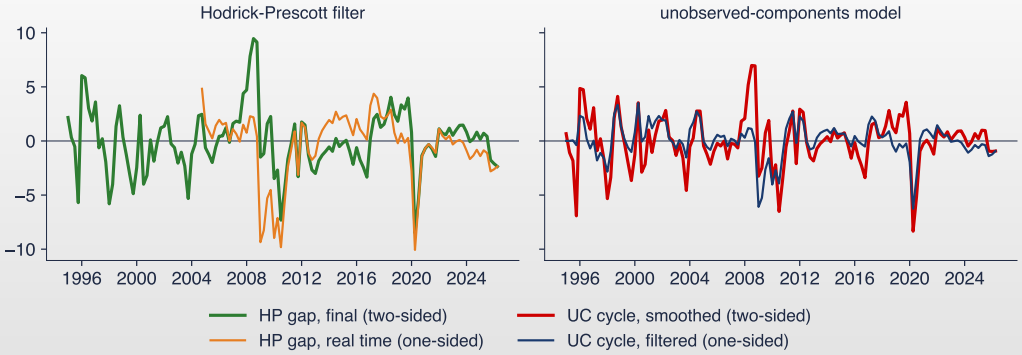
- Gap in % of trend: UC cycle (smoothed), HP gap ($\lambda = 1600$) and the Hamilton filter ($h = 8$, four lags), the first 11 quarters lost

TSA_ch10_trend_cycle

Interpreting the output gaps

- All three see the overheating before 2008 (peaks: UC 7.0% in 2008Q3, HP 9.5%, Hamilton 10.7%) and the 2020 pandemic (UC -8.3%, HP -8.5%)
  - correlations: UC–HP 0.94, UC–Hamilton 0.57, HP–Hamilton 0.69
- Standard deviations: UC 2.4, HP 2.9, Hamilton 6.5: the Hamilton gap also contains every surprise of two years
  - the four regression coefficients sum to 0.97: GDP is close to a random walk with drift
- Latest gap (2026Q2): UC -0.9%, HP -2.4%, Hamilton -4.8%: the sign agrees, the size does not

The output gap in real time and with the full sample



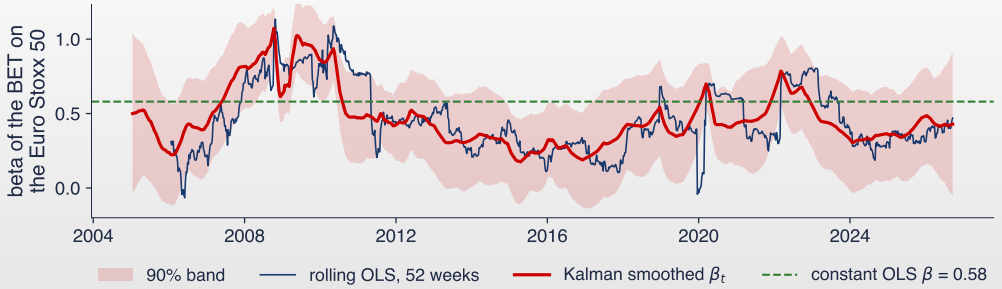
- Left: the HP gap computed with all data (two-sided) and the last value of the HP gap computed with the data available at each date (one-sided, from 2004); right: the UC cycle, filtered and smoothed

TSA_ch10_trend_cycle

Interpreting the revisions

- In 2008Q3 the real-time HP gap was 2.2%; with hindsight it is 9.5%: the overheating was invisible in real time
  - the UC model shows the same problem: filtered 1.1%, smoothed 7.0%
- Root mean square revision since 2004: HP 2.8 points, UC 1.8 points; correlation real time–final: HP 0.57, UC 0.65
  - the end-point problem of Hamilton (2018); the UC model at least reports its own uncertainty
- For policy in real time, use the filtered (one-sided) estimate and its standard error, never the last point of a two-sided filter

A time-varying beta: the BET and the euro area



- Weekly log returns of the BET on the Euro Stoxx 50 (EODHD), 2005–2026 ($T = 1\,126$ weeks)
- Random-walk beta by maximum likelihood, Kalman smoothed, with a 90% band; 52-week rolling OLS; constant OLS

TSA_ch10_tvp_regression

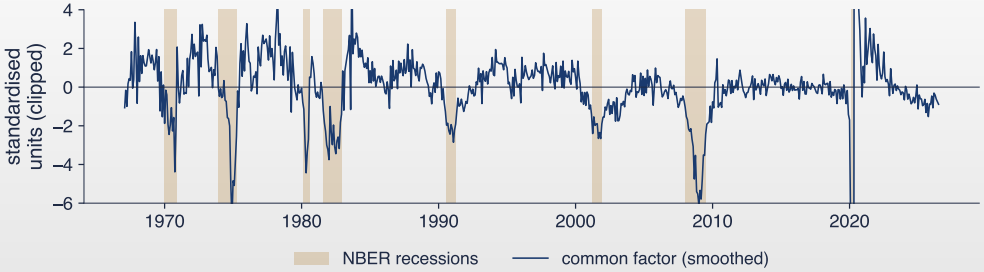
Interpreting the time-varying beta

- The constant OLS beta is 0.58; the smoothed beta peaks at 1.07 (October 2008) and falls to 0.17 (April 2015)
  - in the 2008 crisis the Romanian market moved almost one for one with the euro area; later it decoupled
- Likelihood ratio against a constant beta: 46.1 (one restriction, a variance on the boundary): strong evidence of variation
  - weekly standard deviation of the beta shocks 0.059; latest beta 0.43 with standard error 0.29
- The rolling window jumps when a single extreme week enters or leaves it; the Kalman beta moves only as much as the likelihood allows

Dynamic factor models and nowcasting

- **Dynamic factor model**: $y_{it} = \lambda_i f_t + u_{it}, \quad f_t = \phi_1 f_{t-1} + \phi_2 f_{t-2} + \eta_t$
  - many series $y_{1t}, \ldots, y_{Nt}$ driven by one unobserved common factor f_t (the state) with loadings λ_i
  - u_{it}: the part of series i not explained by the factor; ϕ_1, ϕ_2: the AR coefficients of the factor
  - Stock and Watson (1989): a coincident index of US activity from four monthly indicators
- **Nowcasting** Giannone, Reichlin and Small (2008): estimate the current quarter before GDP is published
  - monthly data arrive at different dates: the latest months of some series are missing (the **ragged edge**)
  - the Kalman filter treats them as missing values and updates the factor with whatever has arrived
- Central banks use such models to nowcast GDP; `statsmodels` has `DynamicFactor` and `DynamicFactorMQ`

A common factor of US activity



- Monthly growth of industrial production, payroll employment, real income less transfers and real manufacturing and trade sales (FRED), standardised
- One AR(2) factor estimated on 1967–2019 and run to July 2026; the last two months of income and sales set to missing

TSA_ch10_dynamic_factor

Interpreting the common factor

- Loadings: employment 0.59, industrial production 0.42, sales 0.29, income 0.23; factor AR(2) with $\hat{\phi}_1 = 0.42$, $\hat{\phi}_2 = 0.41$
  - correlation with the simple average of the four series: 0.86
- Mean of the factor: -3.5 in NBER recession months, 0.4 in expansions; negative in 99% of recession months
  - April 2020: -74 standard units, off the scale of the chart
- Latest value (July 2026): -0.9, computed although two of the four series are not yet available

Recap: Trend, cycle and common factors

- UC models: trend + cycle + seasonal + irregular, each with its own variance
- HP = smoother of a UC model with imposed λ; Hamilton (2018): spurious cycles and end-point revisions
- Romanian output gap: the methods agree on the sign, not on the size; real-time gaps are revised heavily
- TVP regression and dynamic factors are state space models too

Hamilton (1989) and the dating of recessions



The NBER, Cambridge (Massachusetts), whose committee dates US recessions

Photo: Astrophobe (2022); CC BY-SA 4.0; Wikimedia Commons

- The National Bureau of Economic Research (NBER) dates US recessions by a committee, months after the fact
  - a recession is a regime: GDP behaves differently while it lasts
- Hamilton (1989): let the mean of GDP growth depend on an unobserved regime $S_t \in \{1, 2\}$ that follows a Markov chain
  - the data then tell us the probability of a recession in each quarter, without a committee
- The state is now **discrete**; the Kalman filter is replaced by the Hamilton filter

The Markov-switching model

- **Switching mean**: $y_t = \mu_{S_t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2_{S_t}), \quad S_t \in \{1, \ldots, k\}$
  - **Markov chain**: $p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$, constant over time; the rows of the transition matrix sum to 1
- Hamilton's MS-AR(4): $y_t - \mu_{S_t} = \sum_{j=1}^{4} \phi_j (y_{t-j} - \mu_{S_{t-j}}) + \varepsilon_t$
  - the AR dynamics stay the same; only the mean switches
- Variants: switching variance (calm and turbulent markets), switching AR coefficients, $k = 3$ regimes
  - in Python: `MarkovRegression`, `MarkovAutoregression` (`statsmodels`)
- A nonlinear model: the forecast depends on the probability of each regime

Transition matrix, durations and ergodic probabilities

- Two regimes: $\mathbf{P} = \begin{pmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{pmatrix}$
  - h-step transitions: $\mathbf{P}^h$
- **Expected duration** of regime i: the time spent there is geometric, $\Pr(D=d) = p_{ii}^{d-1}(1-p_{ii})$, so $E(D) = 1/(1-p_{ii})$
  - $p_{11} = 0.75$: recessions last $1/0.25 = 4$ quarters on average; $p_{22} = 0.95$: expansions last 20 quarters
- **Ergodic** (long-run) probabilities: $\pi_1 = \dfrac{1-p_{22}}{2-p_{11}-p_{22}}$, the share of time spent in regime 1
  - example: $\pi_1 = 0.05/0.30 = 1/6$: one quarter in six is a recession quarter

The Hamilton filter

- Track $\xi_{t|t} = \Pr(S_t = j \mid Y_t)$, the **filtered probabilities**, for each regime j
  - **prediction**: $\Pr(S_t = j \mid Y_{t-1}) = \sum_i p_{ij} \Pr(S_{t-1} = i \mid Y_{t-1})$
  - **densities**: $f_j(y_t) = \varphi\big((y_t - \mu_j)/\sigma_j\big)/\sigma_j$, the likelihood of y_t in each regime
  - φ: the standard Normal density; μ_j, σ_j: the mean and standard deviation of regime j
  - **update** (Bayes): $\Pr(S_t = j \mid Y_t) = \dfrac{\Pr(S_t = j \mid Y_{t-1})\, f_j(y_t)}{\sum_i \Pr(S_t = i \mid Y_{t-1})\, f_i(y_t)}$
- The denominator is $p(y_t \mid Y_{t-1})$: the log-likelihood is $\sum_t \log p(y_t \mid Y_{t-1})$, as for the Kalman filter
  - the same predict–update cycle, with probabilities instead of means and variances
- Start: the ergodic probabilities π_j

Smoothed probabilities and estimation

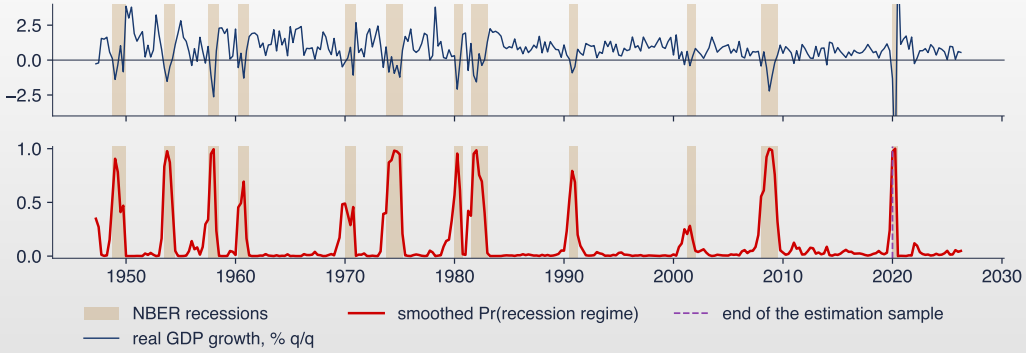
- **Smoothed probabilities** $\Pr(S_t = j \mid Y_n)$: a backward pass Kim (1994), the analogue of the RTS smoother
  - used to date regimes in history; the filtered probabilities are what a forecaster knew at time t
- Estimation: maximum likelihood by numerical optimisation, or the EM algorithm Hamilton (1990)
  - EM: expectation (smoothed probabilities given θ) and maximisation (weighted means, variances, transition counts), repeated
- The likelihood has **several local maxima**: use many random starting values and compare the solutions
  - label switching: regime 1 and regime 2 can swap; name them by their means or variances

Hamilton's specification on today's data

MS-AR(4), US real GDP growth, 1951Q2–1984Q4	Maximum A	Maximum B
mean, regime 1 / regime 2 (% per quarter)	−0.14 / 1.32	−1.14 / 0.96
p_{11}, p_{22}	0.762, 0.892	0.128, 0.952
expected duration, regime 1 / 2 (quarters)	4.2 / 9.3	1.1 / 20.8
log-likelihood	−189.47	−189.16
NBER recession quarters found (smoothed prob. > 0.5)	94%	10%

- Maximum A is Hamilton's answer: a recession regime of about four quarters that matches the NBER dates
- Maximum B has a slightly higher likelihood but a different meaning: isolated sharp quarters
- A likelihood difference of −189.16 against −189.47 cannot decide; economic meaning and robustness checks must

US recessions from GDP growth



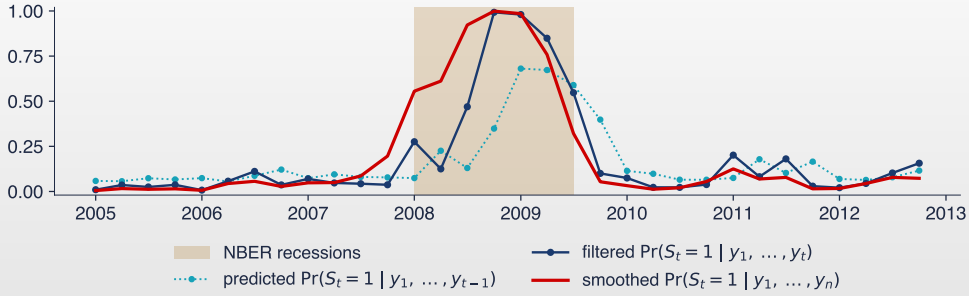
- Two regimes with switching mean, constant variance, estimated on 48 NBER recession quarters out of 1947Q2–2019Q4 (FRED GDPC1, USREC); probabilities after 2019 computed with the same parameters

TSA_ch10_markov_gdp

Interpreting the recession probabilities

- Regime means: -0.42% (recession) and 0.97% (expansion) per quarter, $\hat{\sigma} = 0.79$; $\hat{p}_{11} = 0.685$, $\hat{p}_{22} = 0.949$
  - expected durations: 3.2 quarters of recession, 19.7 quarters of expansion; ergodic share of recession 14%
- Agreement with the NBER: 94% of quarters; 62% of NBER recession quarters detected, 0% false alarms
  - the model misses the mild recessions (2001) and catches the deep ones
- 2020Q2 (-8.2%): probability 1.00, then back to expansion; after 2021 the probability never exceeds 0.12

Filtered against smoothed probabilities: 2008



⊡ Predicted, filtered and smoothed probability of the recession regime, 2005–2012, same model as on the previous chart

TSA_ch10_markov_gdp

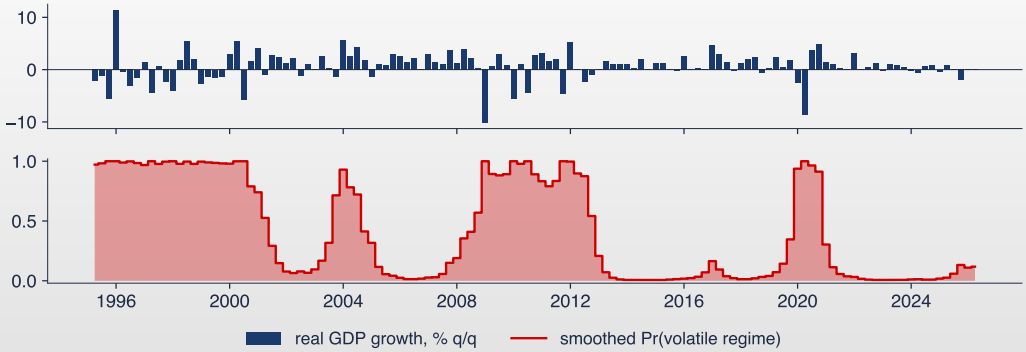
Interpreting the real-time probabilities

- The NBER recession started in 2008Q1; the smoothed probability crosses 0.5 in 2008Q1, the filtered one only in 2008Q4
  - 2008Q1: filtered 0.28, smoothed 0.56; 2008Q3 (growth -0.53%): filtered 0.47, smoothed 0.92
- Concordance with the NBER using only filtered probabilities: 92% of quarters, 54% of recession quarters
  - real-time dating is harder: Chauvet and Piger (2008) compare the methods on real-time data
- Never judge a model's real-time skill from its smoothed probabilities: they use the future

Checks: which regimes did the model find?

- **Switching variance** on 1947–2019: the regimes become high and low volatility, not recession and expansion
  - standard deviations 1.20 and 0.44; durations 45 and 39 quarters; the switch to the calm regime in 1984Q3
  - this is the "Great Moderation" of the mid-1980s; agreement with the NBER falls to 60% (47% false alarms)
- **Including 2020** in the estimation: one regime is the single quarter 2020Q2 (mean -8.2%), the other lasts 315 quarters
  - one extreme observation can capture a whole regime
- Checklist: plot the probabilities against known events; compare several starting values and samples; prefer the specification with a clear economic meaning

Romanian GDP growth regimes



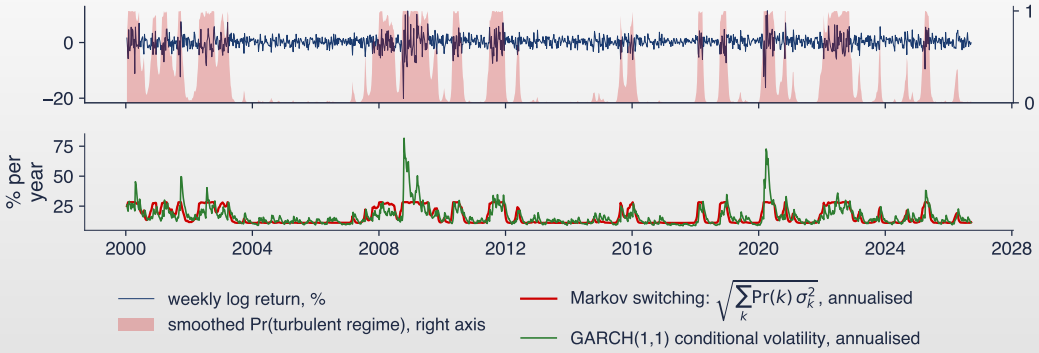
- Quarter-on-quarter growth of Romanian real GDP (Eurostat), two regimes with switching mean and variance; smoothed probability of the volatile regime

TSA_ch10_markov_gdp

Interpreting the Romanian regimes

- Stable regime: mean 1.03% per quarter, standard deviation 1.3; volatile regime: mean 0.07%, standard deviation 3.9
  - expected durations: 15.2 quarters stable, 11.2 quarters volatile; 40% of the quarters are volatile
- Volatile spells: 1995Q2–2001Q2 (transition and the 1997–1999 recession), 2003Q4–2004Q3, 2008Q4–2012Q4 (the crisis and the austerity of 2010), 2020Q1–2020Q4 (the pandemic)
  - latest probability of the volatile regime: 0.12
- With 125 quarters a mean-only model is unstable; regimes of volatility and growth come together in an emerging economy

Calm and turbulent stock markets



⊡ Weekly log returns of the S&P 500, 2000–2026 (EODHD); two regimes with switching mean and variance; bottom: the volatility implied by the regime probabilities against a GARCH(1,1) (Chapter 5), both annualised

TSA_ch10_volatility_regimes

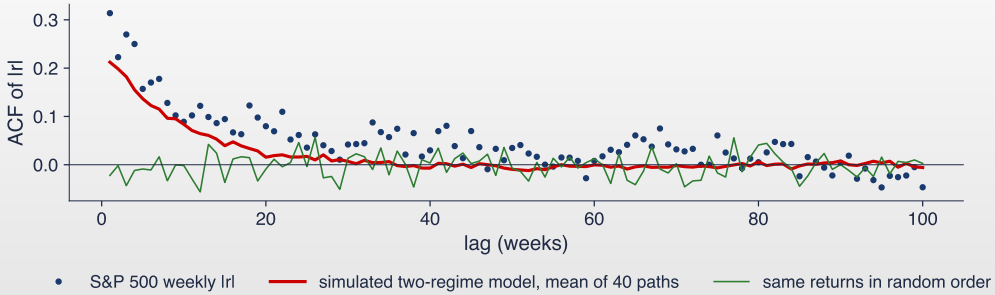
Interpreting the volatility regimes

- Calm regime: volatility 11% per year, mean 0.32% per week; turbulent: 28% per year, mean -0.39% per week
  - durations: 37 weeks calm, 14 weeks turbulent; turbulent in 26% of the weeks (ergodic 28%)
- Most turbulent years: 2022 (90% of the weeks), 2008 (79%), 2009 (69%)
  - correlation with the GARCH(1,1) volatility: 0.68; GARCH: $\hat{\alpha} = 0.26$, $\hat{\beta} = 0.70$
- Two views of the same clustering: GARCH moves volatility continuously, the regime model jumps between two levels

Regimes, GARCH persistence and long memory

- Lamoureux and Lastrapes (1990): shifts in the level of variance make a GARCH look almost integrated ($\alpha + \beta \approx 1$)
  - the near-IGARCH estimates of Chapter 5 may reflect regimes, not one very persistent process
  - Hamilton and Susmel (1994): ARCH within regimes (SWARCH) lowers the persistence
- Diebold and Inoue (2001): rare regime switches produce a slowly decaying ACF and $\hat{d} > 0$ (Chapter 8, spurious long memory)
  - three models, one fact: volatility shocks persist; the next chart simulates the fitted regime model to see how far regimes alone go

Regime switching and spurious long memory



ACF of weekly $|r_t|$ of the S&P 500; mean ACF of 40 series simulated from the fitted two-regime model (constant variance within each regime, no GARCH); the same returns in random order

TSA_ch10_volatility_regimes

Interpreting the simulated regimes

- Data: ACF of $|r_t|$ 0.31 at lag 1, 0.06 at 26 weeks; regime model: 0.21 and 0.01; random order: none
  - the regimes explain most of the short-run clustering, less of the long tail
- Local Whittle $\hat{d}$ of $|r_t|$: 0.32 in the data, 0.29 (SD 0.06) in the simulated regime series, -0.05 in random order
  - a GARCH(1,1) fitted to the simulated series gives $\alpha + \beta = 0.91$ on average (15 series), although there is no GARCH in them
- Persistence and long memory in volatility are consistent with regimes: compare models by likelihood and forecasts, not by one statistic

Recap: Markov switching

- $y_t = \mu_{S_t} + \varepsilon_t$, S_t a Markov chain; durations $1/(1-p_{ii})$; ergodic $\pi_1 = (1-p_{22})/(2-p_{11}-p_{22})$
- Hamilton filter: predict with $\mathbf{P}$, update with Bayes; Kim smoother for history
- US GDP: the model recovers the deep NBER recessions; filtered probabilities react later
- Several maxima, outliers and variance breaks can change what a "regime" means
- Regimes reproduce volatility clustering, GARCH persistence and apparent long memory

Possible contribution of AI

- **Code**: a first draft of a Kalman filter, of the matrices of a new model, of a Markov-switching fit with many starting values
- **Explanation**: a second derivation of the gain, of the steady state, of the Hamilton filter
- **Exploration**: output gaps for many EU countries; regimes in many markets; a nowcast with many indicators
- Example prompt
  - `Write Python code that downloads Romanian quarterly real GDP from Eurostat (namq_10_gdp, Q.CLV10_MEUR.SCA.B1GQ.RO), fits UnobservedComponents with a smooth trend and an AR(2) cycle, compares the smoothed and the filtered cycle with the HP gap and with the Hamilton (2018) regression filter, and reports the revisions of the last eight quarters.`

Checks you must run

- Simulate from known parameters and check that the code recovers them (the Nile values of Durbin and Koopman (2012) are a good test)
- Filtered or smoothed? Real-time claims need filtered estimates; an AI answer that uses `smoothed` for a "forecast" is wrong
- Durations: $1/(1 - p_{ii})$, not $p_{ii}/(1 - p_{ii})$; check the convention of the transition matrix (rows or columns) in the library
- Markov switching: several starting values, label switching, outliers that capture a regime
- The HP filter at the end of the sample: do not read the last gap as a real-time estimate
- Every cited reference: it must exist; check the DOI

Key takeaways

- State space form: a hidden state α_t follows a VAR(1) and is measured with noise; ARMA, ETS, UC, TVP and factor models fit in it
- The Kalman filter predicts and updates; the gain weighs data against model; the local level filter in steady state is SES
- Smoothing for history, filtering for real time; the likelihood comes from the prediction errors; missing data are easy
- Output gaps depend on the method and are revised heavily at the end of the sample
- Markov switching: regime probabilities, durations $1/(1 - p_{ii})$; check what the regimes mean

Key formulas

Quantity	Formula
State space form	$y_t = Z_t\alpha_t + \varepsilon_t, \quad \alpha_{t+1} = T\alpha_t + R\eta_t$
Update	$v_t = y_t - Z_t a_t, \quad F_t = Z_t P_t Z_t' + H, \quad K_t = P_t Z_t' F_t^{-1}, \quad a_{t\|t} = a_t + K_t v_t$
Prediction	$a_{t+1} = Ta_{t\|t}, \quad P_{t+1} = TP_{t\|t}T' + RQR'$
Local level steady state	$\bar{P} = \sigma_\varepsilon^2(q + \sqrt{q^2+4q})/2, \quad \alpha_{SES} = \bar{K} = \bar{P}/(\bar{P} + \sigma_\varepsilon^2)$
Log-likelihood	$-\frac{1}{2}\sum_t(\log 2\pi + \log F_t + v_t^2/F_t)$
HP filter	$\min \sum(y_t - \tau_t)^2 + \lambda \sum(\Delta^2\tau_t)^2, \quad \lambda = \sigma_\varepsilon^2/\sigma_\zeta^2$
Markov chain	$E(D_i) = 1/(1-p_{ii}), \quad \pi_1 = (1-p_{22})/(2-p_{11}-p_{22})$
Hamilton filter	$\Pr(S_t = j \mid Y_t) \propto f_j(y_t)\sum_i p_{ij}\Pr(S_{t-1} = i \mid Y_{t-1})$

Self-assessment

- **Question**: local level with $P_t = 3$, $\sigma_\varepsilon^2 = 1$. What is the Kalman gain?
  - **Answer**: $K_t = 3/(3+1) = 0.75$: the new observation gets weight 0.75
- **Question**: an SES model has $\alpha = 0.5$. Which local level model is behind it?
  - **Answer**: $q = \alpha^2/(1-\alpha) = 0.25/0.5 = 0.5$
- **Question**: $p_{11} = 0.9$ in a two-regime model. How long does regime 1 last on average?
  - **Answer**: $1/(1-0.9) = 10$ periods
- Next: Chapters 11–14 for self-study; Chapter 15, review

References (1/2)

- Chauvet, M., & Piger, J. (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics*, 26(1), 42–49. doi:10.1198/073500107000000296
- Diebold, F. X., & Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1), 131–159. doi:10.1016/S0304-4076(01)00073-2
- Durbin, J., & Koopman, S. J. (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford University Press. doi:10.1093/acprof:oso/9780199641178.001.0001
- Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665–676. doi:10.1016/j.jmoneco.2008.05.010
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384. doi:10.2307/1912559
- Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1–2), 39–70. doi:10.1016/0304-4076(90)90093-9
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. doi:10.2307/j.ctv14jx6sm
- Hamilton, J. D. (2018). Why you should never use the Hodrick–Prescott filter. *The Review of Economics and Statistics*, 100(5), 831–843. doi:10.1162/rest_a_00706
- Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1–2), 307–333. doi:10.1016/0304-4076(94)90067-1
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press. doi:10.1017/CBO9781107049994
- Harvey, A. C., & Jaeger, A. (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics*, 8(3), 231–247. doi:10.1002/jae.3950080302

References (2/2)

- Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16. doi:10.2307/2953682
- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer. doi:10.1007/978-3-031-13584-2
- Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer. doi:10.1007/978-3-540-71918-2
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45. doi:10.1115/1.3662552
- Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1–2), 1–22. doi:10.1016/0304-4076(94)90036-1
- Kim, C.-J., & Nelson, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press. doi:10.7551/mitpress/6444.001.0001
- Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, 8(2), 225–234. doi:10.1080/07350015.1990.10509794
- Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, 55(290), 299–306. doi:10.1080/01621459.1960.10482064
- Rauch, H. E., Tung, F., & Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8), 1445–1450. doi:10.2514/3.3166
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications* (4th ed.). Springer. doi:10.1007/978-3-319-52452-8
- Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, 4, 351–394. doi:10.1086/654119