Lecture 2

Thinking in "Regimes" in economics and finance

Hidden Markov Models

Lars A. Lochstoer

UCLA Anderson School of Management

Overview

- Hidden Markov chain models

- Switching regime models

- Markov-switching State Space Models

- The Unscented Kalman filter for nonlinear SSMs

Motivation

Many economic phenomena can be described as "regimes"

1. Recessions and expansions
2. Bull/bear markets
3. Financial crisis
4. Currency regimes (fixed, floating, crawling peg)
5. High inflation, low inflation
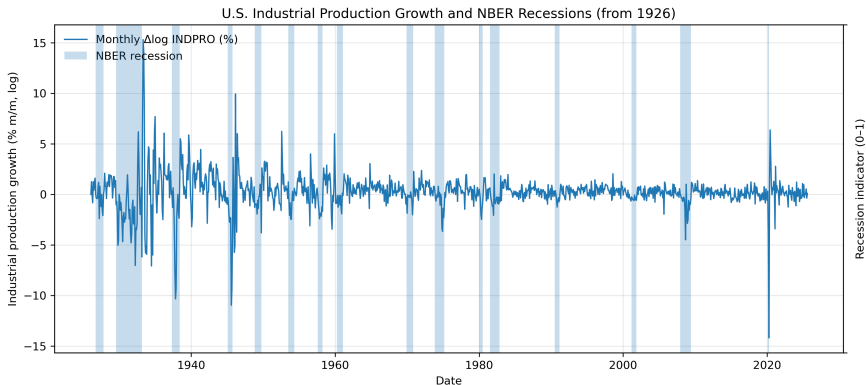6. Uncertainty: technological, geopolitical, financial system, etc.

Common to these: nonlinear dynamics

- Large changes vs small changes
- Different expected duration of different regimes

Motivation: The business cycle

Industrial production data 1926-2025

- Think of this as economic growth data
- Shaded areas denote recessions as determined by the National Bureau of Economic Research (NBER)

U.S. Industrial Production Growth and NBER Recessions (from 1926)

Hidden Markov Models

The Switching Regime (SR) Model

Hidden Markov chain Models (HMMs)

A convenient tool: thinking in terms of states and state transitions

Latent chain (states) $S_t \in \{1, \ldots, K\}$ with transition matrix

$$\Pi = \left[\begin{array}{cccc} \pi_{11} & \pi_{12} & \cdots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{K1} & \pi_{K2} & \cdots & \pi_{KK} \end{array} \right]$$

where $\Pr(S_t = j \mid S_{t-1} = i) = \pi_{ij}$ and $\sum_j \pi_{ij} = 1$.

- Observations y_t have dynamics governed by the state: pdf is $f_{S_t}(y_t; \theta_{S_t})$
- Need prior over the initial state, p_0
- Parameter set $\Theta = \{\Pi, p_0, \theta_1, \ldots, \theta_K\}$.

Simplest case: Two-State Gaussian HMM

Let $f_{S_t}(y_t; \theta_{S_t})$ be $N\left(\mu_{s_t}, \sigma_{s_t}^2\right)$. That is:

$$y_t = \mu_{s_t} + \sigma_{s_t} \varepsilon_t,$$

where $\varepsilon_t \sim N(0, 1)$, and hidden regime $S_t \in \{1, 2\}$ with transition matrix

$$\Pi = \begin{bmatrix} \pi_{11} & 1 - \pi_{11} \\ 1 - \pi_{22} & \pi_{22} \end{bmatrix}$$

Thus, we have 6 parameters to estimate $(\pi_{11}, \pi_{22}, \mu_1, \mu_2, \sigma_1, \sigma_2)$, plus we need a prior for initial state, p_0.

Forecasting with two-state Gaussian HMM

Current beliefs about the state: $p_t(s_t)$ (could write, $p_{t|t}(s_t)$)

- E.g., $p_t(s_t = 1) = 0.7$, $p_t(s_t = 2) = 0.3$

Then

$$
\begin{aligned}
E_t(y_{t+1}) &= E_t(E_t(y_{t+1}|s_t)) \\
&= p_t(s_t = 1) E_t(y_{t+1}|s_t = 1) + p_t(s_t = 2) E_t(y_{t+1}|s_t = 2)
\end{aligned}
$$

Intermediate calculation:

$$
\begin{aligned}
E_t(y_{t+1}|s_t = 1) &= \pi_{11}\mu_1 + (1 - \pi_{11})\mu_2 = e_1'\Pi\mu, \\
E_t(y_{t+1}|s_t = 2) &= (1 - \pi_{22})\mu_1 + \pi_{22}\mu_2 = e_2'\Pi\mu,
\end{aligned}
$$

where $\mu = \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}'$ and e_i is a column vector with a one in element i and zeros otherwise.

Forecasting (cont'd)

Repeating from last slide and define $\mathbf{p}_t = \begin{bmatrix} p_t(s_t = 1) & p_t(s_t = 2) \end{bmatrix}'$.

$$E_t(y_{t+1}) = p_t(s_t = 1) E_t(y_{t+1}|s_t = 1) + p_t(s_t = 2) E_t(y_{t+1}|s_t = 2)$$
$$= \mathbf{p}_t' \Pi \mu,$$

Longer-horizon forecasts:

$$E_t(y_{t+2}) = E_t(E_{t+1}(y_{t+2})) = E_t(\mathbf{p}_{t+1}' \Pi \mu)$$

Thus, we have to forecast future beliefs, $\mathbf{p}_{t+1}$:

$$E_t(\mathbf{p}_{t+1}) = \mathbf{p}_{t+1|t} = \Pi' \mathbf{p}_t$$

and so

$$E_t(y_{t+2}) = \mathbf{p}_t' \Pi^2 \mu$$

which generalizes to

$$E_t(y_{t+j}) = \mathbf{p}_t' \Pi^j \mu, \quad \text{where} \quad j = 1, 2, 3, \ldots$$

Duration of regimes

Expected duration of regime i:

$$D_i = \min \left\{ k \geq 1 : S_{t+k} \neq i | S_t = i \right\},$$

$$\Pr\left(D_i = k\right) = \pi_{ii}^{k-1}\left(1 - \pi_{ii}\right),$$

$$E\left(D_i\right) = \frac{1}{1 - \pi_{ii}}, \quad Var\left(D_i\right) = \frac{\pi_{ii}}{\left(1 - \pi_{ii}\right)^2}$$

Expected length one stays in regime i is exploding as π_{ii} approaches 1

- Thus, with $\pi_{11} \neq \pi_{22}$ one can, as in the data, have recessions (typically) last for a shorter time than expansions
- This is a form of nonlinearity. Compare to AR(1) for instance.

Model estimation

The steps to estimate an HMM share many similarities with the Kalman filter

- Smoothing and filtering probabilities
- Bayes rule
- But, this time not a linear Gaussian setup, which causes differences

We will discuss these steps in order

1. Hamilton filter for forward filtering probabilities
2. Likelihood function
3. Smoothing probabilities and the expectation (E) step
4. Parameter estimates and the maximization (M) step

Hamilton Filter: Forward Probabilities

Predictive regime probabilities:

$$p_{t|t-1}(j) = \sum_{i=1}^{2} \pi_{ij} \, p_{t-1}(i).$$

Observation density by regime:

$$f_j(y_t) = N(y_t; \mu_j, \sigma_j^2).$$

Marginal predictive density:

$$p(y_t|y_{1:t-1}) = \sum_{j=1}^{2} f_j(y_t) \, p_{t|t-1}(j).$$

Bayes update:

$$p_t(j) = \frac{f_j(y_t) \, p_{t|t-1}(j)}{\sum_{m=1}^{2} f_m(y_t) \, p_{t|t-1}(m)}.$$

Exact Log-Likelihood

The log-likelihood function is given by

$$\mathcal{L}(\Theta) = \sum_{t=1}^{T} \log\Big(\sum_{j=1}^{2} N(y_t; \mu_j, \sigma_j^2)\, p_{t|t-1}(j) \Big).$$

Notice that we now at each time t have a probability weighted average of two Normals, corresponding to one for each regime.

EM Algorithm: E-Step

First, we get the smoothing probabilities given a guess of the parameters

$$p_{t|T}(j) = \Pr(S_t = j \mid y_{1:T}), \qquad \xi_{t|T}(i,j) = \Pr(S_{t-1} = i, S_t = j \mid y_{1:T}).$$

Define the backward variable

$$\beta_t(j) = \Pr(y_{t+1}, y_{t+2}, ..., y_T \mid S_t = j.)$$

Then

$$p_{t|T}(j) \propto p_t(j)\,\beta_t(j), \qquad \xi_{t|T}(i,j) \propto p_{t-1}(i)\,\pi_{ij}\,N(y_t; \mu_j, \sigma_j^2)\,\beta_t(j).$$

Normalize per t so probabilities sum to 1.

EM Algorithm: M-Step

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T} y_t p_{t|T}(j)}{\sum_{t=1}^{T} p_{t|T}(j)},$$

$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^{T} (y_t - \hat{\mu}_j)^2 p_{t|T}(j)}{\sum_{t=1}^{T} p_{t|T}(j)},$$

$$\hat{\pi}_{ij} = \frac{\sum_{t=2}^{T} \xi_t(i,j)}{\sum_{t=2}^{T} p_{t-1|T}(i)}.$$

Notice that we take the simple mean and variance using the probability weights for each observation being in regime j.

- The transition probabilities simply count transitions, again using the relevant probability weights.

Iterated on the above 3 equations, until the parameter guesses (which enter into calculations on the right hand side of the equalities) equal the estimates on the left hadn side.

- The EM algorithm is shown to maximize the likelihood function given earlier

Hints for MLE and Identification

- Maximize $\mathcal{L}(\Theta)$ with L-BFGS-B or Newton.

- Reparametrize: $\sigma_j^2 = \exp(\lambda_j)$, $\pi_{12} = \sigma(\alpha)$, $\pi_{21} = \sigma(\beta)$, where $\sigma(\cdot)$ refers to the Sigmoid function

- Label switching: impose $\mu_1 < \mu_2$ or $\sigma_1^2 < \sigma_2^2$.
  - Otherwise, there is not a unique maximum
  - You could label state 1 as state 2 and state 2 as state 1 and the model would yield the same fit

Forecasting: Summary

$$p_{t+1|t}(j) = \sum_{i=1}^{2} \pi_{ij}\, p_t(i),$$

$$E\left(y_{t+1}|y_{1:t}\right) = \sum_{j=1}^{2} \mu_j\, p_{t+1|t}(j),$$

$$V\left(y_{t+1}|y_{1:t}\right) = \sum_{j=1}^{2} (\sigma_j^2 + \mu_j^2)\, p_{t+1|t}(j) - \left(\sum_{j=1}^{2} \mu_j\, p_{t+1|t}(j)\right)^2.$$

Note that state persistence, and likelihood of being transitioned into, are important for how the state mean μ_j is reflected in future forecast

HMMs in Finance: Real-World Examples

Let's look at some HMMs from finance

Will stick to two regimes with iid dynamics conditional on regime

1. Credit spread for high yield bonds
2. Inflation dynamics

Credit: Low vs Stressed Spread Regimes

y_t is the credit spread level. $y_t \mid S_t = j \sim N(\mu_j, \sigma_j^2)$.
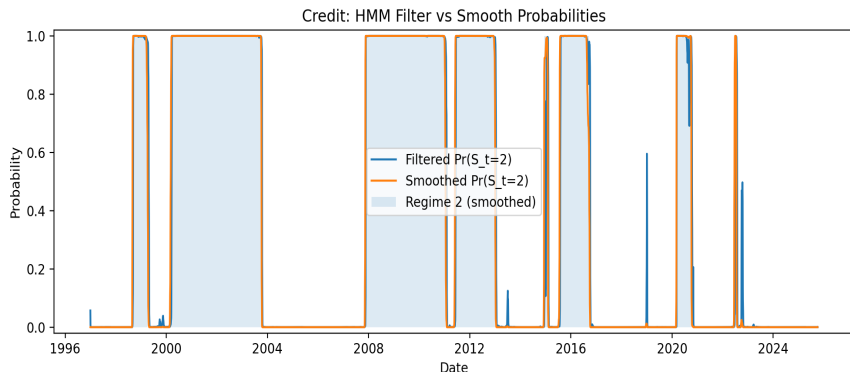
- Monthly data, estimate using ML as above

	μ_1	σ_1	μ_2	σ_2	π_{11}	π_{22}
MLE	3.81	0.71	7.57	2.72	0.998	0.997
(*s.e.*)	(0.013)	(0.009)	(0.055)	(0.036)	(0.001)	(0.001)

- Regime 1: low mean, low volatility
- Regime 2: high mean, high volatility
- Both regimes very persistent

Credit: Filtering and smoothing probabilities

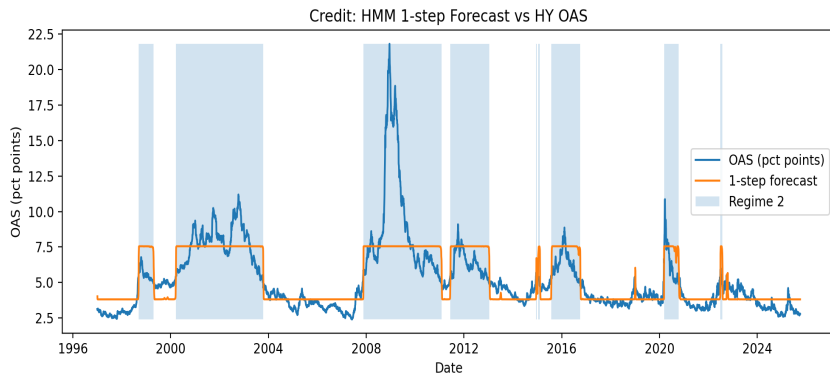y_t is the credit spread level. $y_t \mid S_t = j \sim N(\mu_j, \sigma_j^2)$.

- Regime 1: low mean and vol
- Regime 2: high mean and vol
- Shaded area if $p_{t|T}(s_t = 2) > 0.5$

Credit: HMM Filter vs Smooth Probabilities

Legend:
- Filtered Pr(S_t=2)
- Smoothed Pr(S_t=2)
- Regime 2 (smoothed)

Credit: Low vs Stressed Spread Regimes

y_t is the credit spread level. $y_t \mid S_t = j \sim N(\mu_j, \sigma_j^2)$.

- Regime 1: low mean and vol
- Regime 2: high mean and vol
- Shaded area if $p_{t|T}(s_t = 2) > 0.5$

Credit: HMM 1-step Forecast vs HY OAS

Inflation: Low vs High Volatility Regimes

- YoY inflation, but we will still use the same 2-state SR model here
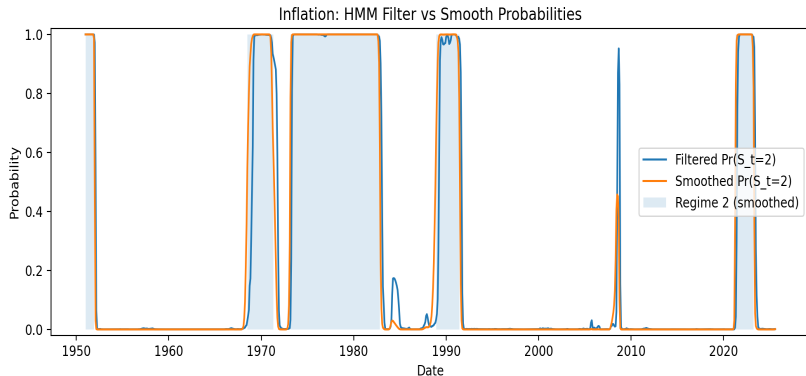- y_t is YoY monthly inflation. $y_t \mid S_t = j \sim N(\mu_j, \sigma_j^2)$.

	μ_1	σ_1	μ_2	σ_2	π_{11}	π_{22}
MLE	2.27	1.21	7.44	2.65	0.992	0.978
(s.e.)	(0.055)	(0.038)	(0.215)	(0.127)	(0.004)	(0.010)

- Regime 1: low mean, low volatility
- Regime 2: high mean, high volatility
- Both regimes very persistent, but regime 2 less so
  - Regime 1: Duration 10.6 years; Regime 2: Duration 3.9 years.

Inflation: Filtering and smoothing probabilities

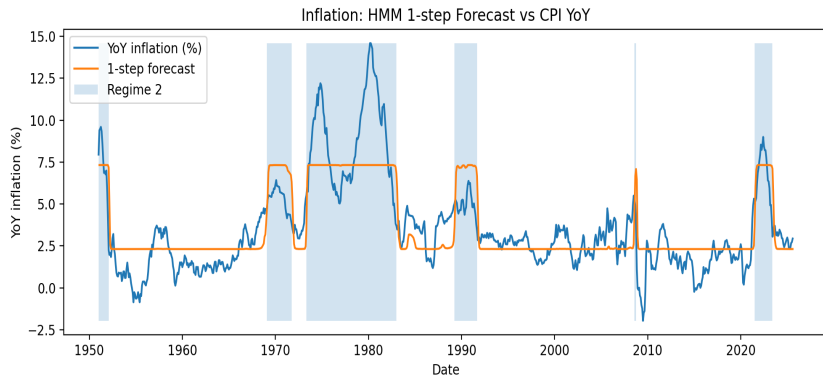y_t is inflation. $y_t \mid S_t = j \sim N(\mu_j, \sigma_j^2)$.

- Regime 1: low mean and vol
- Regime 2: high mean and vol, less persistent
- Shaded area if $p_{t|T}(s_t = 2) > 0.5$

Inflation: HMM Filter vs Smooth Probabilities

Inflation: Regimes with different persistence

y_t is inflation. $y_t \mid S_t = j \sim N(\mu_j, \sigma_j^2)$.

- Regime 1: low mean and vol
- Regime 2: high mean and vol, less persistent
- Shaded area if $p_{t|T}(s_t = 2) > 0.5$

Inflation: HMM 1-step Forecast vs CPI YoY

More complicated regime switching models

A natural extension is the K-regime, AR(p) model:

$$y_t = \phi_{0,s_t} + \phi_{1,s_t} y_{t-1} + ... + \phi_{p,s_t} y_{t-p} + \sigma_{s_t} \varepsilon_t,$$

where $\varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$ and there are K regimes, $S_t \in \{1, \ldots, K\}$.

It is possible to have time-varying regime probabilities as function of observable variables

- Using a multinomial logit and exogenous vector X_t to get regime probabilities.

Want more detail?

- Time series analysis by Hamilton
- Or, how about you ask your favorite LLM for further in-depth discussion. It has read all the textbooks....

State Space + Switching Regime Model

SR-SSMs

Canonical Linear-Gaussian SSM

Observation and state:

$$y_t = Zx_t + d + \varepsilon_t, \qquad\qquad \varepsilon_t \sim N(0, H),$$
$$x_t = Tx_{t-1} + c + R\eta_t, \qquad\qquad \eta_t \sim N(0, Q),$$

with independence over time and across shocks. Initial $x_0 \sim N(a_{0|0}, P_{0|0})$.

Kalman filter:

$$a_{t|t-1} = Ta_{t-1|t-1} + c, \qquad\qquad P_{t|t-1} = TP_{t-1|t-1}T' + RQR',$$
$$v_t = y_t - Za_{t|t-1} - d, \qquad\qquad F_t = ZP_{t|t-1}Z' + H,$$
$$K_t = P_{t|t-1}Z'F_t^{-1}, \qquad\qquad a_{t|t} = a_{t|t-1} + K_tv_t,$$
$$P_{t|t} = P_{t|t-1} - K_tF_tK_t'.$$

Log-likelihood: $\ell(\Theta) = \sum_{t=1}^{T} \log N(v_t; 0, F_t)$.

SR-SSM: Model (K States)

Regime $S_t \in \{1, \ldots, K\}$, transition matrix $\Pi = [\pi_{ij}]$.

Conditional on $S_t = j$:

$$y_t = Z_j x_t + d_j + \varepsilon_t, \qquad\qquad \varepsilon_t \sim N(0, H_j),$$
$$x_t = T_j x_{t-1} + c_j + R_j \eta_t, \qquad\qquad \eta_t \sim N(0, Q_j).$$

Initials: $S_0 \sim \pi_0$, $x_0 | S_0 = j \sim N(a_{0|0}^{(j)}, P_{0|0}^{(j)})$.

Parameters: $\Theta = \{(Z_j, d_j, H_j, T_j, c_j, R_j, Q_j)_{j=1}^{K}, \Pi\}$.

Hamilton–Kim Mixture Kalman Filter

Prediction from $i \to j$:

$$a_{t|t-1}^{(j|i)} = T_j a_{t-1|t-1}^{(i)} + c_j,$$

$$P_{t|t-1}^{(j|i)} = T_j P_{t-1|t-1}^{(i)} T_j' + R_j Q_j R_j'.$$

Mixing:

$$p_{t|t-1}(j) = \sum_i \pi_{ij} p_{t-1}(i), \qquad \omega_{ij,t} = \frac{\pi_{ij} p_{t-1}(i)}{p_{t|t-1}(j)},$$

$$a_{t|t-1}^{(j)} = \sum_i \omega_{ij,t} a_{t|t-1}^{(j|i)},$$

$$P_{t|t-1}^{(j)} = \sum_i \omega_{ij,t} \left(P_{t|t-1}^{(j|i)} + (a_{t|t-1}^{(j|i)} - a_{t|t-1}^{(j)})(a_{t|t-1}^{(j|i)} - a_{t|t-1}^{(j)})' \right).$$

Update in j:

$$v_t^{(j)} = y_t - Z_j a_{t|t-1}^{(j)} - d_j, \quad F_t^{(j)} = Z_j P_{t|t-1}^{(j)} Z_j' + H_j,$$

$$K_t^{(j)} = P_{t|t-1}^{(j)} Z_j' (F_t^{(j)})^{-1},$$

$$a_{t|t}^{(j)} = a_{t|t-1}^{(j)} + K_t^{(j)} v_t^{(j)}, \quad P_{t|t}^{(j)} = P_{t|t-1}^{(j)} - K_t^{(j)} F_t^{(j)} K_t^{(j)'}.$$

Regime and State Smoothing

Regime smoothing:

$$p_{t|T}(j) = \frac{p_t(j) \sum_{\ell=1}^{K} \frac{\pi_{j\ell} \times p_{t+1|T}(\ell)}{p_{t+1}(\ell)}}{\sum_{m=1}^{K} p_t(m) \sum_{\ell=1}^{K} \frac{\pi_{m\ell} \times p_{t+1|T}(\ell)}{p_{t+1}(\ell)}}, \quad t = T-1, \ldots, 1.$$

State smoothing (Kim collapse + RTS):

$$\omega_{j\ell,t+1}^{*} = \frac{\pi_{j\ell} \times p_{t+1|T}(\ell) / p_{t+1|t}(\ell)}{\sum_{m=1}^{K} \pi_{jm} \times p_{t+1|T}(m) / p_{t+1|t}(m)}, \quad a_{t+1|T}^{(j,*)} = \sum_{\ell} \omega_{j\ell,t}^{*} a_{t+1|T}^{(\ell)},$$

$$P_{t+1|T}^{(j,*)} = \sum_{\ell} \omega_{j\ell,t}^{*} \left[P_{t+1|T}^{(\ell)} + \left(a_{t+1|T}^{(\ell)} - a_{t+1|T}^{(j,*)} \right) \left(a_{t+1|T}^{(\ell)} - a_{t+1|T}^{(j,*)} \right)' \right]$$

$$J_t^{(j)} = P_{t|t}^{(j)} T_j' (P_{t+1|t}^{(j)})^{-1},$$

$$\hat{a}_{t|T}^{(j)} = a_{t|t}^{(j)} + J_t^{(j)} (\hat{a}_{t+1|T}^{(j,*)} - a_{t+1|t}^{(j)}),$$

$$\hat{P}_{t|T}^{(j)} = P_{t|t}^{(j)} + J_t^{(j)} (\hat{P}_{t+1|T}^{(j,*)} - P_{t+1|t}^{(j)}) J_t^{(j)'}.$$

Collapse:

$$\hat{a}_{t|T} = \sum_j p_{t|T}(j) \hat{a}_{t|T}^{(j)}, \quad \hat{P}_{t|T} = \sum_j p_{t|T}(j) (\hat{P}_{t|T}^{(j)} + (\hat{a}_{t|T}^{(j)} - \hat{a}_{t|T})(\hat{a}_{t|T}^{(j)} - \hat{a}_{t|T})').$$

2-State Inflation Model

Scalar x_t, inflation y_t, regime $S_t \in \{1, 2\}$:

$$y_t = d_j + x_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, h_j),$$
$$x_t = c_j + \phi_j x_{t-1} + \eta_t, \quad \eta_t \sim N(0, q_j),$$

with $j = S_t$. $\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix}$.

Filter scalars:

$$a_{t|t-1}^{(j|i)} = c_j + \phi_j a_{t-1|t-1}^{(i)}, \quad P_{t|t-1}^{(j|i)} = \phi_j^2 P_{t-1|t-1}^{(i)} + q_j,$$

$$v_t^{(j)} = y_t - d_j - a_{t|t-1}^{(j)}, \quad F_t^{(j)} = P_{t|t-1}^{(j)} + h_j, \quad K_t^{(j)} = \frac{P_{t|t-1}^{(j)}}{F_t^{(j)}}.$$

Identification and Initialization

Only $\mu_j = d_j + \frac{c_j}{1-\phi_j}$ is identified without restrictions.

- Fix $d_1 = 0$ or impose $d_1 = d_2$.
- Constrain $|\phi_j| < 1$ via $\phi_j = \tanh(\varphi_j)$.
- Often set $h_1 = h_2 = h$.

Initialization options:

- Diffuse: large $P_{0|0}^{(j)}$, $a_{0|0}^{(j)} = 0$, p_0 stationary in Π.
- Regime-stationary: $a_{0|0}^{(j)} = \frac{c_j}{1-\phi_j}$, $P_{0|0}^{(j)} = \frac{q_j}{1-\phi_j^2}$.

Exact Likelihood and Direct MLE

Predictive mixture: $p(y_t|y_{1:t-1}) = \sum_{j=1}^{2} N\big(v_t^{(j)}; 0, F_t^{(j)}\big)\, p_{t|t-1}(j)$.

Log-likelihood: $\mathcal{L}(\Theta) = \sum_{t=1}^{T} \log\Big(\sum_{j=1}^{2} f_j\,(y_t)\, p_{t|t-1}(j)\Big)$.

Use L-BFGS-B or Newton with reparametrizations:
$\phi_j = \tanh\varphi_j$, $h_j = e^{\lambda_j}$, $q_j = e^{\kappa_j}$, $\pi_{12} = \sigma(\alpha)$, $\pi_{21} = \sigma(\beta)$.

Cost penalty for number of regimes $O(TK^2) = O(4T)$.

EM Algorithm

E-step: compute $p_{t|T}(j) = \Pr(S_t = j | y_{1:T})$, $\xi_t(i,j) = \Pr(S_{t-1} = i, S_t = j | Y)$, and smoothed moments $\overline{x_t}^{(j)}, \overline{x_t^2}^{(j)}, \overline{x_t x_{t-1}}^{(j)}$.

M-step:

$$\hat{\pi}_{ij} = \frac{\sum_t \xi_t(i,j)}{\sum_t p_{t-1|T}(i)}, \quad \hat{d}_j = \frac{\sum_t p_{t|T}(j)(y_t - \overline{x_t}^{(j)})}{\sum_t p_{t|T}(j)},$$

$$\hat{h}_j = \frac{\sum_t p_{t|T}(j)\left[(y_t - \hat{d}_j)^2 - 2(y_t - \hat{d}_j)\overline{x_t}^{(j)} + \overline{x_t^2}^{(j)}\right]}{\sum_t p_{t|T}(j)},$$

$$\begin{bmatrix} \hat{c}_j \\ \hat{\phi}_j \end{bmatrix} = \left(\sum_t p_{t|T}(j)\begin{bmatrix} 1 \\ \overline{x_{t-1}}^{(j)} \end{bmatrix}\begin{bmatrix} 1 & \overline{x_{t-1}}^{(j)} \end{bmatrix}\right)^{-1}\left(\sum_t p_{t|T}(j)\begin{bmatrix} \overline{x_t}^{(j)} \\ \overline{x_t x_{t-1}}^{(j)} \end{bmatrix}\right),$$

$$\hat{q}_j = \frac{1}{\sum_t p_{t|T}(j)}\sum_t p_{t|T}(j)\, E\left[(x_t - \hat{c}_j - \hat{\phi}_j x_{t-1})^2 \mid y_{1:T}, S_t = j\right].$$

Inflation: 2-state SR-SSM estimation results

	d_1	d_2	c_1	c_2	q_1	q_2
MLE	0.54	0.41	0.06	0.02	0.39	0.07
(*s.e.*)	(297)	(297)	(6.99)	(0.193)	(0.044)	(0.005)

	ϕ_1	ϕ_2	h	π_{11}	π_{22}
MLE	0.976	0.999	1.25	0.957	0.984
(*s.e.*)	(0.012)	(0.001)	(0.001)	(0.021)	(0.007)

- Regime 2: low mean, low volatility
- Regime 1: high mean, high volatility
- Both regimes very persistent, but regime 1 less so
  - Regime 1: Duration 1.9 years; Regime 2: Duration 5.4 years.

Does anything look off to you in the estimates?

Important: you cannot blindly trust LLM output

Despite having read papers on the SR-SSM, the estimation has issues

1. The observation equation means d_1 and d_2 have enormous standard errors
2. So does c_1 and to some extent c_2
3. Persistence ϕ_2 statistically unit root
4. h just one number, not separate volatility for each state in observation equation
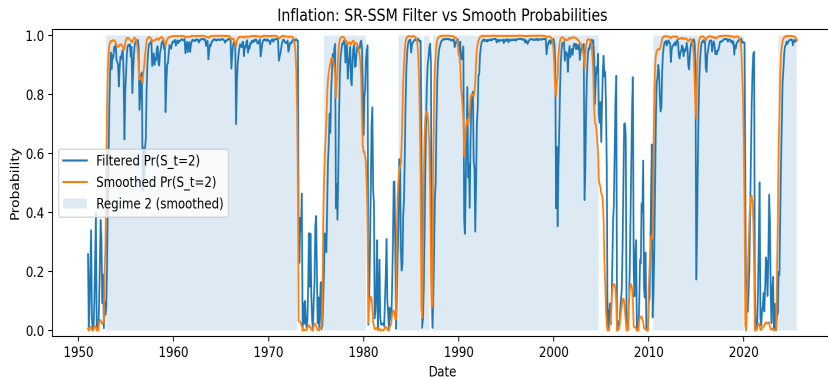
What's going on? Model is overidentified and has close to nonstationary elements.

- Need to choose two parameters of c_1, c_2, d_1, d_2
    - E.g., constrain $c_1 = c_2 = 0$.
- Perhaps constrain $\left| \phi_j \right| < 1 - 2 \times s.e. \left(\phi_j \right)$ for stationarity?
- Perhaps help with starting values for h_1, h_2 set equal to variance of observable, or similar?
- Ask LLM for suggestion for help with better identification of variables in model?

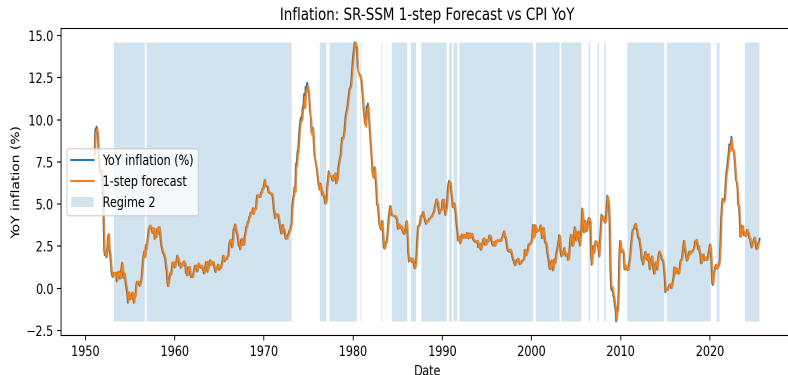Modeling and estimation is an iterative procedure

Inflation: Filtering and smoothing probabilities

- Regime 1: low mean and vol
- Regime 2: high mean and vol, less persistent
- Shaded area if $p_{t|T}(s_t = 2) > 0.5$

Inflation: SR-SSM Filter vs Smooth Probabilities

Inflation: Regimes with different persistence

- Regime 2: low mean and vol
- Regime 1: high mean and vol, less persistent
- Notice much better forecast of inflation than simple SR earlier

Inflation: SR-SSM 1-step Forecast vs CPI YoY

Another Nonlinear Version of the SSM

The Unscented Kalman Filter

UKF: Nonlinear SSM

UKF: Unscented Kalman FIlter

$$y_t = h(x_t) + \varepsilon_t, \quad \varepsilon_t \sim N(0, R),$$
$$x_t = f(x_{t-1}) + \eta_t, \quad \eta_t \sim N(0, Q).$$

Here, $h(\cdot)$ and $f(\cdot)$ are known (you choose) nonlinear functions.

UKF propagates moments via "sigma points" instead of linearization.

UKF: Sigma Points

Given $x \sim N(m, P)$, construct $\{\chi^{(i)}\}_{i=0}^{2n}$ with weights $\{W^{(i)}\}$:

$$\chi^{(0)} = m, \quad \chi^{(i)} = m + \left[\sqrt{(n+\lambda)P}\right]_i, \; i = 1, \ldots, n,$$

$$\chi^{(i)} = m - \left[\sqrt{(n+\lambda)P}\right]_{i-n}, \; i = n+1, \ldots, 2n,$$

$W_m^{(0)} = \frac{\lambda}{n+\lambda}$, $W_c^{(0)} = \frac{\lambda}{n+\lambda} + 1 - \alpha^2 + \beta$, $W_m^{(i)} = W_c^{(i)} = \frac{1}{2(n+\lambda)}$,
$\lambda = \alpha^2(n+\kappa) - n$.

Main idea: use Kalman filter for each sigma point, then average across sigma points using weights W

UKF: Updates

Time update:

$$\chi_{t|t-1}^{(i)} = f(\chi_{t-1|t-1}^{(i)}), \quad m_{t|t-1} = \sum_i W_m^{(i)} \chi_{t|t-1}^{(i)},$$

$$P_{t|t-1} = \sum_i W_c^{(i)} (\chi_{t|t-1}^{(i)} - m_{t|t-1})(\chi_{t|t-1}^{(i)} - m_{t|t-1})' + Q.$$

Measurement update:

$$\zeta_t^{(i)} = h(\chi_{t|t-1}^{(i)}), \quad \hat{y}_t = \sum_i W_m^{(i)} \zeta_t^{(i)},$$

$$S_t = \sum_i W_c^{(i)} (\zeta_t^{(i)} - \hat{y}_t)(\zeta_t^{(i)} - \hat{y}_t)' + R,$$

$$C_t = \sum_i W_c^{(i)} (\chi_{t|t-1}^{(i)} - m_{t|t-1})(\zeta_t^{(i)} - \hat{y}_t)',$$

$$K_t = C_t S_t^{-1}, \quad m_{t|t} = m_{t|t-1} + K_t(y_t - \hat{y}_t),$$

$$P_{t|t} = P_{t|t-1} - K_t S_t K_t'.$$

Conclusion

Combining the SSM with non-linear features provides very flexible non-linear modeling tools

Maximum likelihood estimation is standard and has a large well-developed literature that you can draw on

- For instance, you can add a penalty for number of parameters and ridge/lasso type penalties like we did for the logistic regression in Data Analytics in Spring
- In this case, you would need some type of cross-validation to find the optimal constraint for best out-of-sample performance

These SSM-based models are typically relative intuitive and help us understand the underlying mechanisms driving the economy/the process at hand

The downside is that they (largely) assume you know the functional form (e.g., linear + regimes).

- Machine learning methods like neural nets and deep learning are more flexible and aimed at allowing for very flexible nonlinear patterns