On Bayesian Filtering for Markov Regime Switching Models

Disclaimer: The views expressed do not necessarily reflect those of Norges Bank

Nigar Hashimzade[†] *Oleg Kirsanov*[*] *Tatiana Kirsanova*[*] *Junior Maih*[‡]

[†]Brunel University, [*]University of Glasgow, [‡]Norges Bank

Summer 2024

What the paper is about

This paper presents a framework for empirical analysis of dynamic macroeconomic models using Bayesian filtering, with a specific focus on New Keynesian DSGE models with regime switching in the state space form.

Motivation

- **Dynamic Economic Realities:** Economic conditions shift between crises and stability, conventional and unconventional policies, necessitating models that capture such dynamics.
- **Popularity of Regime-Switching Models:** The proliferation of regime-switching models addresses nonlinearities in economic processes, especially crucial for discerning 'good luck' from 'good policy.' However, their estimation is intricate due to two fundamental challenges:
  - Dimensionality: with regime switches the number of possible histories grows exponentially
  - Approximation: Non-standard switching filters and smoothers with some approximations are required even for linear Gaussian models.
- **Unknown Economic Applicability:** Despite various filters being widely used in engineering, their properties in economic applications remain unexplored.

Complexity of filtering under regime-switching I

A constant-parameter state-space model

$$\alpha_t = T(\alpha_{t-1}, \eta_t)$$
$$y_t = Z(\alpha_t, \varepsilon_t)$$

- A simple linear **filter** [e.g. Kalman Filter, Kalman (1975)] can be described:
  - Use the estimated model and the data up to period $t-1$ to predict the observation (and its variance) in period t;
  - Use the difference between the actual observation in period t and its prediction to adjust the estimated model parameters (using Bayesian updating);
- Continue to the latest observation (and add the new ones as they arrive).
- A **smoother** is applied 'backwards', thus adding the information about the subsequent ('future') observation to the information about the previous ('past') observation that was used in the filter.

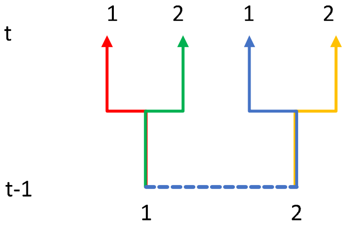
Complexity of filtering under regime-switching II

For $s_t = 1, 2, ..., h$ a switching state-space model is given by

$$\begin{aligned} \alpha_t &= T_{s_t}(\alpha_{t-1}, \eta_t) \\ y_t &= Z_{s_t}(\alpha_t, \varepsilon_t) \end{aligned}$$

In this case, exact filtration becomes intractable.
Consider a simple example with $h = 2$ (1="good policy', 2='bad policy'):

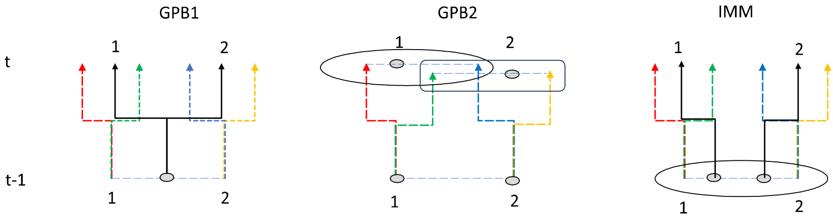
- We don't know whether today is state 1 or 2;
- Whichever state is today, the world can go into either state 1 or 2 tomorrow;
- In each next period the number of histories doubles;



Complexity of filtering under regime-switching II

- In general, if there are h states of the world, the number of histories is multiplied by h in every period.
- Markov property reduces the set we condition on, but not the number of histories. Pr(state | history of states) vs. Pr(history of states | observations)
- We need some approximation to avoid the exponential growth of the paths.

Complexity of filtering under regime-switching III



The Literature

- **Theoretical Advances in Economics:** *Kim and Nelson* (KN) *filter and smoother* (Kim, 1994, Kim and Nelson, 1999)
- **Common Practice in Economics:**
  - Many economic studies employ the Kim and Nelson (KN) filter (Kim, 1994, Kim and Nelson, 1999) without explicit discussion or justification. See e.g Davig and Doh (2014), Chang, Maih and Tan 2021, Chen, Leeper and Leith (2022)
  - The KN smoother (smoothed *variables*, not probabilities) is computationally unstable: No evidence of using Markov-switching smoother to uncover latent variables

The Literature

- **Engineering Literature Connection:** The KN filter, *identified as GPB2*, belongs to the Generalised Pseudo-Bayesian (GPB) filter class, widely known in engineering.
- **Engineering Perspectives:** In engineering literature, the Interactive Multiple Model (IMM) filters, initiated by Blom and Bar-Shalom (1988), became predominant from the 1990s onwards.
- **Engineering Objectives:** no interest in maximising likelihood, presence of observation errors, no structural shocks.
- **Engineering Consensus:** Engineering literature suggests:
  - GPB1 is suboptimal
  - GPB2 and IMM exhibit similar accuracy.
  - GPB2 is computationally more intensive than IMM.
- **Engineering Smoothers:** rely on invertibility of observation errors covariance matrix. Not applicable in Economics

The Literature

- **Recent Developments in Economics:** some evidence of using the IMM or its variants:
  - Liu, Wang and Zha (2013) applied a variant of IMM,
  - Binning and Maih (2015) introduced Regime-switching Sigma-point filters based on IMM, and some other RISE-based work of Maih
  - Leith, Kirsanova, Machado, and Ribeiro (2024) is an application.

Our Contributions and Results

- Generalization IMM and GPB –> IMM(N) and GPB(N).
- Derivation of corresponding multiple-regime smoothers: 25% improvement over updating.
- Implementation of the algorithms in the RISE toolbox
- Validation through rigorous simulation exercises using a prototypical New Keynesian DSGE model:
  - IMM is superior in terms of speed and as accurate as KN-GPB(2)
  - smoothing is crucial 25% improvement over updating in RMSE
  - long sample: RMSEs of *persistent* smoothed states reduce after initial 50-100 observations
  - Misspecification of the model is less of a problem if smoothing is applied
- Application to U.S. macroeconomic time series, identifying significant policy shifts and crises including those related to the post-Covid-19 era.

The modified Fernandez-Villaverde et.al. (2015) model: brief description

Standard NK DSGE model with households, monopolistically competitive firms, and monetary policymaker with a Taylor rule

- Households derive utility from consumption, real money balances and disutility from labor. They own capital that they rent to firms and own shares in the firms. They face preference shocks, shocks to their disutility of labor, and investment-specific technology shocks.
- Firms: produce goods using capital and labor and face technology shocks. They set prices for their goods under Calvo pricing frictions.
- The central bank sets the interest rate following a Taylor rule

Switching mechanisms

Two Markov switching processes:

- hawkish vs dovish policymaker (good vs. bad policy): $s_{1,t}$

$$\frac{r_t}{r_{ss}} = \left(\frac{r_{t-1}}{r_{ss}}\right)^{\gamma_r(s_{1,t})} \left[\left(\frac{\pi_t}{\pi_{targ}}\right)^{\gamma_{\pi(s_{1,t})}} \left(\frac{Y_{d,t}}{\lambda_{yd} Y_{d,t-1}}\right)^{\gamma_{y(s_{1,t})}}\right]^{1-\gamma_r(s_{1,t})} \sigma_\xi \xi_t$$

- large vs small shocks (bad vs. good luck): $s_{2,t}$
  - $\sigma_a(s_{2,t})$: Neutral technology shock
  - $\sigma_\mu(s_{2,t})$: Investment-specific technology shock
  - $\sigma_\varphi(s_{2,t})$: Labor supply shock
  - $\sigma_d(s_{2,t})$: Preference shock
  - $\sigma_\xi(s_{2,t})$: Monetary policy shock

Observed variables

The variables observed are:

- Output growth
- Price Inflation
- Wage Inflation
- Feds Funds rate
- Relative price of investment goods

Simulation design

- Given the calibration of the parameters, we simulate 500 samples of 1000 observations of artificial data.
- We run filtering and smoothing algorithms, visualize results and compute characteristics of filter performance.
- We use the simulation results to investigate the performance of the discussed filters, controlling for the sample length. e.g. 300, 500, 1000
- We use the RMSE as the evaluation criterion for accuracy

$$\mathcal{R}_{\varphi} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\frac{x_t - x_{\varphi}}{x_{ss}} \right)^2},$$

  where updated variables indexed $\varphi = t \mid t$ in case of a filter, and smoothed variables indexed $\varphi = t \mid T$ in case of a smoother.
- Computer: Ryzen 3950X with 64GB RAM using MATLAB, R2022b.

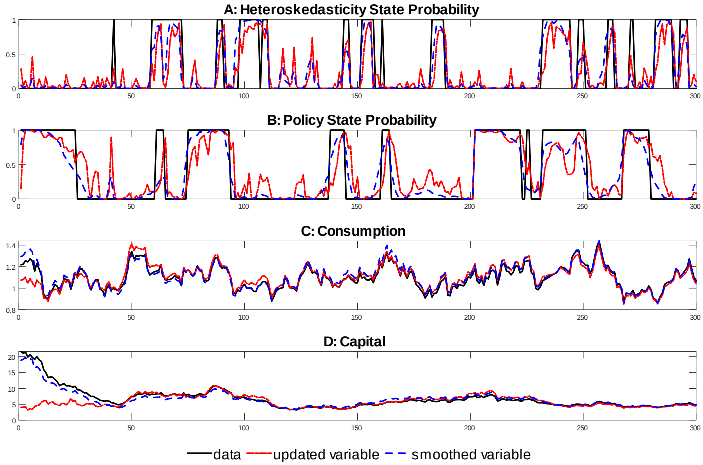
The Best Filter

vars	Relative RMSEs			
	IMM	GPB1	GPB2	GPB3
consum.	1.0003	1.030	1	1.0002
capital	1.0001	1.051	1	1.0006
output	1.0005	1.016	1.00004	1
real wage	1.0002	1.035	1	1.0003
Tobin's Q	1.0001	1.025	1	1.0001
invest.	1	1.065	1.00004	1.0010
lab sup.	1.0005	1.016	1.00004	1
pref shock	1.0001	1.015	1	1.00006
lab sup sh.	1.0005	1.019	1.0001	1
tech shock	1.0004	1.066	1	1.0015
shock prob	1.00002	1.002	1	1.0004
policy prob	1	1.023	1.00001	1.00003

Computational times for filtering 1000 observations

A: IMM(1) vs. GPB(2)-KN			B: Relative speed			
	updating only sec	updating and smoothing sec		ratio	updating only	ratio
			GPB(1)	0.28		
IMM(1)	0.27	1.49	GPB(2)	1	IMM(1)	1
GPB(2)	1.38	2.59	GPB(3)	4.21	IMM(2)	5.81
			GPB(4)	17.74	IMM(3)	52.70
			GPB(5)	79.97	IMM(4)	691.14

Updating and smoothing produced by IMM filter



Accuracy improvement by smoothing

$$1 - R_{t|T} / R_{t|t}$$

vars:	IMM	GPB(2)	GPB(1)	GPB(3)
consumption	0.29	0.27	0.31	0.27
capital	0.21	0.21	0.23	0.21
output	0.42	0.37	0.47	0.38
real wage	0.18	0.18	0.18	0.18
Tobin's Q	0.25	0.25	0.23	0.25
investment	0.30	0.28	0.32	0.28
labour supply	0.42	0.37	0.47	0.38
preference shock	0.14	0.13	0.14	0.13
labour supply shock	0.37	0.33	0.40	0.34
technology shock	0.10	0.10	0.04	0.10
shock state probs	0.15	0.15	0.15	0.15
policy state probs	0.18	0.18	0.16	0.18

RMSE accuracy: Information Matters I

- How important it is to work with a long sample?

vars:	$\mathcal{R}_{t\|t}$	$\mathcal{R}_{t\|300}$	$\mathcal{R}_{t\|1000}$
consumption	0.051	0.036	0.033
capital	0.352	0.277	0.266
output	0.050	0.029	0.026
investment	0.466	0.317	0.300
labour supply	0.049	0.028	0.026
preference shock	0.051	0.042	0.041
labour supply shock	0.110	0.066	0.061
technology shock	0.002	0.001	0.001
shock regime probs	0.265	0.225	0.225
policy regime probs	0.335	0.275	0.275

- When smoothing starts at t=1000, RMSE's for economic variables reduce by about 10%
- Smoothing of probabilities is not very affected.

RMSE accuracy: Information Matters II

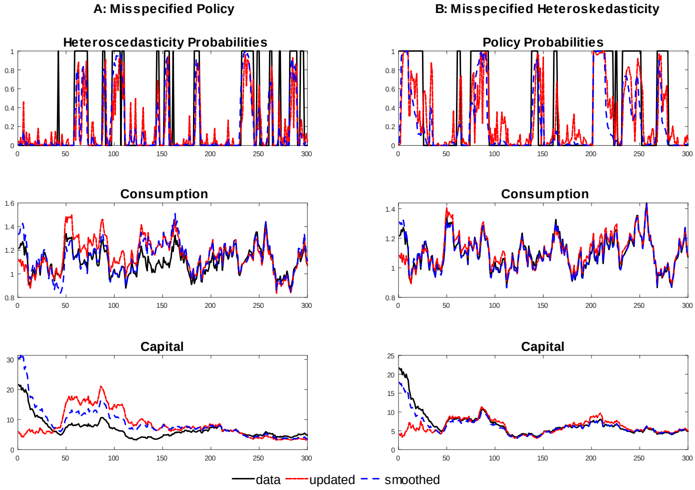
- How long should the sample be?

	$1-\frac{\mathcal{R}_{t\|300}}{\mathcal{R}_{t\|t}}$	$1-\frac{\mathcal{R}_{t\|250}}{\mathcal{R}_{t\|t}}$	$1-\frac{\mathcal{R}_{t\|200}}{\mathcal{R}_{t\|t}}$	$1-\frac{\mathcal{R}_{t\|100}}{\mathcal{R}_{t\|t}}$
consumption	0.30	0.30	0.30	0.26
capital	0.21	0.21	0.20	0.14
output	0.43	0.43	0.42	0.33
investment	0.32	0.32	0.32	0.27
labour supply	0.43	0.42	0.42	0.32
pref. shock	0.18	0.19	0.20	0.19
lab. supp. shock	0.40	0.40	0.40	0.30
techn. shock	0.10	0.10	0.11	0.11
shock state probs	0.15	0.15	0.15	0.15
policy state probs	0.18	0.18	0.18	0.19

- The effectiveness of smoother falls if the sample size is 100 or shorter.
- No difference in performance between 200-300 observations.

Model Misspecification I

- Consider two cases
  - Researcher assumes there is one policy regime
  - Researcher assumes there is one shock regime



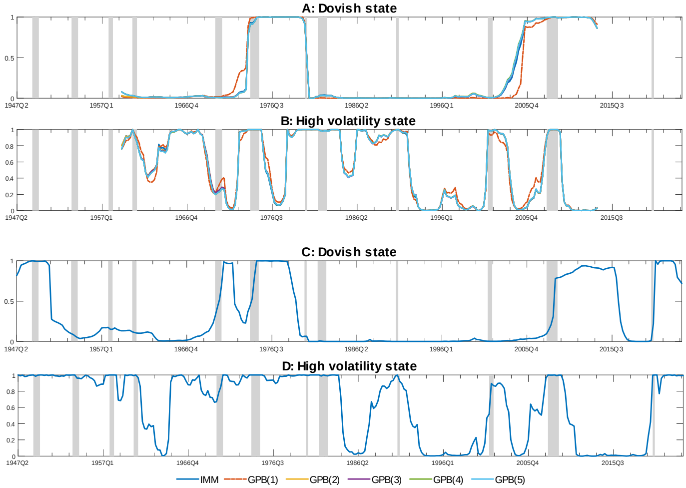
Interim Summary

- IMM outperforms the other filters in speed and its accuracy is comparable to that of KN-GPB2 (0.005% difference), which in turn is more accurate than GPB1 (2-6% difference).
- A long sample is very beneficial for empirical investigation
  - Smoothing reduces RMSE well for sample size above 100.
  - Longer data series allow to obtain better smoothed RMSEs for initial observations in the sample.
- Smoothing of regime probabilities is important – it reduces RMSEs by about 20% at each sample.
- Smoothing of state variables is particularly important – using 1000 observations sample, the RMSEs of the first 300 observations are reduced by about 20-30%.
- Some MRSEs due to model misspecifications are reduced if smoothing is applied.

Estimation results: Fernandez-Villaverde et al. (2015) dataset

- Estimate the model with IMM and using original data 1959Q2-2013Q4
- Problem potentially multimodal with observationally-equivalent configurations of parameters, especially because we use wide priors.
- To help identification we normalize the states under estimation partially through the priors but also through a direct choice of the Hawkish state and the High-volatility state
- Matlab's fmincon is not up to the task. To estimate the mode of the posterior distribution, we use the Artificial Bee Colony algorithm Karaboga and Basturk (2007), as implemented in RISE, which has very powerful exploration capabilities.

Smoothed state probabilities

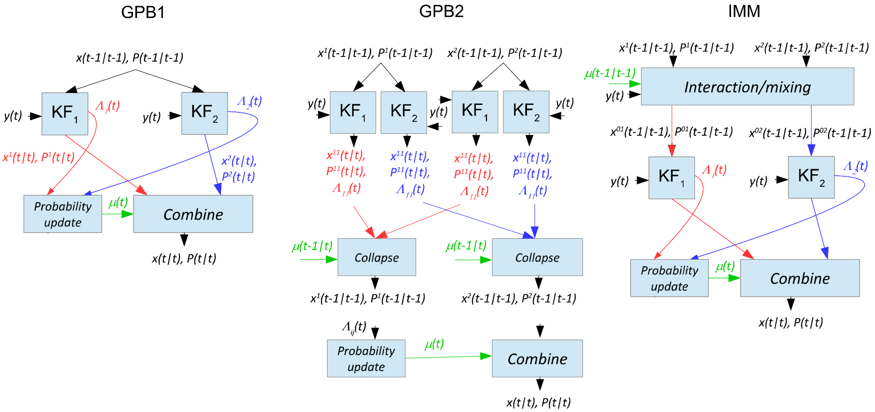


- Great Inflation, Volker disinflation. Great financial crisis, Recessions
- New data: Early start - better identification of the Great Inflation, ZLB period, Post-Covid period

Conclusions

- **Optimal Filter:** suggests that IMM trounces GPBN in terms of speed and competes closely with GPB2 in terms of accuracy.
- **Methodological Advancements:** Introduces IMMN and GPBN algorithms and a versatile smoothing procedure to enhance state-space model filtering.
- **Empirical Significance:** unveils doveish policy periods and heightened volatility, notably during the COVID crisis.
- **Overall Impact:** significantly refines regime-switching model estimation, offering valuable insights for economic analyses.

Approximate filtering procedures



A Smoothing algorithm for the state vector I

1. Initialise the smoother by setting $r^i_{T|T-1} = F'_{i,T}\left(H_{i,T}\right)^{-1}\eta^i_{T|T-1}$, $i = 1, \dots, M$ and $x^i_{T|T}$ as the $t = T$ output of the corresponding filter
2. Compute the smoothed estimates of the state vector for each regime, $i = 1 \dots M$, using recursion:

$$
\begin{aligned}
L^{ij}_{t+1,t} &= A_{j,t+1}\left(I - K_{i,t}F_{i,t}\right) \\
r^i_{t|t-1} &= F^{i\prime}_t\left(H^i_t\right)^{-1}\eta^i_{t|t-1} + \sum_{j=1}^{M} p^{ij} L^{ij\prime}_{t+1,t} r^j_{t+1|t} \\
x^i_{t|T} &= x^i_{t|t-1} + P^i_{t|t-1} r^i_{t|t-1},
\end{aligned}
$$

   for $t = T-1, T-2, \dots, 1$.
3. Use the smoothed probabilities, $\mu^i_{t|T}$, to ‘fuse’ the smoothed state vectors:

$$x_{t|T} = \sum_{i=1}^{M} \mu^i_{t|T} x^i_{t|T}.$$

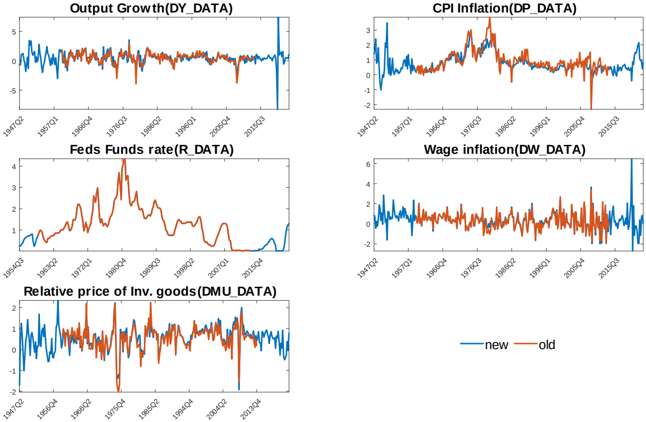
RMSE accuracy: Sensitivity to differences in states and transitionsl

policy description	γ_π	$q_{HD} = q_{DH}$	upd-d	sm-d
more dist. regimes, less dist. feedback	1.5	0.05	0.366	0.310
less dist. regimes, less dist. feedback	1.5	0.1	0.412	0.380
more dist. regimes, more dist. feedback	1.7	0.05	0.335	0.275
less dist. regimes, more dist. feedback	1.7	0.1	0.386	0.349

- Nomenclature
  - More distinct feedback : $\gamma_\pi = (1.7, 0.9)$
  - Less distinct feedback : $\gamma_\pi = (1.5, 0.9)$
  - More distinct regime: prob to leave doveish/hawkish regime $= 0.05$
  - Less distinct regime: prob to leave doveish/hawkish regime $= 0.1$
- The greater the difference between the simulated regimes, the better identification is

Updating the dataset (no re-estimation) I

- Update the data to 1947Q2-2023Q3



- Some revisions and new events - Covid period

Filtering Problem I

State-space representation of a dynamic linear model with regime switches in both the measurement and the transition equations:

$$
\begin{aligned}
y_t &= F_{S_t} x_t + \beta_{S_t} z_t + e_t \\
x_t &= A_{S_t} x_{t-1} + \gamma_{S_t} z_t + G_{S_t} v_t
\end{aligned}
$$

where

$$
\begin{pmatrix} e_t \\ v_t \end{pmatrix} \sim \mathrm{N}\left(0, \begin{pmatrix} R & 0 \\ 0 & Q \end{pmatrix}\right).
$$

A first-order Markov process with M regimes is governed by the following transition matrix:

$$
p = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{pmatrix}
$$

Filtering Problem II

- Find

$$
\begin{aligned}
x_{t|t-1} \quad &: \quad = \mathbb{E}\left[x_t \mid \psi_{t-1}\right], \\
P_{t|t-1} \quad &: \quad = \mathbb{E}\left[\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' \mid \psi_{t-1}\right] \\
& \quad\quad f\left(y_t \mid \psi_{t-1}\right)
\end{aligned}
$$

for $\psi_{t-1} \equiv \left(y'_{t-1}, y'_{t-2}, \ldots, y'_1, z'_t, z'_{t-1}, \ldots, z_1\right)'$ the vector of observations

- Standard Kalman filter

$$
\begin{aligned}
x_{t|t-1} &= Ax_{t-1|t-1} + \gamma z_t, & P_{t|t-1} &= AP_{t-1|t-1}A' + GQG', \\
\eta_{t|t-1} &= y_t - Fx_{t|t-1} - \beta z_t, & x_{t|t} &= x_{t|t-1} + K_t \eta_{t|t-1}, \\
H_t &:= FP_{t|t-1}F' + R, & P_{t|t} &= \left(I - P_{t|t-1}F'H_t^{-1}F\right) P_{t|t-1}
\end{aligned}
$$

needs to be adapted for multiple regimes.

The IMM algorithm I

1. Start with $x^i_{t-1|t-1}, P^i_{t-1|t-1}, \mu^i_{t-1|t-1} := \Pr\left[S_{t-1} = i \mid \psi_{t-1}\right]$
2. Compute the *mixing probabilities* defined as

$$\mu^{i|j}_{t-1|t-1} := \Pr\left[S_{t-1} = i \mid \psi_{t-1}, S_t = j\right] = \frac{p^{ij}\mu^i_{t-1|t-1}}{\sum_{k=1}^{M} p^{kj}\mu^k_{t-1|t-1}}$$

3. Compute the *mixed* state vectors and MSE matrices for each regime:

$$\begin{aligned}
\mathbf{x}^{0j}_{t-1|t-1} &= \sum_{i=1}^{M} \mu^{i|j}_{t-1|t-1} x^i_{t-1|t-1}, \\
P^{0j}_{t-1|t-1} &= \sum_{i=1}^{M} \mu^{i|j}_{t-1|t-1} \left\{ \begin{array}{c} P^i_{t-1|t-1} + \left(x^i_{t-1|t-1} - x^{0i}_{t-1|t-1}\right) \\ \times \left(x^i_{t-1|t-1} - x^{0i}_{t-1|t-1}\right)' \end{array} \right\}.
\end{aligned}$$

The IMM algorithm I

4. For each regime j compute the standard KF to obtain

$$\left(x^j_{t|t-1}, P^j_{t|t-1}, \eta^j_{t|t-1}, x^j_{t|t}, P^j_{t|t}\right) = \mathcal{K}\left(\begin{array}{c} F_{j,t}, \beta_{j,t}, A_{j,t}, \gamma_{j,t}, G_{j,t}, Q, R; \\ x^{0j}_{t-1|t-1}, P^{0j}_{t-1|t-1}; y_t, z_t \end{array}\right)$$

   with the associated likelihood

$$\begin{aligned} \Lambda^j_t &= f\left(y_t \mid S_t = j, \psi_{t-1}\right) \\ &= (2\pi)^{-N/2} \left|H^j_t\right|^{-1/2} \exp\left(-\frac{1}{2}\eta^{j\prime}_{t|t-1} H^{j^{-1}}_t \eta^j_{t|t-1}\right) \end{aligned}$$

5. Update the probabilities:

$$\mu^j_{t|t} == \frac{\Lambda^j_t \sum_{i=1}^M p^{ij} \mu^i_{t-1|t-1}}{\sum_{j=1}^M \Lambda^j_t \sum_{i=1}^M p^{ij} \mu^i_{t-1|t-1}}.$$

The t-increment likelihood (maximised in the process of estimation) is

$$L_t = \log f\left(y_t \mid \psi_{t-1}\right)$$

Smoothed probabilities/ Hamilton filter I

The smoothing algorithm is implemented by backward recursion as the following.

1. Initialise $\mu^k_{T|T} = \Pr\left[S_T = k \mid \psi_T\right], \; k = 1, ..., M.$
2. Combine (**??**) with (**??**) and use filtered $\mu^j_{t|t}$ to make the step back:

$$\mu^j_{t|T} = \sum_{k=1}^{M} \mu^k_{t+1|T} \frac{\mu^j_{t|t} p^{jk}}{\sum_{m=1}^{M} p^{mk} \mu^m_{t|t}}, \;\; t = T-1, T-2, ..., 1.$$

Smoothed states

1. Initialise the smoother by setting $r^i_{T|T-1} = F'_{i,T}\left(H_{i,T}\right)^{-1}\eta^i_{T|T-1}$, $i = 1, \ldots, M$ and $x^i_{T|T}$ as the $t = T$ output of the corresponding filter
2. Compute the smoothed estimates of the state vector for each regime, $i = 1 \ldots M$, using recursion:

$$\begin{aligned} L^{ij}_{t+1,t} &= A_{j,t+1}\left(I - K_{i,t}F_{i,t}\right) \\ r^i_{t|t-1} &= F^{i\prime}_t\left(H^i_t\right)^{-1}\eta^i_{t|t-1} + \sum_{j=1}^{M} p^{ij} L^{ij\prime}_{t+1,t} r^j_{t+1|t} \\ x^i_{t|T} &= x^i_{t|t-1} + P^i_{t|t-1} r^i_{t|t-1}, \end{aligned}$$

for $t = T-1, T-2, ..., 1$.

3. Use the smoothed probabilities, $\mu^i_{t|T}$, to 'fuse' the smoothed state vectors:

$$x_{t|T} = \sum_{i=1}^{M} \mu^i_{t|T} x^i_{t|T}.$$