# IV. Markov-switching models

A. Introduction to Markov-switching models

- Many economic series exhibit dramatic breaks:
  - recessions
  - financial panics
  - currency crises
- Questions to be addressed:
  - how handle econometrically
  - how incorporate into economic theory

# Economic recessions as changes in regime

$y_t$ = real GDP growth in quarter $t$

$s_t = 1$ when economy is in expansion

$s_t = 2$ when economy is in recesion

$y_t = m_{s_t} + \varepsilon_t$

$\varepsilon_t \sim N(0, \sigma^2)$

$\text{Prob}(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots, y_{t-1}, y_{t-2}, \ldots)$
$= p_{ij}$

$y_t = m_{s_t} + \varepsilon_t$

If $s_t$ is observed, $m_{s_t} \sim$ AR(1)

$m_{s_t} = a + \lambda m_{s_{t-1}} + v_t$

$a = p_{21} m_1 + p_{12} m_2$

$\lambda = p_{11} - p_{21}$

$v_t \sim$ martingale difference sequence

If only $\Omega_t = \{y_t, y_{t-1}, \ldots, y_1\}$ is observed, Prob$(s_t = 1|\Omega_t)$ is nonlinear in $\Omega_t$.

Given Prob$(s_{t-1} = j|\Omega_{t-1})$, can calculate Prob$(s_t = j|\Omega_t)$ (and likelihood $f(y_t|\Omega_{t-1})$) recursively:

$$\text{Prob}(s_t = j|\Omega_{t-1}) = p_{1j}\text{Prob}(s_{t-1} = 1|\Omega_{t-1}) + p_{2j}\text{Prob}(s_{t-1} = 2|\Omega_{t-1})$$

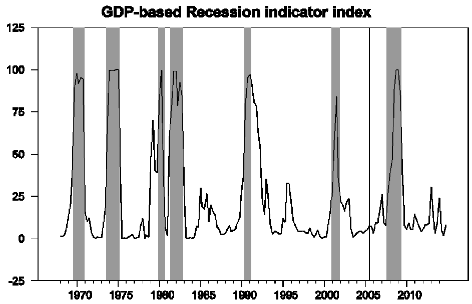
$$f(y_t|\Omega_{t-1}) = \sum_{i=1}^{2}\text{Prob}(s_t = i|\Omega_{t-1})f(y_t|s_t = i, \Omega_{t-1})$$

$$\text{Prob}(s_t = j|\Omega_t) = \frac{\text{Prob}(s_t=j|\Omega_{t-1})f(y_t|s_t=j,\Omega_{t-1})}{f(y_t|\Omega_{t-1})}$$

Could choose population parameters $\theta = (m_1, m_2, \sigma, p_{11}, p_{22})'$ by maximizing likelihood.

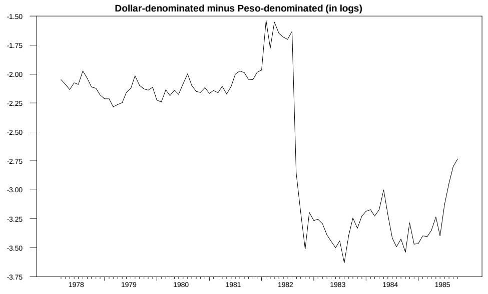
Plot of Prob($s_t = 2|\Omega_{t+1}, \hat{\boldsymbol{\theta}}_{t+1}$) with simulated real-time inference (historical real-time data sets from ALFRED) through 2005.

Plot of actual real-time inference (announced publicly at each date $t+1$) since 2005.




| Date of announcement | Announcement |
|---|---|
| Simulated (through June 2005) | |
| May 1970 | recession began 1969:Q2 |
| Aug 1971 | recession ended 1970:Q4 |
| May 1974 | recession began 1973:Q4 |
| Feb 1976 | recession ended 1975:Q1 |
| Nov 1979 | recession began 1979:Q2 |
| May 1981 | recession ended 1980:Q2 |
| Feb 1982 | recession began 1981:Q2 |
| Aug 1983 | recession ended 1982:Q4 |
| Feb 1991 | recession began 1989:Q4 |
| Feb 1993 | recession ended 1991:Q4 |
| Feb 2002 | recession began 2001:Q1 |
| Aug 2002 | recession ended 2001:Q3 |
| Actual real time (since July 2005) | |
| Jan 30, 2009 | recession began 2007:Q4 |
| Apr 30, 2010 | recession ended 2009:Q2 |

## Another example of change in regime




Model of structural change:

$$y_t - \mu_1 = \phi(y_{t-1} - \mu_1) + \varepsilon_t \qquad t \leq t_0$$
$$y_t - \mu_2 = \phi(y_{t-1} - \mu_2) + \varepsilon_t \qquad t > t_0$$

Questions:

1) How forecast with this model?
2) What caused change at $t_0$?
3) What is probability law for $\{y_t\}$?

$$s_t^* = 1 \quad t = 1, 2, .., t_0$$
$$s_t^* = 2 \quad t = t_0 + 1, t_0 + 2, \ldots$$
$$y_t - \mu_{s_t^*} = \phi(y_{t-1} - \mu_{s_{t-1}^*}) + \varepsilon_t$$

Need: probability law for $s_t^*$

Markov chain:

$$P(s_t^* = j | s_{t-1}^* = i, s_{t-2}^* = k, \ldots)$$
$$= P(s_t^* = j | s_{t-1}^* = i)$$
$$= p_{ij}$$

Transition from 1 to 2 is permanent

$$\Rightarrow p_{21} = 0$$

In general, if $s_t$ is a Markov chain taking on one of the values $s_t = 1, 2, \ldots, N$, let $p_{ij} = P(s_t = j | s_{t-1} = i)$. Collect in matrix $\mathbf{P} = [p_{ji}]$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}$$

In general, if $s_t$ is a Markov chain taking on one of the values $s_t = 1, 2, \ldots, N$, let $p_{ij} = P(s_t = j | s_{t-1} = i)$. Collect in matrix $\mathbf{P} = [p_{ji}]$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}$$

Let $\xi_t = \mathbf{e}_i$ (the $i$th column of $\mathbf{I}_N$) when $s_t = i$. Then

$$E(\xi_{t+1} | \xi_t = \mathbf{e}_i) = \begin{bmatrix} P(s_{t+1} = 1 | s_t = i) \\ P(s_{t+1} = 2 | s_t = i) \\ \vdots \\ P(s_{t+1} = N | s_t = i) \end{bmatrix}$$

$$= \mathbf{P}\mathbf{e}_i$$

$$= \mathbf{P}\xi_t$$

Suppose we had a set of observations $\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1\}$ that gave us an imperfect inference about $s_t$ summarized as

$$\hat{\xi}_{t|t} = E(\xi_t | \Omega_t) = \begin{bmatrix} P(s_t = 1 | \Omega_t) \\ P(s_t = 2 | \Omega_t) \\ \vdots \\ P(s_t = N | \Omega_t) \end{bmatrix}$$

Then

$$\hat{\xi}_{t+1|t} = E(\xi_{t+1}|\Omega_t) = \mathbf{P}\hat{\xi}_{t|t}$$

(e.g., row $j$ states that

$$P(s_{t+1} = j|\Omega_t)$$
$$= p_{1j}P(s_t = 1|\Omega_t) + p_{2j}P(s_t = 2|\Omega_t)$$
$$+ \cdots + p_{Nj}P(s_t = N|\Omega_t)$$

Return to original example of interest:

$$y_t - \mu_{s_t^*} = \phi(y_{t-1} - \mu_{s_{t-1}^*}) + \varepsilon_t$$

$\varepsilon_t \sim$ i.i.d. $N(0,\sigma^2)$

$$P(s_t^* = j|s_{t-1}^* = i) = p_{ij}^* \quad i,j = 1,2$$

$\{s_t^*\}_{t=1}^T$ independent of $\{\varepsilon_t\}_{t=1}^T$

$$\Omega_t = \{y_t, y_{t-1}, \ldots, y_1\}$$

Implication:

$$y_t|\Omega_{t-1}, s_t^*, s_{t-1}^* \sim N\left(\mu_{s_t^*} + \phi(y_{t-1} - \mu_{s_{t-1}^*}), \sigma^2\right)$$

Convenient to summarize $\{s_t^*, s_{t-1}^*\}$ with a single Markov chain:

$s_t = 1$ if $s_t^* = 1$ and $s_{t-1}^* = 1$

$s_t = 2$ if $s_t^* = 2$ and $s_{t-1}^* = 1$

$s_t = 3$ if $s_t^* = 1$ and $s_{t-1}^* = 2$

$s_t = 4$ if $s_t^* = 2$ and $s_{t-1}^* = 2$

$\xi_t = \mathbf{e}_i$ ($i$th column of $\mathbf{I}_4$) when $s_t = i$

$\hat{\xi}_{t+1|t} = \mathbf{P}\hat{\xi}_{t|t}$

$$\mathbf{P} = \begin{bmatrix} p_{11}^* & 0 & p_{11}^* & 0 \\ p_{12}^* & 0 & p_{12}^* & 0 \\ 0 & p_{21}^* & 0 & p_{21}^* \\ 0 & p_{22}^* & 0 & p_{22}^* \end{bmatrix}$$

$$p(y_t|s_t = 3, \Omega_{t-1})$$
$$= p(y_t|\mu_{s_t^*} = \mu_1, \mu_{s_{t-1}^*} = \mu_2, \Omega_{t-1})$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{\frac{-[y_t - \mu_1 - \phi(y_{t-1} - \mu_2)]^2}{2\sigma^2}\right\}$$

Collect the densities that might be associated with each of the $N = 4$ states in an $(N \times 1)$ vector

$$\boldsymbol{\eta}_t = \begin{bmatrix} p(y_t|s_t = 1, \Omega_{t-1}) \\ p(y_t|s_t = 2, \Omega_{t-1}) \\ \vdots \\ p(y_t|s_t = N, \Omega_{t-1}) \end{bmatrix}$$

Recall that

$$\mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1} = \begin{bmatrix} P(s_t = 1|\Omega_{t-1}) \\ P(s_t = 2|\Omega_{t-1}) \\ \vdots \\ P(s_t = N|\Omega_{t-1}) \end{bmatrix}$$

Thus

$$\boldsymbol{\eta}_t \odot \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1} = \begin{bmatrix} p(y_t|s_t = 1, \Omega_{t-1})P(s_t = 1|\Omega_{t-1}) \\ p(y_t|s_t = 2, \Omega_{t-1})P(s_t = 2|\Omega_{t-1}) \\ \vdots \\ p(y_t|s_t = N, \Omega_{t-1})P(s_t = N|\Omega_{t-1}) \end{bmatrix}$$

Summing the elements of this vector gives

$$\mathbf{1}'(\eta_t \odot \mathbf{P}\hat{\xi}_{t-1|t-1})$$
$$= \sum_{j=1}^{N} p(y_t, s_t = j|\Omega_{t-1})$$
$$= p(y_t|\Omega_{t-1}),$$

the conditional likelihood of $t$th observation.

The result of dividing the $j$th element of $(\eta_t \odot \mathbf{P}\hat{\xi}_{t-1|t-1})$ by the conditional likelihood is

$$\frac{p(y_t, s_t = j|\Omega_{t-1})}{p(y_t|\Omega_{t-1})} = P(s_t = j|y_t, \Omega_{t-1})$$

$$\frac{(\eta_t \odot \mathbf{P}\hat{\xi}_{t-1|t-1})}{\mathbf{1}'(\eta_t \odot \mathbf{P}\hat{\xi}_{t-1|t-1})} = \hat{\xi}_{t|t}$$

$$\frac{(\eta_t \odot \mathbf{P}\hat{\xi}_{t-1|t-1})}{\mathbf{1}'(\eta_t \odot \mathbf{P}\hat{\xi}_{t-1|t-1})} = \hat{\xi}_{t|t}$$

Iterative algorithm similar to Kalman filter:

Input for step $t$:

$\hat{\xi}_{t-1|t-1}$

(an $N \times 1$ vector whose $j$th element is $P(s_t = j|y_t, y_{t-1}, \ldots, y_1)$).

Output for step $t$:

$\hat{\xi}_{t|t}$

Options for initial value $\hat{\xi}_{0|0}$:

(1) If Markov chain is ergodic, use ergodic probabilities

$$\hat{\xi}_{0|0} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{e}_{N+1}$$

$$\underset{(N+1)\times N}{\mathbf{A}} = \begin{bmatrix} \mathbf{I}_N - \mathbf{P} \\ \mathbf{1}' \end{bmatrix}$$

(2) Set $\hat{\xi}_{0|0} = \boldsymbol{\rho}$, a vector of free parameters to be estimated by maximum likelihood or Bayesian methods along with the other parameters.

(3) Set $\hat{\xi}_{0|0} = N^{-1}\mathbf{1}$.

(4) Set $\hat{\xi}_{0|0}$ based on prior beliefs.

Above assumed we knew parameters $\boldsymbol{\theta}$ appearing in $\boldsymbol{\eta}_t = [p(y_t|s_t = j, \Omega_{t-1}; \boldsymbol{\theta})]_{j=1}^N$ (in this case $\boldsymbol{\theta} = (\phi, \mu_1, \mu_2, \sigma^2)'$) and $\mathbf{p}$ appearing in $\mathbf{P}$ (in this case $\mathbf{p} = (p_{11}, p_{22})'$).

However, as byproduct of step $t$ of iteration we ended up calculating $p(y_t|\Omega_t; \boldsymbol{\theta}, \mathbf{p})$ and so we've calculated log likelihood

$$\mathcal{L}(\boldsymbol{\theta}, \mathbf{p}) = \sum_{t=1}^T \log p(y_t|\Omega_t; \boldsymbol{\theta}, \mathbf{p})$$

which can be maximized numerically with respect to $\boldsymbol{\theta}$ and $\mathbf{p}$ by numerical methods.

Note– during numerical search we'd want to be choosing $\lambda_{11}$ and $\lambda_{22}$ rather than $p_{11}$ and $p_{22}$ where

$$p_{11} = \frac{\lambda_{11}^2}{1+\lambda_{11}^2}$$

$$p_{22} = \frac{\lambda_{22}^2}{1+\lambda_{22}^2}$$

General case:

$$\boldsymbol{\eta}_t = \begin{bmatrix} p(\mathbf{y}_t|s_t = 1, \Omega_{t-1}) \\ p(\mathbf{y}_t|s_t = 2, \Omega_{t-1}) \\ \vdots \\ p(\mathbf{y}_t|s_t = N, \Omega_{t-1}) \end{bmatrix}$$

$$p(\mathbf{y}_t|\Omega_{t-1}) = \mathbf{1}'(\boldsymbol{\eta}_t \odot \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1})$$

$$\mathcal{L}(\mathbf{y}_1, \ldots, \mathbf{y}_T; \boldsymbol{\theta}) = \sum_{t=1}^{T} \log p(\mathbf{y}_t|\Omega_{t-1})$$

# IV. Markov-switching models

A. Introduction to Markov-switching models
B. Economic theory and changes in regime

B.1. Closed-form solution of DSGE's and asset-pricing implications

Lucas tree model with CRRA utility:

$P_t$ = price of stock

$D_t$ = dividend

$\gamma$ = coefficient of relative risk aversion

$$P_t = D_t^{-\gamma} \sum_{k=1}^{\infty} \beta^k E_t D_{t+k}^{(1+\gamma)}$$

Cecchetti, Lam and Mark (1990):

$$\log D_t - \log D_{t-1} = m_{s_t} + \varepsilon_t$$

$$P_t = \rho_{s_t} D_t$$

B.2. Linear rational expectations models with changes in regime

$$\mathbf{A}_{s_t} E(\mathbf{y}_{t+1} | \Omega_t, s_t, s_{t-1}, \ldots, s_1) = \mathbf{d}_{s_t} + \mathbf{B}_{s_t} \mathbf{y}_t + \mathbf{C}_{s_t} \mathbf{x}_t$$

$\mathbf{A}_j = (n_y \times n_y)$ matrix of parameters when $s_t = j$.

Davig and Leeper (2007):

Let $\mathbf{y}_{jt}$ = value of $\mathbf{y}_t$ when $s_t = j$

$$\underset{(Nn_y \times 1)}{\mathbf{Y}_t} = \begin{bmatrix} \underset{(n_y \times 1)}{\mathbf{y}_{1t}} \\ \vdots \\ \underset{(n_y \times 1)}{\mathbf{y}_{Nt}} \end{bmatrix}$$

$$E(\mathbf{y}_{t+1}|s_t = i, \Omega_t) = \sum_{j=1}^{N} E(\mathbf{y}_{t+1}|s_{t+1} = j, s_t = i, \Omega_t) p_{ij}$$

Hence when $s_t = i$,

$$\mathbf{A}_{s_t} E(\mathbf{y}_{t+1}|s_t, \Omega_t) = (\mathbf{p}_i' \otimes \mathbf{A}_i) E(\mathbf{Y}_{t+1}|\mathbf{Y}_t)$$

$$\underset{(Nn_y \times Nn_y)}{\mathbf{A}} = \begin{bmatrix} \underset{(1\times N)}{\mathbf{p}_1'} \otimes \underset{(n_y \times n_y)}{\mathbf{A}_1} \\ \vdots \\ \underset{(1\times N)}{\mathbf{p}_N'} \otimes \underset{(n_y \times n_y)}{\mathbf{A}_N} \end{bmatrix} \qquad \underset{(Nn_y \times 1)}{\mathbf{d}} = \begin{bmatrix} \underset{(n_y \times 1)}{\mathbf{d}_1} \\ \vdots \\ \underset{(n_y \times 1)}{\mathbf{d}_N} \end{bmatrix}$$

$$\underset{(Nn_y \times Nn_y)}{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_N \end{bmatrix} \qquad \underset{(Nn_y \times n_x)}{\mathbf{C}} = \begin{bmatrix} \underset{(n_y \times n_x)}{\mathbf{C}_1} \\ \vdots \\ \underset{(n_y \times n_x)}{\mathbf{C}_N} \end{bmatrix}$$

Consider non-regime-changing system

$\mathbf{A}E(\mathbf{Y}_{t+1}|\mathbf{Y}_t) = \mathbf{d} + \mathbf{B}\mathbf{Y}_t + \mathbf{C}\mathbf{x}_t$

If we can find a stable solution of the form

$$\underset{(Nn_y\times 1)}{\mathbf{Y}_t} = \underset{(Nn_y\times 1)}{\mathbf{h}} + \underset{(Nn_y\times n_x)}{\mathbf{H}}\underset{(n_x\times 1)}{\mathbf{x}_t}$$

then the $i$th block

$$\underset{(n_y\times 1)}{\mathbf{y}_t} = \underset{(n_y\times 1)}{\mathbf{h}_{s_t}} + \underset{(n_y\times n_x)}{\mathbf{H}_{s_t}}\underset{(n_x\times 1)}{\mathbf{x}_t}$$

is a stable solution to our original equation of interest.

- However, even if we find a unique stable solution to the invariant system, there may be other stable solutions to the original system
  - Farmer, Waggoner, and Zha (2010)

B.3. Multiple equilibria

- Multiplicity of stable equilibria could itself be of interest
  - Coordination externalities (Cooper and John, 1988; Cooper, 1994)
  - Equilibria indexed by expectations (Kurz and Motolese, 2001)
- Regime-switching model could describe transitions between equilibria
  - Kirman (1993); Chamley (1999)

B.4. Tipping points and financial crises

- In other models, there is a unique equilibrium, but small change in fundamentals can cause big change in outcome

  - Acemoglu and Scott (1997); Moore and Schaller (2002); Guo, Miao, and Morelle (2005); Veldkamp (2005); Startz (1998); Hong, Stein, and Yu (2007); Branch and Evans (2010)
- Financial crises

  - Brunnermeier and Sannikov (2014); Hamilton (2005); Asea and Blomberg (1998); Hubrich and Tetlow (2013)

B.5. Currency crises and sovereign debt crises

- Currency crises

  - Jeanne and Masson (2000); Peria (2002); Cerra and Saxena (2005)
- Sovereign debt crises

  - Greenlaw, et. al. (2013); Davig, Leeper and Walker (2011); Bi (2012)

B.6. Changes in policy as the source of changes in regime

- Monetary policy: hawks vs. doves

  Owyang and Ramey (2004); Schorfheide (2005); Liu, Waggoner, and Zha (2011); Bianchi (2013)
- Unsustainable fiscal policy and inflation

  Ruge-Murcia (1995, 1999)

# IV. Markov-switching models

A. Introduction to Markov-switching models
B. Economic theory and changes in regime
C. Extensions

C.1. Selecting the number of regimes

Smith, Naik and Tsai (2006):

$y_t = \mathbf{x}_t'\boldsymbol{\beta}_{s_t} + \sigma_{s_t}\varepsilon_t$

$\hat{T}_i = \sum_{t=1}^{T} \text{Prob}(s_t = i|\Omega_T;\hat{\boldsymbol{\theta}}_{MLE})$

$MSC = -2\mathcal{L}(\hat{\boldsymbol{\theta}}_{MLE}) + \sum_{i=1}^{N} \frac{\hat{T}_i(\hat{T}_i+Nk)}{\hat{T}_i-Nk-2}$

- Calculate nonstandard properties of likelihood ratio test
  - Hansen (1992)
  - Garcia (1998)
- Use general specification tests of null of *N* regimes that have power against *N* +1
  - Hamilton (1996)
  - Carrasco, Hu and Ploberger (2014)

C.2. Chib's multiple change-point model

$$\mathbf{P} = \begin{bmatrix} p_{11} & 0 & 0 & \cdots & 0 & 0 \\ 1-p_{11} & p_{22} & 0 & \cdots & 0 & 0 \\ 0 & 1-p_{22} & p_{33} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p_{N-1,N-1} & 0 \\ 0 & 0 & 0 & \cdots & 1-p_{N-1,N-1} & 1 \end{bmatrix}$$

C.3. Allowing any parameters of distribution to change

$$\boldsymbol{\eta}_t = \begin{bmatrix} p(\mathbf{y}_t|s_t = 1, \Omega_{t-1}) \\ p(\mathbf{y}_t|s_t = 2, \Omega_{t-1}) \\ \vdots \\ p(\mathbf{y}_t|s_t = N, \Omega_{t-1}) \end{bmatrix}$$

$$p(\mathbf{y}_t|\Omega_{t-1}) = \mathbf{1}'(\boldsymbol{\eta}_t \odot \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1})$$

$$\mathcal{L}(\mathbf{y}_1,\ldots,\mathbf{y}_T;\boldsymbol{\theta}) = \sum_{t=1}^{T} \log p(\mathbf{y}_t|\Omega_{t-1})$$

Example: Dueker (JBES, 1997). AR(1) with Student t innovations whose degrees of freedom change with regime:

$$p(y_t|s_t = j, \Omega_{t-1}; \boldsymbol{\theta}) = \frac{\Gamma[(\nu_j+1)/2]}{(\pi\nu_j)^{1/2}\Gamma(\nu_j/2)\sigma} \left[1 + \frac{(y_t - c - \phi y_{t-1})^2}{\sigma\nu_j}\right]^{-(\nu_j+1)/2}$$

$$\boldsymbol{\theta} = (c, \phi, \sigma, \nu_1, \nu_2)'$$

Example: Krolzig (*Markov-Switching Vector Autoregressions,* Springer 1997): Gaussian VAR(1) with lag coefficients changing:

$$p(\mathbf{y}_t|s_t = j, \Omega_{t-1}, \boldsymbol{\theta}) = (2\pi)^{-n/2}|\boldsymbol{\Omega}|^{-1/2}$$
$$\exp[-(1/2)(\mathbf{y}_t - \mathbf{c}_j - \boldsymbol{\Phi}_j\mathbf{y}_{t-1})'\boldsymbol{\Omega}^{-1}$$
$$(\mathbf{y}_t - \mathbf{c}_j - \boldsymbol{\Phi}_j\mathbf{y}_{t-1})]$$
$$\boldsymbol{\theta} = (\mathbf{c}_1', \mathbf{c}_2', (\mathrm{vec}\,\Phi_1')', (\mathrm{vec}\,\Phi_2')', (\mathrm{vech}\,\boldsymbol{\Omega})')'$$

Can also allow transition probabilities to be parametric function of exogenous or lagged dependent variables $\mathbf{z}_t$:

$$\mathbf{P} = [P(s_t = j|s_{t-1} = i, \mathbf{z}_t; \mathbf{p})]_{i,j=1}^{N}.$$

Example:

$$\mathbf{P} = \begin{bmatrix} \frac{\exp(\boldsymbol{\gamma}'\mathbf{z}_t)}{1+\exp(\boldsymbol{\gamma}'\mathbf{z}_t)} & \frac{1}{1+\exp(\boldsymbol{\delta}'\mathbf{z}_t)} \\ \frac{1}{1+\exp(\boldsymbol{\gamma}'\mathbf{z}_t)} & \frac{\exp(\boldsymbol{\delta}'\mathbf{z}_t)}{1+\exp(\boldsymbol{\delta}'\mathbf{z}_t)} \end{bmatrix}$$

$$\mathbf{p} = (\boldsymbol{\gamma}', \boldsymbol{\delta}')'$$

What can't we do? Models where $y_t$ depends on a growing number of states:

$$p(y_t|\Omega_{t-1}, s_t^*, s_{t-1}^*, \ldots, s_1^*; \boldsymbol{\theta})$$

Example: ARMA process

$$y_t = \varepsilon_t + \theta_{s_{t-1}^*}\varepsilon_{t-1}$$
$$\Rightarrow y_t = \varepsilon_t + \theta_{s_{t-1}^*}y_{t-1} - \theta_{s_{t-1}^*}\theta_{s_{t-2}^*}y_{t-2}$$
$$+\theta_{s_{t-1}^*}\theta_{s_{t-2}^*}\theta_{s_{t-3}^*}y_{t-3} + \cdots$$

Example: GARCH process:

$$y_t = \sqrt{h_t}\,\varepsilon_t$$

$$\begin{aligned} h_t &= \zeta_{s_t} + \alpha y_{t-1}^2 + \delta h_{t-1} \\ &= \zeta_{s_t} + \alpha y_{t-1}^2 + \delta(\zeta_{s_{t-1}} + \alpha y_{t-2}^2) \\ &\quad + \delta^2(\zeta_{s_{t-2}} + \alpha y_{t-3}^2) + \cdots \end{aligned}$$

Solution:

numerical Bayesian methods

## IV. Markov-switching models

A. Introduction to Markov-switching models
B. Economic theory and changes in regime
C. Extensions
D. Bayesian analysis of Markov-switching models

Example:

$y_t = \boldsymbol{\beta}_{s_t}' \mathbf{x}_t + \varepsilon_t$

$\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$

$P(s_t = j | s_{t-1} = i) = p_{ij} \quad i,j = 1,2$

(does not depend on $\mathbf{x}_{t-k}, \varepsilon_{t-k}, s_{t-k-1}$ for $k = 0, 1, 2, \ldots$)

Gibbs sampler:

$\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2', \boldsymbol{\theta}_3', \boldsymbol{\theta}_4')'$

$\theta_1 = \sigma^{-2}$

$\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$

$\boldsymbol{\theta}_3 = (p_{11}, p_{22})'$

$\boldsymbol{\theta}_4 = (s_1, s_2, \ldots, s_T)'$

(1) Generating $\theta_1 = \sigma^{-2}$ from $p(\theta_1 | \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X})$.

Prior: $\sigma^{-2} \sim \Gamma(N, \lambda)$

Conditioning on $\boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X}$ is equivalent to observing $\{\varepsilon_t\}_{t=1}^T$ for

$\varepsilon_t = y_t - \boldsymbol{\beta}_{s_t}' \mathbf{x}_t$

Posterior:

$\sigma^{-2} | \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X} \sim \Gamma(N + T, \lambda + S)$

$S = \sum_{t=1}^T \varepsilon_t^2$

(2) Generating $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ from $p(\boldsymbol{\theta}_2 | \theta_1, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X})$.

Priors:

$\boldsymbol{\beta}_i | \sigma^{-2} \sim N(\mathbf{m}_i, \sigma^2 \mathbf{M}_i) \quad i = 1, 2$

(independent of each other)

Posterior:

Conditioning on $\{s_t\}_{t=1}^{T}$, only those observations $t$ for which $s_t = 1$ are relevant for posterior distribution of $\boldsymbol{\beta}_1$.

$$\boldsymbol{\beta}_i | \theta_1, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X} \sim N(\mathbf{m}_i^*, \sigma^2 \mathbf{M}_i^*)$$

$$\mathbf{M}_i^* = \left( \mathbf{M}_i^{-1} + \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \delta_{(s_t = i)} \right)^{-1}$$

$$\mathbf{m}_i^* = \mathbf{M}_i^* \left( \mathbf{M}_i^{-1} \mathbf{m}_i + \sum_{t=1}^{T} \mathbf{x}_t y_t \delta_{(s_t = i)} \right)$$

Label-switching problem:

If switch $\boldsymbol{\beta}_1$ with $\boldsymbol{\beta}_2$ and $p_{11}$ with $p_{22}$, value of likelihood $p(\mathbf{Y} | \mathbf{X}, \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ is identical.

Implication: if priors for $\boldsymbol{\beta}_i$ and $p_{ii}$ are same for $i = 1, 2$, then true posterior distribution is bimodal and perfectly symmetric around the two modes.

Presume we have interpretive (as opposed to numerical) labels for regimes. E.g., regime 2 = "recession", should have faster GDP growth, so that, say, $\beta_1(1)$, first element of $\boldsymbol{\beta}_1$, should be bigger $\beta_2(1)$, the first element of $\boldsymbol{\beta}_2$.

Strategy (1): Intentionally use symmetric priors for regimes 1 and 2 and intentionally randomly perturb parameter draw $j$ to switch across modes so as to get multimodal posterior distribution, and apply normalization rule to this.

Strategy (2): Impose normalization requirement $\beta_1(1) > \beta_2(1)$ at every draw.

Drawback to (2): not clear it's same distribution as (1).

Drawback to either approach: Even though normalized posterior distribution has unique global mode, may still have local modes resulting from label switching.

Recommendation: plot posterior distributions to check for this.

(3) Generating $\boldsymbol{\theta}_3 = (p_{11}, p_{22})'$ from $p(\boldsymbol{\theta}_3 | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X})$.

Priors:

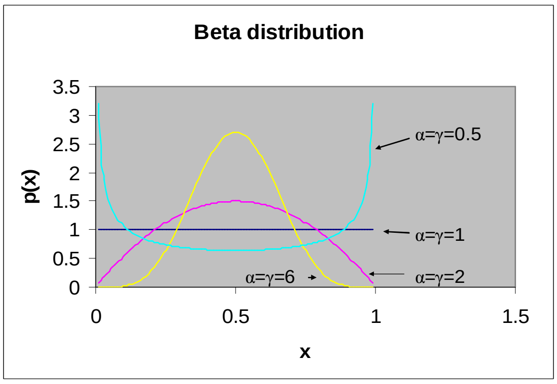
A variable $x$ is said to have a beta distribution with parameters $\alpha > 0$ and $\gamma > 0$, denoted $x \sim \mathrm{Beta}(\alpha, \gamma)$, if

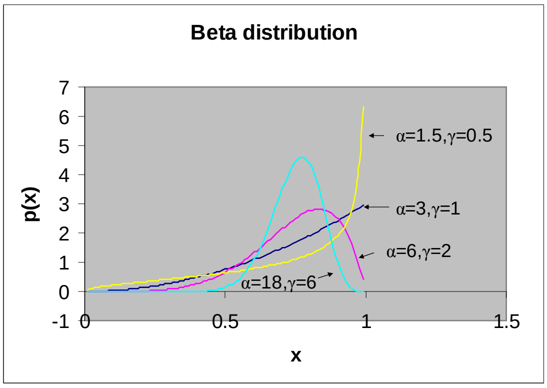
$$p(x|\alpha,\gamma) = \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\Gamma(\gamma)} x^{\alpha-1}(1-x)^{\gamma-1}$$

for $0 < x < 1$ and $p(x|\alpha,\gamma) = 0$ elsewhere.

$$E(x) = \frac{\alpha}{\alpha+\gamma}$$

$$V(x) = \frac{\alpha\gamma}{(\alpha+\gamma)^2(\alpha+\gamma+1)}$$




Priors:

$$p_{ii} \sim \text{Beta}(\alpha_i, \gamma_i) \qquad i = 1, 2$$

(independent of each other)

Posterior:

Observation of $\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X}$ only affects inference about $p_{ii}$ through $\boldsymbol{\theta}_4 = (s_1, s_2, \ldots, s_T)'$.

Assume that initial probability $P(s_1 = 1)$ does not depend on $p_{ii}$ (don't use ergodic probabilities).

Suppose that in the sequence $\boldsymbol{\theta}_4 = (s_1, s_2, \ldots, s_T)'$ state $s_t = 1$ is observed to be followed by $s_{t+1} = 1$ a total of $n_{11}$ times, whereas state $s_t = 1$ is followed by $s_{t+1} = 2$ a total of $n_{12}$ times.

Then, for purposes of inference about $p_{11}$, can view the data $\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X}$ solely as a sample of $n_{11} + n_{12}$ observations from a Bernoulli variable with probability of success $p_{11}$:

$$p(\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X} | \boldsymbol{\theta}_3) \propto p_{11}^{n_{11}} (1 - p_{11})^{n_{12}}$$

data:

$p(\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X} | \boldsymbol{\theta}_3) \propto p_{11}^{n_{11}}(1 - p_{11})^{n_{12}}$

prior: $p_{ii} \sim \text{Beta}(\alpha_i, \gamma_i)$

$p(\boldsymbol{\theta}_3) \propto p_{11}^{\alpha_1 - 1}(1 - p_{11})^{\gamma_1 - 1}$

posterior: $p_{ii} \sim \text{Beta}(\alpha_i^*, \gamma_i^*)$

$\alpha_1^* = \alpha_1 + n_{11}$

$\gamma_1^* = \gamma_1 + n_{12}$

$\alpha_2^* = \alpha_2 + n_{22}$

$\gamma_2^* = \gamma_2 + n_{21}$

(4) Generating $\boldsymbol{\theta}_4 = (s_1, s_2, \ldots, s_T)'$ from $p(\boldsymbol{\theta}_4 | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{Y}, \mathbf{X})$.

Calculate $P(s_T = 1 | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{Y}, \mathbf{X})$ from first element of $\hat{\boldsymbol{\xi}}_{T|T}$.

Generate $U_T \sim U(0,1)$ and set $S_T = 1$ if $U_T < \mathbf{e}_1' \hat{\boldsymbol{\xi}}_{T|T}$.

Consider

$$P(S_t = i | S_{t+1} = j, S_{t+2} = k, \ldots, S_T = z, \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{Y}, \mathbf{X})$$
$$= P(S_t = i | S_{t+1} = j, \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)$$

for $\Omega_t = \{y_t, y_{t-1}, \ldots, y_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1\}$

$$P(S_t = i | S_{t+1} = j, \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)$$

$$= \frac{P(S_t = i, S_{t+1} = j | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)}{P(S_{t+1} = j | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)}$$

$$= \frac{P(S_{t+1} = j | S_t = i) P(S_t = i | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)}{P(S_{t+1} = j | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)}$$

$$= \frac{p_{ij} \mathbf{e}_i' \hat{\boldsymbol{\xi}}_{t|t}}{\mathbf{e}_j' \mathbf{P} \hat{\boldsymbol{\xi}}_{t|t}}$$

Iterating backwards $t = T-1, T-2, \ldots$ we generate the sequence $\boldsymbol{\theta}_4 = (s_1, s_2, \ldots, s_T)'$ from $p(\boldsymbol{\theta}_4 | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{Y}, \mathbf{X})$.

Generalization: $N$-state Markov chain.

$x \sim \text{Beta}(\alpha_1, \alpha_2)$

$$p(x | \alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}$$

$$0 < x < 1$$

$$(x_1, x_2, \ldots, x_N)' \sim \text{Dirichlet}(\alpha_1, \alpha_2, \ldots, \alpha_N)$$

$$p(x|\alpha_1, \ldots, \alpha_N) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_N)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_N)} x_1^{\alpha_1 - 1} \cdots x_N^{\alpha_N - 1}$$

$0 < x_i < 1$

$x_1 + \cdots + x_N = 1$

$\alpha_i > 0$

prior:

$(p_{11}, p_{12}, \ldots, p_{1N})'$

$\sim \text{Dirichlet}(\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1N})$

data:

$n_{1j}$ = number of times $s_t = 1$ is followed by $s_{t+1} = j$

posterior:

$(p_{11}, p_{12}, \ldots, p_{1N})' | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{Y}, \mathbf{X}$

$\sim \text{Dirichlet}(\alpha_{11} + n_{11}, \ldots, \alpha_{1N} + n_{1N})$

Generalization: time-dependent transition probabilities:

$P(s_t = j | s_{t-1} = i, \mathbf{x}_t)$

$\mathbf{x}_t$ predetermined at $t$

Convenient framework for Gibbs sampling: latent variable $z_t^*$

$$z_t^* = \gamma_0 s_{t-1} + \boldsymbol{\gamma}' \mathbf{x}_t + u_t$$

$$u_t \sim N(0,1)$$

$$s_t = \begin{cases} 1 & \text{if } z_t^* < 0 \\ 2 & \text{if } z_t^* \geq 0 \end{cases}$$

Gibbs sampler:

$$\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2', \boldsymbol{\theta}_3', \boldsymbol{\theta}_4')'$$

$$\theta_1 = \sigma^{-2}$$

$$\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$$

$$\boldsymbol{\theta}_3 = (s_1, s_2, \ldots, s_T, z_1^*, z_2^*, \ldots, z_T^*)'$$

$$\boldsymbol{\theta}_4 = (\gamma_0, \boldsymbol{\gamma}')'$$

Draws from $p(\theta_1 | \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X})$ and $p(\boldsymbol{\theta}_2 | \theta_1, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X})$ same as before (conditioning on $\{z_t^*\}$ adds no information beyond that in $\{s_t\}$).

Will draw $\boldsymbol{\theta}_3$ using

$$p(\boldsymbol{\theta}_3|\theta_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_4,\mathbf{Y},\mathbf{X}) = p(s_1,\ldots,s_T|\theta_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_4,\mathbf{Y},\mathbf{X}) \times p(z_1^*,\ldots,z_T^*|s_1,\ldots,s_T,\theta_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_4,\mathbf{Y},\mathbf{X})$$

(a) To draw from $p(s_1,\ldots,s_T|\theta_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_4,\mathbf{Y},\mathbf{X})$ use modification of filter for time-varying probabilities

Recall filter with constant probabilities:

$j$th element of $\boldsymbol{\eta}_t$:

$$\eta_{jt} = p(y_t|s_t = j,\Omega_{t-1})$$

$j$th element of $\hat{\boldsymbol{\xi}}_{t|t-1} = \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1}$:

$$p(s_t = j|\Omega_{t-1})$$

$j$th element of $\boldsymbol{\eta}_t \odot \hat{\boldsymbol{\xi}}_{t|t-1}$:

$$p(y_t,s_t = j|\Omega_{t-1})$$

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\boldsymbol{\eta}_t \odot \hat{\boldsymbol{\xi}}_{t|t-1}}{\mathbf{1}'(\boldsymbol{\eta}_t \odot \hat{\boldsymbol{\xi}}_{t|t-1})}$$

Filter with time-varying probabilities:

$$z_t^* = \gamma_0 s_{t-1} + \boldsymbol{\gamma}'\mathbf{x}_t + u_t$$

$$u_t \sim N(0,1)$$

$s_t = 1$ if $z_t^* < 0$

$$
\begin{aligned}
p(s_t = 1|\Omega_{t-1}) = \\
&p(s_t = 1|s_{t-1} = 1, \Omega_{t-1})p(s_{t-1} = 1|\Omega_{t-1}) \\
&+ p(s_t = 1|s_{t-1} = 2, \Omega_{t-1})p(s_{t-1} = 2|\Omega_{t-1}) \\
= p(\gamma_0 + \boldsymbol{\gamma}'\mathbf{x}_t + u_t < 0)p(s_{t-1} = 1|\Omega_{t-1}) \\
&+ p(2\gamma_0 + \boldsymbol{\gamma}'\mathbf{x}_t + u_t < 0)p(s_{t-1} = 2|\Omega_{t-1}) \\
= \Phi(-\gamma_0 - \boldsymbol{\gamma}'\mathbf{x}_t)p(s_{t-1} = 1|\Omega_{t-1}) \\
&+ \Phi(-2\gamma_0 - \boldsymbol{\gamma}'\mathbf{x}_t)p(s_{t-1} = 2|\Omega_{t-1})
\end{aligned}
$$

So we just replace $\hat{\boldsymbol{\xi}}_{t|t-1} = \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1}$ in the regular filter with

$$\hat{\boldsymbol{\xi}}_{t|t-1} = \begin{bmatrix} p(s_t = 1|\Omega_{t-1}) \\ p(s_t = 2|\Omega_{t-1}) \end{bmatrix}$$

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\boldsymbol{\eta}_t \odot \hat{\boldsymbol{\xi}}_{t|t-1}}{\mathbf{1}'(\boldsymbol{\eta}_t \odot \hat{\boldsymbol{\xi}}_{t|t-1})}$$

(i) Run through filter to calculate $\{\hat{\boldsymbol{\xi}}_{t|t}, \hat{\boldsymbol{\xi}}_{t|t-1}\}_{t=1}^{T}$

(ii) Generate $s_T = 1$ with probability $\mathbf{e}_1'\hat{\boldsymbol{\xi}}_{T|T}$ and $s_T = 2$ with probability $\mathbf{e}_2'\hat{\boldsymbol{\xi}}_{T|T}$.

(iii) To get $s_t$ for $t = T-1, T-2, \ldots$ use modification of earlier smoothing algorithm:

$$
\begin{aligned}
P(S_t = i|S_{t+1} = j, \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \Omega_t) \\
&= \frac{P(S_t = i, S_{t+1} = j|\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \Omega_t)}{P(S_{t+1} = j|\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \Omega_t)} \\
&= \frac{P(S_{t+1} = j|S_t = i)P(S_t = i|\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \Omega_t)}{P(S_{t+1} = j|\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \Omega_t)} \\
&= \begin{cases} \dfrac{\Phi(-\gamma_0 i - \boldsymbol{\gamma}'\mathbf{x}_t)\mathbf{e}_i'\hat{\boldsymbol{\xi}}_{t|t}}{\mathbf{e}_1'\hat{\boldsymbol{\xi}}_{t+1|t}} & \text{if } j = 1 \\ \dfrac{[1-\Phi(-\gamma_0 i - \boldsymbol{\gamma}'\mathbf{x}_t)]\mathbf{e}_i'\hat{\boldsymbol{\xi}}_{t|t}}{\mathbf{e}_2'\hat{\boldsymbol{\xi}}_{t+1|t}} & \text{if } j = 2 \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
p(\boldsymbol{\theta}_3|\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X}) = \\
&p(s_1, \ldots, s_T|\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X}) \\
&\times\ p(z_1^*, \ldots, z_T^*|s_1, \ldots, s_T, \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X})
\end{aligned}
$$

(b) To draw from

$p(z_1^*, \ldots, z_T^*|s_1, \ldots, s_T, \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \mathbf{Y}, \mathbf{X})$

$z_t^* = \gamma_0 s_{t-1} + \boldsymbol{\gamma}'\mathbf{x}_t + u_t$

Generate $z_t^* \sim N(\gamma_0 s_{t-1} + \boldsymbol{\gamma}'\mathbf{x}_t, 1)$

keep if $sign(z_t^*) = sign(s_t - 1.5)$

otherwise draw new $z_t^*$

Finally, to generate a value for $\boldsymbol{\theta}_4 = (\gamma_0, \boldsymbol{\gamma}')'$ given $\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{Y}, \mathbf{X}$, notice that conditional on having observed $\{s_{t-1}, z_t^*\}_{t=1}^T$, generating $\boldsymbol{\theta}_4$ is standard regression problem:

$$z_t^* = \gamma_0 s_{t-1} + \boldsymbol{\gamma}'\mathbf{x}_t + u_t$$

# IV. Markov-switching models

A. Introduction to Markov-switching models
B. Economic theory and changes in regime
C. Extensions
D. Bayesian analysis of Markov-switching models
E. State-space models with Markov switching

$$\xi_{t+1} = \mathbf{F}_{s_t}\xi_t + \mathbf{v}_{t+1} \qquad E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}') = \mathbf{Q}_{s_t}$$

$$\mathbf{y}_t = \mathbf{A}_{s_t}'\mathbf{x}_t + \mathbf{H}_{s_t}'\xi_t + \mathbf{w}_t \qquad E(\mathbf{w}_t\mathbf{w}_t') = \mathbf{R}_{s_t}$$

$$P(s_{t+1} = j | s_t = i) = p_{ij}$$

$\{s_t\}, \{\mathbf{v}_t\}, \{\mathbf{w}_t\}$ independent

$\mathbf{x}_t$ predetermined exogenous

$\theta_1$ = unknown elements of $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{R}_1, \mathbf{R}_2$

$\theta_2$ = unknown elements of $\mathbf{F}_1, \mathbf{F}_2,$
$\mathbf{A}_1, \mathbf{A}_2, \mathbf{H}_1, \mathbf{H}_2$

$\theta_3 = (p_{11}, p_{22})'$

$\theta_4 = (s_0, s_1, s_2, \ldots, s_T)'$

$\theta_5$ = unknown elements of $\{\xi_0, \xi_1, \ldots, \xi_T\}$

(1) Generating $\theta_1$ $\{\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{R}_1, \mathbf{R}_2\}$ given $\{\mathbf{Y}, \mathbf{X}, \theta_2, \theta_3, \theta_4, \theta_5\}$.

$$\xi_{t+1} = \mathbf{F}_{s_t}\xi_t + \mathbf{v}_{t+1} \qquad E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}') = \mathbf{Q}_{s_t}$$

Notice for purposes of estimating $\mathbf{Q}_1$, the likelihood satisfies

$$p(\mathbf{Y}|\mathbf{X}, \theta_2, \theta_3, \theta_4, \theta_5) \propto \prod_{t=1}^{T} |\mathbf{Q}_{s_{t-1}}|^{-1/2}$$
$$\exp\left\{-(1/2)\sum_{t=1}^{T}(\xi_t - \mathbf{F}_{s_t-1}\xi_{t-1})'\right.$$
$$\left.\mathbf{Q}_{s_{t-1}}^{-1}(\xi_t - \mathbf{F}_{s_t-1}\xi_{t-1})\right\}$$
$$\propto \prod_{t=1}^{T} |\mathbf{Q}_1|^{-(1/2)\delta_{s_{t-1}=1}}$$
$$\exp\left\{-(1/2)\sum_{t=1}^{T}(\xi_t - \mathbf{F}_{s_t-1}\xi_{t-1})'\right.$$
$$\left.\mathbf{Q}_1^{-1}(\xi_t - \mathbf{F}_{s_t-1}\xi_{t-1})\delta_{s_{t-1}=1}\right\}$$

prior:

$$\mathbf{Q}_1^{-1} \sim W(N_{Q_1}, \Lambda_{Q_1})$$

posterior:

$$\mathbf{Q}_1^{-1}|\theta_2, \theta_3, \theta_4, \theta_5, \mathbf{Y}, \mathbf{X}$$
$$\sim W(N_{Q_1} + T_1, \Lambda_{Q_1} + \mathbf{S}_{Q_1})$$
$$T_1 = \sum_{t=1}^{T}\delta_{s_{t-1}=1}$$
$$\mathbf{S}_{Q_1} = \sum_{t=1}^{T}\mathbf{v}_t\mathbf{v}_t'\delta_{s_{t-1}=1}$$
$$\mathbf{v}_t = \xi_t - \mathbf{F}_{s_t-1}\xi_{t-1}$$

(2) Generating $\theta_2$ $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{A}_1, \mathbf{A}_2, \mathbf{H}_1, \mathbf{H}_2\}$ given $\{\mathbf{Y}, \mathbf{X}, \theta_1, \theta_3, \theta_4, \theta_5\}$.

$$\xi_{t+1} = \mathbf{F}_{s_t}\xi_t + \mathbf{v}_{t+1} \qquad E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}') = \mathbf{Q}_{s_t}$$

prior: $\mathbf{f}_2|\mathbf{Q}_2 \sim N(\mathbf{m}_{F_2}, \mathbf{Q}_2 \otimes \mathbf{M}_{F_2})$

posterior: $\mathbf{f}_2|\mathbf{Y}, \mathbf{X}, \theta_1, \theta_3, \theta_4, \theta_5$

$$\sim N(\mathbf{m}_{F_2}^*, \mathbf{Q}_2 \otimes \mathbf{M}_{F_2}^*)$$

$$\mathbf{M}_{F_2}^* = \left(\mathbf{M}_{F_2}^{-1} + \sum_{t=1}^{T} \xi_{t-1}\xi_{t-1}'\delta_{s_{t-1}=2}\right)^{-1}$$

$$\mathbf{m}_{F_2}^* = (\mathbf{I}_r \otimes \mathbf{M}_{F_2}^*\mathbf{M}_{F_2}^{-1})\mathbf{m}_{F_2} + \left(\mathbf{I}_r \otimes \mathbf{M}_{F_2}^* \sum_{t=1}^{T} \xi_{t-1}\xi_{t-1}'\delta_{s_{t-1}=2}\right)\hat{\mathbf{f}}_2$$

$$\hat{\mathbf{f}}_2 = \mathrm{vec}(\widehat{\mathbf{F}_2'})$$

$$\widehat{\mathbf{F}_2'} = \left(\sum_{t=1}^{T} \xi_{t-1}\xi_{t-1}'\delta_{s_{t-1}=2}\right)^{-1} \left(\sum_{t=1}^{T} \xi_{t-1}\xi_t'\delta_{s_{t-1}=2}\right)$$

(3) Generating $\theta_3 = (p_{11}, p_{22})'$ given $\{\mathbf{Y}, \mathbf{X}, \theta_1, \theta_2, \theta_4, \theta_5\}$.

(4) Generating $\theta_4 = (s_0, s_1, s_2, \ldots, s_T)'$ given $\{\mathbf{Y}, \mathbf{X}, \theta_1, \theta_2, \theta_3, \theta_5\}$.

Exactly same as for other Markov-switching models.

(5) Generating $\theta_5$ $\{\xi_0, \xi_1, \ldots, \xi_T\}$ given $\{\mathbf{Y}, \mathbf{X}, \theta_1, \theta_2, \theta_3, \theta_4\}$.

$$\xi_{t+1} = \mathbf{F}_{s_t}\xi_t + \mathbf{v}_{t+1} \qquad E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}') = \mathbf{Q}_{s_t}$$

$$\mathbf{y}_t = \mathbf{A}_{s_t}'\mathbf{x}_t + \mathbf{H}_{s_t}'\xi_t + \mathbf{w}_t \qquad E(\mathbf{w}_t\mathbf{w}_t') = \mathbf{R}_{s_t}$$

Conditional on $\{s_0, s_1, \ldots, s_T\}$, this is just a Kalman filter problem where we use different $\mathbf{F}, \mathbf{Q}, \mathbf{A}, \mathbf{H}, \mathbf{R}$ for different dates.

$$\mathbf{P}_{t+1|t} = \mathbf{F}_{s_t}\mathbf{P}_{t|t}\mathbf{F}_{s_t}' + \mathbf{Q}_{s_t}$$

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|t} - \{\mathbf{P}_{t+1|t}\mathbf{H}_{s_{t+1}} (\mathbf{H}_{s_{t+1}}'\mathbf{P}_{t+1|t}\mathbf{H}_{s_{t+1}} + \mathbf{R}_{s_{t+1}})^{-1}\mathbf{H}_{s_{t+1}}'\mathbf{P}_{t+1|t}\}$$

$$\hat{\xi}_{t+1|t} = \mathbf{F}_{s_t}\hat{\xi}_{t|t}$$

$$\hat{\varepsilon}_{t+1|t} = \mathbf{y}_{t+1} - \mathbf{A}'_{s_{t+1}}\mathbf{x}_{t+1} - \mathbf{H}'_{s_{t+1}}\hat{\xi}_{t+1|t}$$

$$\hat{\xi}_{t+1|t+1} = \hat{\xi}_{t+1|t} + \{\mathbf{P}_{t+1|t}\mathbf{H}_{s_{t+1}} (\mathbf{H}'_{s_{t+1}}\mathbf{P}_{t+1|t}\mathbf{H}_{s_{t+1}} + \mathbf{R}_{s_{t+1}})^{-1}\hat{\varepsilon}_{t+1|t}\}$$

$$\xi_T|\mathbf{Y},\mathbf{X},\theta_1,\theta_2,\theta_3,\theta_4 \sim N(\hat{\xi}_{T|T},\mathbf{P}_{T|T})$$

$$\xi_t|\xi_{t+1},\mathbf{Y},\mathbf{X},\theta_1,\theta_2,\theta_3,\theta_4 \sim N(\xi^*_{t|t},\mathbf{P}^*_{t|t})$$

$$\mathbf{J}_t = \mathbf{P}_{t|t}\mathbf{F}'_{s_t}\mathbf{P}^{-1}_{t+1|t}$$

$$\mathbf{P}^*_{t|t} = \mathbf{P}_{t|t} - \mathbf{J}_t\mathbf{F}_{s_t}\mathbf{P}_{t|t}$$