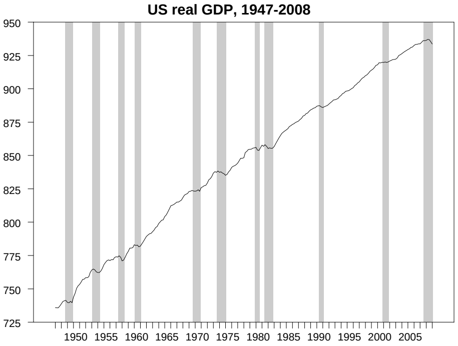
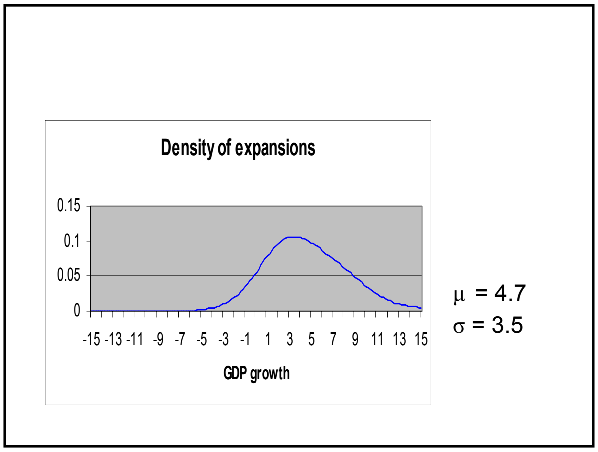
IV. Markov-switching models

A. Introduction to Markov-switching models

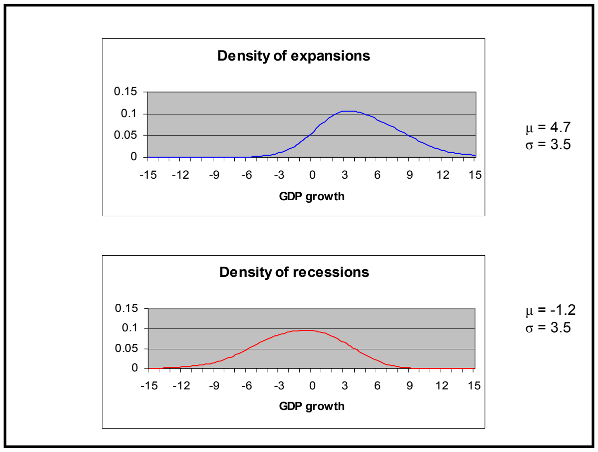
Measured as 100 times natural log
(means vertical change of 10 is approximately a 10% increase)



cumulative drop, 2007:Q3 to 2009:Q1 = 2.4%

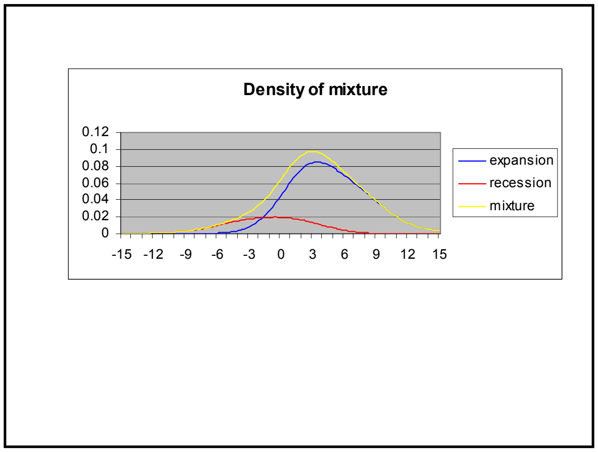


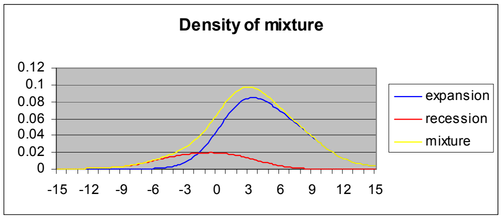
μ = 4.7
σ = 3.5



μ = 4.7
σ = 3.5

μ = -1.2
σ = 3.5



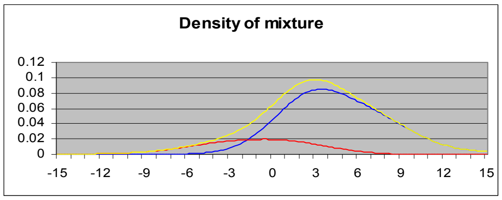


Density of mixture

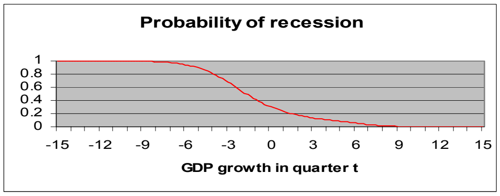
$$\Pr(S_t = 2 \mid y_t) = \frac{\Pr(S_t = 2, y_t)}{f(y_t)}$$

$$= \frac{\Pr(S_t = 2, y_t)}{\Pr(S_t = 1, y_t) + \Pr(S_t = 2, y_t)}$$

$$\Pr(S_t = 2, y_t) = \Pr(S_t = 2) \cdot f(y_t \mid S_t = 2)$$



Density of mixture



Probability of recession

Prob(NBER expansion in *t* + 1| NBER expansion in *t*)
= 164/174 = 0.94
Prob(NBER contraction in *t* +1 | NBER contraction in *t*)
= 35/44 = 0.80

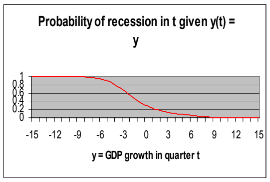
$$\Pr(S_{t+1}=2 \mid y_{t+1}, y_t) = \frac{\Pr(S_{t+1}=2, y_{t+1} \mid y_t)}{f(y_{t+1} \mid y_t)}$$

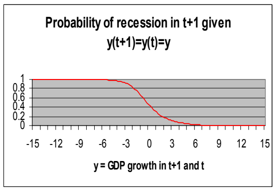
$$= \frac{\Pr(S_{t+1}=2, y_{t+1} \mid y_t)}{\Pr(S_{t+1}=1, y_{t+1} \mid y_t) + \Pr(S_{t+1}=2, y_{t+1} \mid y_t)}$$

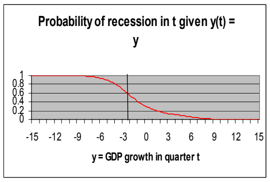
$$\Pr(S_{t+1}=2, y_{t+1} \mid y_t) = \sum_{s=1}^{2} \Pr(S_{t+1}=2, S_t = s, y_{t+1} \mid y_t)$$

$$\Pr(S_{t+1}=2, S_t = s, y_{t+1} \mid y_t) =$$

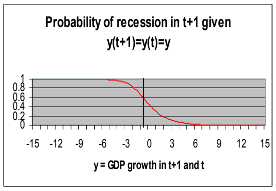
$$\underbrace{\Pr(S_t = s \mid y_t)}_{\substack{\text{known from}\\ \text{previous slide}}} \cdot \underbrace{\Pr(S_{t+1}=2 \mid S_t = s, y_t)}_{\substack{0.80 \text{ for s} = 2\\ 0.06 \text{ for s} = 1}} \cdot \underbrace{f(y_{t+1} \mid S_{t+1}=2, S_t = s, y_t)}_{\substack{\text{known from}\\ \text{density of recessions}}}$$


Probability of recession in t given y(t) = y


Probability of recession in t+1 given y(t+1)=y(t)=y


Probability of recession in t given y(t) = y

Prob = 0.6
at y = -2.2


Probability of recession in t+1 given y(t+1)=y(t)=y

Prob = 0.6
At y = -0.8

Can also calculate “smoothed probability”

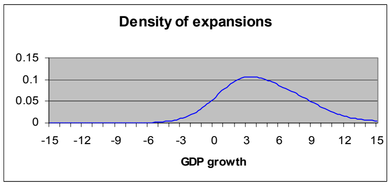
$$\Pr(S_t = 2 \mid y_t, y_{t+1})$$

Nonparametric summary of NBER labels:

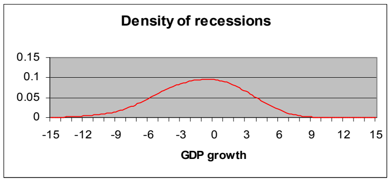
$f(y_t \mid S_t = 1)$ estimated nonparametrically

$$\Pr(S_{t+1} = 1 \mid S_t = 1) = 0.94$$

$$\Pr(S_{t+1} = 2 \mid S_t = 2) = 0.80$$


Density of expansions

$\mu = 4.7$
$\sigma = 3.5$


Density of recessions

$\mu = -1.2$
$\sigma = 3.5$

Could also approach parametrically:

$$f(y_t \mid S_t = 1) \sim N(\mu_1, \sigma^2)$$

$$f(y_t \mid S_t = 2) \sim N(\mu_2, \sigma^2)$$

$$\Pr(S_{t+1} = 1 \mid S_t = 1) = p_{11}$$

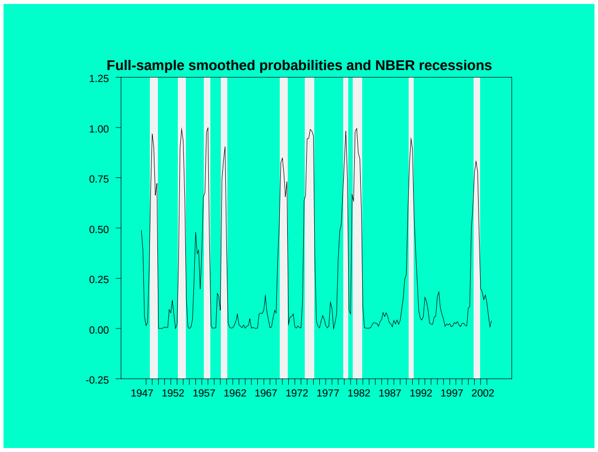
$$\Pr(S_{t+1} = 2 \mid S_t = 2) = p_{22}$$

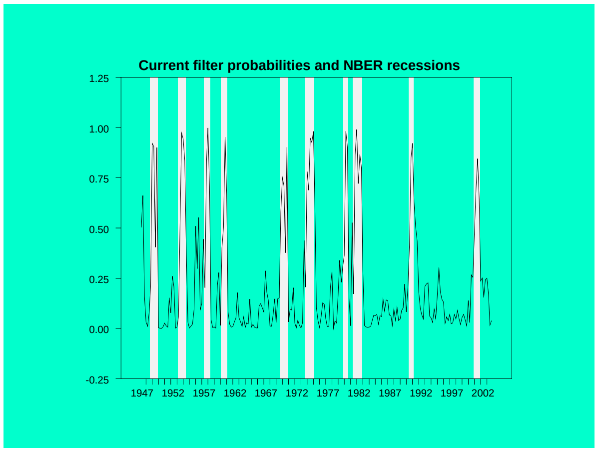
$$\Pr(S_{t+1} = 2, S_t = s, y_{t+1} \mid y_t) =$$

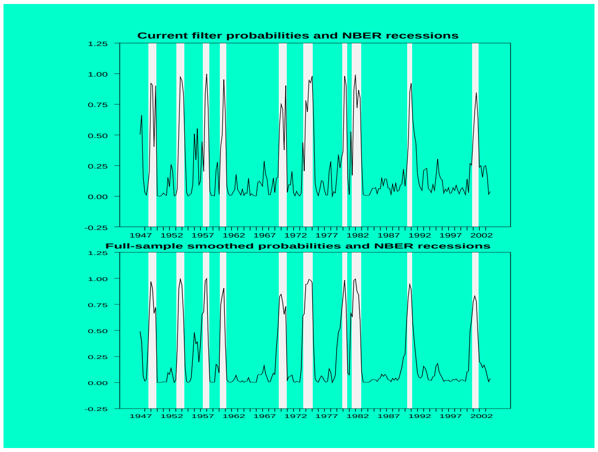
$$\underbrace{\Pr(S_t = s \mid y_t)}_{\substack{\text{known from}\\ \text{previous step}}} \cdot \underbrace{\Pr(S_{t+1} = 2 \mid S_t = s, y_t)}_{\substack{\text{p22 for s = 2}\\ \text{p11 for s = 1}}} \cdot \underbrace{f(y_{t+1} \mid S_{t+1} = 2, S_t = s, y_t)}_{N(\mu_2, \sigma^2)}$$

$$\begin{aligned} &f(y_{t+1} \mid y_t) = \\ &\Pr(S_{t+1} = 1, S_t = 1, y_{t+1} \mid y_t) + \\ &\Pr(S_{t+1} = 1, S_t = 2, y_{t+1} \mid y_t) + \\ &\Pr(S_{t+1} = 2, S_t = 1, y_{t+1} \mid y_t) + \\ &\Pr(S_{t+1} = 2, S_t = 2, y_{t+1} \mid y_t) \end{aligned}$$

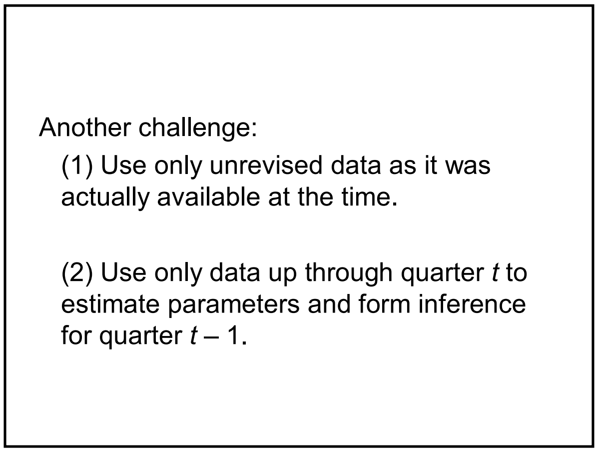
Parameter	Estimated from NBER dates and GDP	Estimated from GDP alone
μ_1	4.73	4.62
μ_2	-1.19	-0.57
σ	3.47	3.35
p_{11}	0.94	0.92
p_{22}	0.80	0.74


Full-sample smoothed probabilities and NBER recessions


Current filter probabilities and NBER recessions


Current filter probabilities and NBER recessions

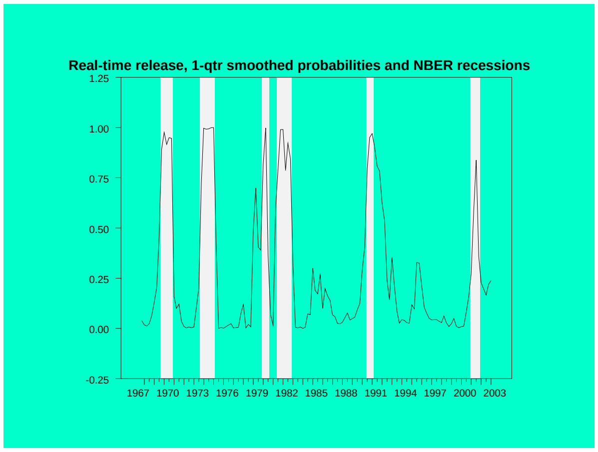
Full-sample smoothed probabilities and NBER recessions



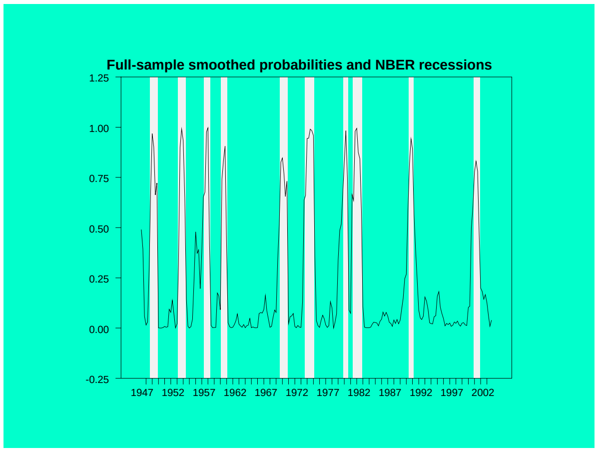
Another challenge:

(1) Use only unrevised data as it was actually available at the time.

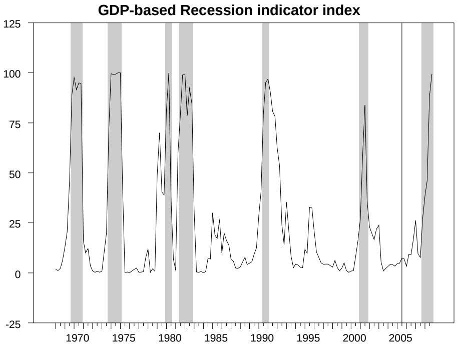
(2) Use only data up through quarter *t* to estimate parameters and form inference for quarter *t* – 1.

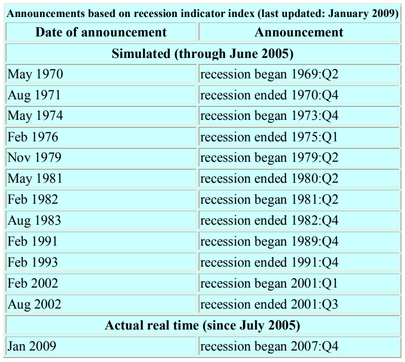


Real-time release, 1-qtr smoothed probabilities and NBER recessions

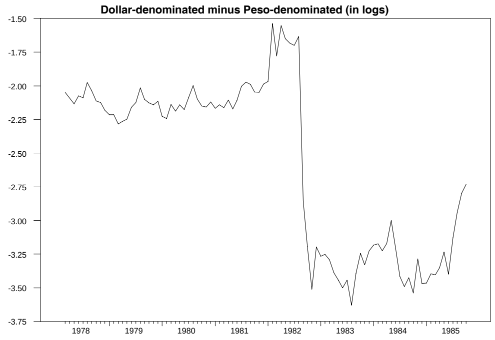


Full-sample smoothed probabilities and NBER recessions





Announcements based on recession indicator index (last updated: January 2009)	
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 2009	recession began 2007:Q4



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 $\{y_t\}$?

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

Need: probability law for s_t^*

Markov chain:

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

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 $\mathbf{P} = [p_{ji}]$

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

Let $\boldsymbol{\xi}_t = \mathbf{e}_i$ (the ith column of $\mathbf{I}_N$) when $s_t = i$. Then

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

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

$$= \mathbf{P}\boldsymbol{\xi}_t$$

More generally, 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

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

Then

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

(e.g., row j states that

$$\begin{aligned} &P(s_{t+1} = j|\Omega_t) \\ &\quad = p_{1j}P(s_t = 1|\Omega_t) + p_{2j}P(s_t = 2|\Omega_t) \\ &\qquad + \cdots + p_{Nj}P(s_t = N|\Omega_t) \end{aligned}$$

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 $s_{t-1}^* = 2$

$\boldsymbol{\xi}_t = \mathbf{e}_i$ (ith column of $\mathbf{I}_4$) when $s_t = i$

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

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

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

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

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

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

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

the conditional likelihood of tth observation.

The result of dividing the jth element of $(\boldsymbol{\eta}_t \odot \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1})$ by the conditional likelihood is

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

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

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

Iterative algorithm similar to Kalman filter:

Input for step t:

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

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

Output for step t:

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

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

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

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

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

---

(2) Set $\hat{\boldsymbol{\xi}}_{0|0} = \boldsymbol{\rho}$, 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 λ_{11} and λ_{22} rather than p_{11} and p_{22} where

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

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

Changes in regime are characterized by differences between the different rows of $\boldsymbol{\eta}_t = [p(y_t|s_t = j,\Omega_{t-1};\boldsymbol{\theta})]_{j=1}^{N}$. In above example, this was Gaussian AR(1) with different constant terms when $s_t = 1,\ldots,4$. Could also have different AR coefficients, variances, non-Gaussian distributions, vector processes, etc.

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:

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

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. 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 $k = 0, 1, 2, \dots$)

Gibbs sampler:

$$\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2', \boldsymbol{\theta}_3', \boldsymbol{\theta}_4')'$$
$$\theta_1 = \sigma^{-2}$$
$$\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$$
$$\boldsymbol{\theta}_3 = (p_{11}, p_{22})'$$
$$\boldsymbol{\theta}_4 = (s_1, s_2, \dots, s_T)'$$

---

(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:

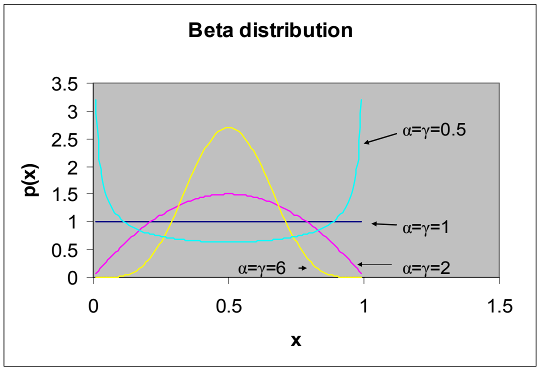
A variable x is said to have a beta distribution with parameters $\alpha > 0$ and $\gamma > 0$, denoted $x \sim \mathrm{Beta}(\alpha, \gamma)$, if

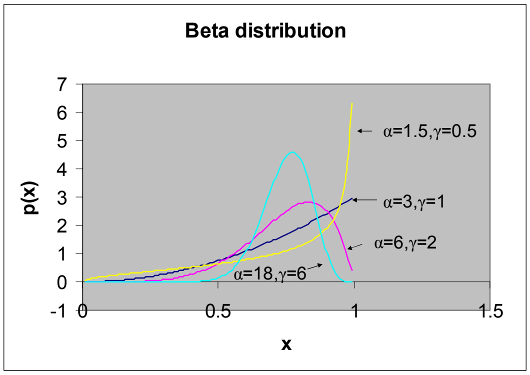
$$p(x|\alpha,\gamma) = \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\Gamma(\gamma)} x^{\alpha-1}(1-x)^{\gamma-1}$$

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:

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

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

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

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

$$\alpha_1^* = \alpha_1 + n_{11}$$
$$\gamma_1^* = \gamma_1 + n_{12}$$
$$\alpha_2^* = \alpha_2 + n_{22}$$
$$\gamma_2^* = \gamma_2 + n_{21}$$

---

(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

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

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

$$
\begin{aligned}
&P(S_t = i | S_{t+1} = j, \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t) \\
&\quad = \frac{P(S_t = i, S_{t+1} = j | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)}{P(S_{t+1} = j | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)} \\
&\quad = \frac{P(S_{t+1} = j | S_t = i) P(S_t = i | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)}{P(S_{t+1} = j | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \Omega_t)} \\
&\quad = \frac{p_{ij} \mathbf{e}_i' \hat{\boldsymbol{\xi}}_{t|t}}{\mathbf{e}_j' \mathbf{P} \hat{\boldsymbol{\xi}}_{t|t}}
\end{aligned}
$$

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:

$$\begin{aligned} \boldsymbol{\theta} &= (\theta_1, \boldsymbol{\theta}_2', \boldsymbol{\theta}_3', \boldsymbol{\theta}_4', \boldsymbol{\theta}_5')' \\ \theta_1 &= \sigma^{-2} \\ \boldsymbol{\theta}_2 &= (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')' \\ \boldsymbol{\theta}_3 &= (s_1, s_2, \ldots, s_T)' \\ \boldsymbol{\theta}_4 &= (z_1^*, z_2^*, \ldots, z_T^*)' \\ \boldsymbol{\theta}_5 &= (\gamma_0, \boldsymbol{\gamma}')' \end{aligned}$$

Draws from $p(\theta_1 | \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5, \mathbf{Y}, \mathbf{X})$ and $p(\boldsymbol{\theta}_2 | \theta_1, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5, \mathbf{Y}, \mathbf{X})$ same as before (conditioning on $\{z_t^*\}$ adds no information beyond that in $\{s_t\}$).

Draws of $\boldsymbol{\theta}_3 = (s_1, s_2, \ldots, s_T)'$ conditional on $\boldsymbol{\theta}_4 = (z_1^*, z_2^*, \ldots, z_T^*)'$ trivial from accounting identity:

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

Distribution of z_t^* conditional on $\{s_t, \mathbf{x}_t\}_{t=1}^T$ and $(\gamma_0, \boldsymbol{\gamma}')'$:

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

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

If $s_t = 1$, then z_t^* was a draw from a $N(\gamma_0 s_{t-1} + \boldsymbol{\gamma}'\mathbf{x}_t, 1)$ distribution that came out to be negative.

To generate such a truncated Normal variable, generate $N(\gamma_0 s_{t-1} + \boldsymbol{\gamma}'\mathbf{x}_t, 1)$ variables until one comes up negative, use this for z_t^*.

When $s_t = 2$, generate $N(\gamma_0 s_{t-1} + \boldsymbol{\gamma}'\mathbf{x}_t, 1)$ and take the first nonnegative value as simulated z_t^*.

Doing this for $t = 1, \ldots, T$ gives a draw from $p(\boldsymbol{\theta}_4 | \theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_5, \mathbf{Y}, \mathbf{X})$.

Finally, to generate a value for $\boldsymbol{\theta}_5 = (\gamma_0, \boldsymbol{\gamma}')'$ given $\theta_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_5, \mathbf{Y}, \mathbf{X}$, notice that conditional on having observed $\{s_{t-1}, z_t^*\}_{t=1}^T$, generating $\boldsymbol{\theta}_5$ 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. Bayesian analysis of Markov-switching models
C. 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

θ_1 = unknown elements of $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{R}_1, \mathbf{R}_2$

θ_2 = unknown elements of $\mathbf{F}_1, \mathbf{F}_2,$
$\mathbf{A}_1, \mathbf{A}_2, \mathbf{H}_1, \mathbf{H}_2$

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

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

θ_5 = unknown elements of $\{\xi_0, \xi_1, \ldots, \xi_T\}$

(1) Generating $\boldsymbol{\theta}_1$ $\{\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{R}_1, \mathbf{R}_2\}$ given $\{\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\}$.

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}_{s_t}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} \qquad E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}') = \mathbf{Q}_{s_t}$$

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

$$p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5) \propto \prod_{t=1}^{T} |\mathbf{Q}_{s_{t-1}}|^{-1/2}$$
$$\exp\left\{-(1/2)\sum_{t=1}^{T}(\boldsymbol{\xi}_t - \mathbf{F}_{s_{t-1}}\boldsymbol{\xi}_{t-1})'\right.$$
$$\left.\mathbf{Q}_{s_{t-1}}^{-1}(\boldsymbol{\xi}_t - \mathbf{F}_{s_{t-1}}\boldsymbol{\xi}_{t-1})\right\}$$
$$\propto \prod_{t=1}^{T} |\mathbf{Q}_1|^{-(1/2)\delta_{s_{t-1}=1}}$$
$$\exp\left\{-(1/2)\sum_{t=1}^{T}(\boldsymbol{\xi}_t - \mathbf{F}_{s_{t-1}}\boldsymbol{\xi}_{t-1})'\right.$$
$$\left.\mathbf{Q}_1^{-1}(\boldsymbol{\xi}_t - \mathbf{F}_{s_{t-1}}\boldsymbol{\xi}_{t-1})\delta_{s_{t-1}=1}\right\}$$

prior:

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

posterior:

$$\mathbf{Q}_1^{-1}|\boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5, \mathbf{Y}, \mathbf{X}$$
$$\sim W(N_{Q_1} + T_1, \boldsymbol{\Lambda}_{Q_1} + \mathbf{S}_{Q_1})$$
$$T_1 = \sum_{t=1}^{T} \delta_{s_{t-1}=1}$$
$$\mathbf{S}_{Q_1} = \sum_{t=1}^{T} \mathbf{v}_t\mathbf{v}_t'\delta_{s_{t-1}=1}$$
$$\mathbf{v}_t = \boldsymbol{\xi}_t - \mathbf{F}_{s_{t-1}}\boldsymbol{\xi}_{t-1}$$

(2) Generating θ_2 $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{A}_1, \mathbf{A}_2, \mathbf{H}_1, \mathbf{H}_2\}$ given $\{\mathbf{Y}, \mathbf{X}, \theta_1, \theta_3, \theta_4, \theta_5\}$.

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}_{s_t}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} \qquad E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}^{'}) = \mathbf{Q}_{s_t}$$

---

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

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

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

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

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

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

$$\widehat{\mathbf{F}_2^{\prime}} = \left(\sum_{t=1}^{T} \boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t-1}^{'}\delta_{s_{t-1}=2}\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t}^{'}\delta_{s_{t-1}=2}\right)$$

(3) Generating $\boldsymbol{\theta}_3 = (p_{11}, p_{22})'$ given $\{\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\}$.

(4) Generating $\boldsymbol{\theta}_4 = (s_0, s_1, s_2, \ldots, s_T)'$ given $\{\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_5\}$.

Exactly same as for other Markov-switching models.

---

(5) Generating $\boldsymbol{\theta}_5$ $\{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T\}$ given $\{\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4\}$.

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}_{s_t}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} \qquad E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}') = \mathbf{Q}_{s_t}$$

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

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

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

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

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

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}}_{t+1|t} &= \mathbf{y}_{t+1} - \mathbf{A}'_{s_{t+1}}\mathbf{x}_{t+1} - \mathbf{H}'_{s_{t+1}}\hat{\boldsymbol{\xi}}_{t+1|t} \\ \hat{\boldsymbol{\xi}}_{t+1|t+1} &= \hat{\boldsymbol{\xi}}_{t+1|t} + \{\mathbf{P}_{t+1|t}\mathbf{H}_{s_{t+1}} \\ &\qquad (\mathbf{H}'_{s_{t+1}}\mathbf{P}_{t+1|t}\mathbf{H}_{s_{t+1}} + \mathbf{R}_{s_{t+1}})^{-1}\hat{\boldsymbol{\varepsilon}}_{t+1|t}\}\end{aligned}$$

$$\begin{aligned}&\boldsymbol{\xi}_T|\mathbf{Y},\mathbf{X},\theta_1,\theta_2,\theta_3,\theta_4 \sim N(\hat{\boldsymbol{\xi}}_{T|T},\mathbf{P}_{T|T}) \\ &\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1},\mathbf{Y},\mathbf{X},\theta_1,\theta_2,\theta_3,\theta_4 \sim N(\boldsymbol{\xi}^*_{t|t},\mathbf{P}^*_{t|t}) \\ &\mathbf{J}_t = \mathbf{P}_{t|t}\mathbf{F}'_{s_t}\mathbf{P}^{-1}_{t+1|t} \\ &\mathbf{P}^*_{t|t} = \mathbf{P}_{t|t} - \mathbf{J}_t\mathbf{F}_{s_t}\mathbf{P}_{t|t}\end{aligned}$$