

Bayesian estimation of Markov Switching models

Gianni Amisano (ECB, U Brescia)

Frankfurt, 26 March 2010

The Markov Switching model

- Markov Switching (MS) model (Lindgren, 1978, Hamilton, 1994, Kim and Nelson, 1999)

$$\begin{aligned}
 p(y_t | s_t) &= N(\mathbf{x}'_t \boldsymbol{\beta}_{s_t}, h_{s_t}^{-1}) \\
 p(s_t = j | s_{t-1} = i, \mathbf{s}_{t-2}, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}) &= \\
 &= p(s_t = j | s_{t-1} = i) = p_{ij}
 \end{aligned}$$

- observables: y_t, \mathbf{x}_t
- parameters: $\theta = [\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_m, h_1, \dots, h_m, p_{11}, \dots, p_{1m}, \dots, p_{m1}, \dots, p_{mm}]'$

An example (I)

- Hamilton (1989), univariate business cycle model

$$\begin{aligned}\phi(L) \left(\Delta y_t - \mu_{s_t} \right) &= \sigma_{s_t} e_t \\ \phi(L) &= 1 - \phi_1 L - \dots - \phi_p L^p\end{aligned}$$

- asymmetric model

An example (II): mixture model for financial returns

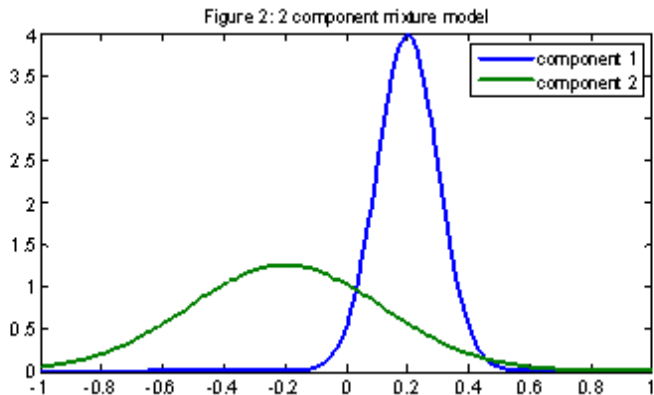
- Example of MS model

$$(y_t | s_t = 1) \sim N(0.02, 0.01) \text{ (high returns, low volatility)}$$

$$(y_t | s_t = 2) \sim N(-0.02, 0.1) \text{ (low returns, high volatility)}$$

$$p_{11} = .99, p_{22} = .8$$

An example (II): mixture model for financial returns, cont'd



Inferential problems

We can define

- Extended likelihood (conditional on latent variables)

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{s})$$

most of the time complete likelihood has very simple form

- Likelihood

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{s})p(\mathbf{s}|\boldsymbol{\theta})d\mathbf{s} = \sum_{s_i \in \mathcal{S}} p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{s}_i)p(\mathbf{s}_i|\boldsymbol{\theta})$$

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{s}) = \text{Extended likelihood}$$

$$p(\mathbf{s}|\boldsymbol{\theta}) = \text{law of motion of latent states}$$

- NEED A CLEVER WAY TO DO THE INTEGRATION, otherwise summation over m^T possibilities!
- Bayesian approach: might be useful to condition on latent variables to make inference on $\boldsymbol{\theta}$

Recap: Bayesian inference on latent variable models

- 2 approaches
 - ① integrate latent variables out and form likelihood $p(\mathbf{y}|\boldsymbol{\theta})$
 - ② simulate also the latent variable and construct MCMC which simulates also latent variables: data augmentation

Data augmentation (Gibbs with latent variables): I

- Treat latent variables in the same way as parameters. Focus on

$$p(\boldsymbol{\theta})p(\mathbf{s}|\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{s})$$

as a "joint posterior" and use a Gibbs Sampling algorithm (data augmentation)

- In all models with latent variables we have 3 groups of objects
 - 1 data \mathbf{y} (observable)
 - 2 parameters $\boldsymbol{\theta}$ (unobservable)
 - 3 latent variables \mathbf{z} (unobservable)

Data augmentation (Gibbs with latent variables): II

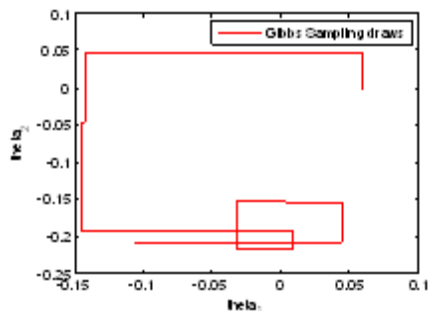
- Parameters and latent variables are in turn simulated from their conditional posterior distributions

$$p(\boldsymbol{\theta}|\mathbf{s}, \mathbf{y})$$

$$p(\mathbf{s}|\boldsymbol{\theta}, \mathbf{y})$$

this a MC converging to joint posterior distribution of $\boldsymbol{\theta}$ and \mathbf{s}

Data augmentation (Gibbs with latent variables): III



Gibbs Sampling Draws in a 2 block
example (θ_1 and θ_2)

The MS univariate linear regression model (ULRM)

- Assume 2 state MS model

$$\begin{aligned}
 p(y_t | s_t) &= N(\mathbf{x}'_t \boldsymbol{\beta}_{s_t}, h_{s_t}^{-1}) \\
 p(s_t = j | s_{t-1} = i, \mathbf{s}_{t-2}, \mathbf{y}_{t-1}, \mathbf{x}_{t-1}) &= \\
 &= p(s_t = j | s_{t-1} = i) = p_{ij}, i, j = 1, 2 \\
 \mathbf{P} &= \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}
 \end{aligned}$$

- MS linear univariate regression

The MS univariate linear regression model (ULRM)

II

- parameters:

$$\boldsymbol{\theta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ h_1 \\ h_2 \\ p_{11} \\ p_{22} \end{bmatrix}$$

- p_{ij} bounded between 0 and 1. Which prior do we specify for these two parameters?

The Beta distribution: I

- Use Beta distribution:

$$p(p_{ii} | r_{i1}, r_{i2}) = \frac{p_{ii}^{r_{i1}-1} (1 - p_{ii})^{r_{i2}-1}}{B(r_{i1}, r_{i2})}$$

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

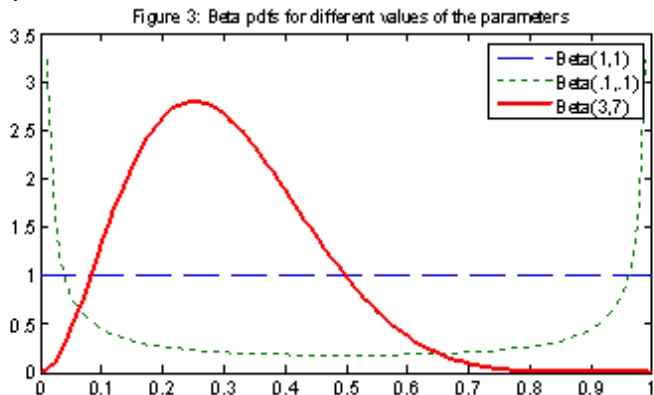
$$E(p_{ii}) = \frac{r_{i1}}{r_{i1} + r_{i2}}$$

$$V(x) = \frac{r_{i1} r_{i2}}{(r_{i1} + r_{i2})^2 (r_{i1} + r_{i2} + 1)}$$

$$\text{Mode}(x) = \frac{r_{i1} - 1}{(r_{i1} + r_{i2} - 2)}$$

The Beta distribution: II

Examples of beta distributions



Simulation strategy

In order to apply Gibbs Sampling, we describe now conditional posterior distributions

- of parameters given states (we can partition the parameter vector further)
- of states given parameters

The regression parameters: I

$$p(\beta_i | \beta_j, h_i, h_j, \underline{s}_T, \mathbf{y})$$

$$p(h_i | \beta_i, \beta_j, h_j, \underline{s}_T, \mathbf{y})$$

- Conditional on data and latent variables, in the extended likelihood parameters β_i, h_i appear only in those observations allocated to state i .
- Therefore use only subsample of observations allocated to state i .
- Same results as in ULRM, ie β_i distributed as Gaussians and h_i distributed as Gammas

The regression parameters: II

Example: $T = 4$

If we condition on vector of states variables

$$\underline{s}_T = [1 \quad 1 \quad 2 \quad 2]$$

\Rightarrow only observations 1 and 2 will contribute to determine conditional posterior of β_1 and h_1 .

\Rightarrow only observations 3 and 4 will contribute to determine conditional posterior of β_2 and h_2 .

The transition probabilities: I

- Conditional on data, latent variables and other parameters, in the "joint" posterior distribution

$$p(\boldsymbol{\theta})p(\underline{\mathbf{s}}_T|\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}, \underline{\mathbf{s}}_T)$$

- we note that

$$\begin{aligned} p(\boldsymbol{\theta}) &= p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_1, h_2)p(\mathbf{P}) \\ p(\underline{\mathbf{s}}_T|\boldsymbol{\theta}) &= p(\underline{\mathbf{s}}_T|\mathbf{P}) \\ p(\mathbf{y}|\boldsymbol{\theta}, \underline{\mathbf{s}}_T) &= p(\mathbf{y}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_1, h_2, \underline{\mathbf{s}}_T) \end{aligned}$$

The transition probabilities: II

- Hence conditional on latent states, data does not give additional information on transition probabilities and conditional posterior of free elements of \mathbf{P} is given by product

$$\begin{aligned}
 & p(\mathbf{P}) \times p(\underline{\mathbf{s}}_T | \mathbf{P}) \\
 \propto & p_{11}^{r_{11}-1} (1 - p_{11})^{r_{12}-1} \times p_{11}^{T_{11}} (1 - p_{11})^{T_{12}} \times \\
 & p_{22}^{r_{21}-1} (1 - p_{22})^{r_{22}-1} \times p_{22}^{T_{22}} (1 - p_{22})^{T_{21}}
 \end{aligned}$$

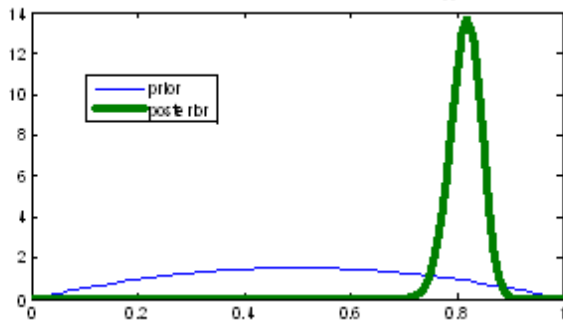
- where T_{ij} is number of times transition from state i to state j is observed in the state vector $\underline{\mathbf{s}}_T$.

$\Rightarrow p_{ij}$ is $Beta(r_{i1} + T_{i1}, r_{i2} + T_{i2})$ distributed.

The transition probabilities: III

Example: suppose \underline{s}_T such that $T_{11} = 140$, $T_{12} = 30$ and $r_{11} = 2$, $r_{12} = 2$

Figure 4: prior and conditional posterior for p_1



The cond. posterior distribution of states, I

Here object of interest is

$$p(\underline{\mathbf{s}}_T | \underline{\mathbf{y}}_T, \boldsymbol{\theta})$$

from which we want to draw

The cond. posterior distribution of states, II

Best way to do this: use Multi-move GS (Carter-Kohn, 1994, Fruewirth-Schnatter, 1994, Chib, 1996): draw whole sequence from $p(\underline{\mathbf{s}}_T | \underline{\mathbf{y}}_T, \boldsymbol{\theta})$ using backward (in time) sequential partition

$$\begin{aligned} \underline{\mathbf{s}}_T &\sim p(\underline{\mathbf{s}}_T | \underline{\mathbf{y}}_T, \boldsymbol{\theta}) \\ p(\underline{\mathbf{s}}_T | \underline{\mathbf{y}}_T, \boldsymbol{\theta}) &= p(s_T | \underline{\mathbf{y}}_T, \boldsymbol{\theta}) \times \\ &\quad \times \prod_{t=1}^{T-1} p(s_t | s_{t+1}, \dots, s_T, \underline{\mathbf{y}}_T, \boldsymbol{\theta}) \end{aligned}$$

The cond. posterior distribution of states, III

- Given Markov property of latent variables, typical element in the recursion is:

$$p(s_t | s_{t+1}, \dots, s_T, \underline{\mathbf{y}}_T, \boldsymbol{\theta}) = p(s_t | s_{t+1}, \underline{\mathbf{y}}_t, \boldsymbol{\theta})$$

- (need only to know s_{t+1} and data history up to time t).

The cond. posterior distribution of states, IV

- Using Bayes' Theorem

$$\begin{aligned}
 p(s_t = i | s_{t+1} = j, \underline{\mathbf{y}}_t, \boldsymbol{\theta}) &= \frac{p(s_t = i, s_{t+1} = j, \underline{\mathbf{y}}_t, \boldsymbol{\theta})}{p(s_{t+1} = j, \underline{\mathbf{y}}_t, \boldsymbol{\theta})} = \\
 &= \frac{p(s_{t+1} = j | s_t = i, \boldsymbol{\theta}) \times p(s_t = i | \underline{\mathbf{y}}_t, \boldsymbol{\theta})}{p(s_{t+1} = j, \underline{\mathbf{y}}_t, \boldsymbol{\theta})} \\
 &= \frac{p_{ij} \times \pi_{i,t|t}}{\pi_{j,t+1|t}}
 \end{aligned}$$

- where

$$\begin{aligned}
 \pi_{i,t|t} &= p(s_t = i | \underline{\mathbf{y}}_t, \boldsymbol{\theta}) \text{ filtered prob} \\
 \pi_{j,t+1|t} &= p(s_{t+1} = j, \underline{\mathbf{y}}_t, \boldsymbol{\theta}) \text{ proj prob}
 \end{aligned}$$

- need filtering!

Filtering in MS models, I

- Filtering delivers probabilities of latent variables at each time t given sample information up to this point.
- As we have seen before filtering also yields likelihood of the model.
- Filtering entails integrating latent variables out. In the MS model, since latent variable has discrete support and integration means just summing over finite number of states, filtering is analytically feasible and very easy.

Filtering in MS models, II

Filtering is carried out with following algorithm

- start from initialisation at $t = 0$

$$\begin{aligned} p(s_0 = i | \underline{\mathbf{y}}_0, \boldsymbol{\theta}) &= p(s_0 = i | \boldsymbol{\theta}) = \pi_{i,0|0}, \\ i &= 1, 2, \dots, m \end{aligned}$$

- for $t = 0 : T - 1$ iterate the following steps (see next page)

Filtering in MS models, III

1) Prediction (predict following state on the grounds of current sample info)

$$\begin{aligned}
 \pi_{j,t+1|t} &= p(s_{t+1} = j | \underline{\mathbf{y}}_t, \boldsymbol{\theta}) = \sum_{i=1}^m p(s_{t+1} = j, s_t = i | \underline{\mathbf{y}}_t, \boldsymbol{\theta}) \\
 &= \sum_{i=1}^m p(s_{t+1} = j | s_t = i, \underline{\mathbf{y}}_t, \boldsymbol{\theta}) \times p(s_t = i | \underline{\mathbf{y}}_t, \boldsymbol{\theta}) = \\
 &= \sum_{i=1}^m p_{ij} \times \pi_{i,t|t}
 \end{aligned}$$

Filtering in MS models, IV

2) Update: use data information y_{t+1} to update predicted probabilities (Bayes theorem)

$$\begin{aligned}
 \pi_{j,t+1|t+1} &= p(s_{t+1} = j | \underline{\mathbf{y}}_{t+1}, \boldsymbol{\theta}) = \\
 &= p(s_{t+1} = j | \underline{\mathbf{y}}_t, y_{t+1}, \boldsymbol{\theta}) \\
 &= \frac{\pi_{j,t+1|t} \times p(y_{t+1} | s_{t+1} = j, \boldsymbol{\theta})}{p(y_{t+1} | \underline{\mathbf{y}}_t, \boldsymbol{\theta})}
 \end{aligned}$$

$$p(y_{t+1} | \underline{\mathbf{y}}_t, \boldsymbol{\theta}) = \sum_{i=1}^m \left[\pi_{j,t+1|t} \times p(y_{t+1} | s_{t+1} = j, \boldsymbol{\theta}) \right]$$

Filtering in MS models, a numerical example, I

Example: suppose

$$(y_t | s_t = 1) \sim N(0.02, 0.01) \text{ (high returns, low volatility)}$$

$$(y_t | s_t = 2) \sim N(-0.02, 0.1) \text{ (low returns, high volatility)}$$

$$p_{11} = .99, p_{22} = .8$$

and

$$p(s_{t-1} = 1 | \underline{y}_{t-1}) = 0.5$$

$$y_t = 0.05$$

Filtering in MS models, a numerical example, II

1 projection:

$$\begin{aligned} p(s_t = 1 | \underline{\mathbf{y}}_{t-1}) &= \sum_{i=1}^2 p_{i1} \times p(s_{t-1} = i | \underline{\mathbf{y}}_{t-1}) = \\ &= 0.99 \times 0.5 + 0.2 \times 0.5 = 0.595 \end{aligned}$$

2 update:

$$\begin{aligned} p(s_t = 1 | \underline{\mathbf{y}}_t) &= \frac{p(y_t | s_t = 1, \underline{\mathbf{y}}_{t-1}) \times p(s_t = 1 | \underline{\mathbf{y}}_{t-1})}{\sum_{i=1}^2 p(y_t | s_t = i, \underline{\mathbf{y}}_{t-1}) \times p(s_t = i | \underline{\mathbf{y}}_{t-1})} = \\ &= \frac{3.8139 \times 0.595}{3.8139 \times 0.595 + 1.2310 \times 0.405} = 0.8199 \end{aligned}$$

Some remarks I

- 1 Denominator in update equation is likelihood of y_{t+1} conditioned on past observables only (beside parameters), typical term for computing likelihood

$$p(\underline{\mathbf{y}}_T | \boldsymbol{\theta}) = \prod_{t=1}^T p(y_t | \underline{\mathbf{y}}_{t-1}, \boldsymbol{\theta})$$

- 2 With multivariate model (i.e. \mathbf{y}_t a vector) we use same formulae.
- 3 Initialisation of the filter using ergodic distribution of s_t i.e. setting $\pi_{j,0|0}$ equal to $p(s_t = j | \boldsymbol{\theta}) = \pi_j$:

Some remarks II, continued

$$p(s_t = j | \boldsymbol{\theta}) = \pi_j = \sum_{i=1}^m p_{ij} \times \pi_i$$

$$\Leftrightarrow \boldsymbol{\pi}' = \boldsymbol{\pi}' \mathbf{P}$$

when $m=2$:

$$\pi_1 = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \quad \pi_2 = \frac{p_{22}}{2 - p_{11} - p_{22}}$$

Summing up

To simulate joint posterior distribution of latent variables and parameters in a MS model, we can use a GS algorithm which works as follows

- ① initialise by drawing parameters from arbitrary distribution (eg prior);
- ② run filter and backward recursion to compute probabilities $p(s_t = i | s_{t+1} = j, \underline{\mathbf{y}}_t, \boldsymbol{\theta})$;
- ③ use these probabilities to draw latent variables;
- ④ conditional on latent variables draw parameters (in separate blocks β_i, h_i , each row of \mathbf{P});

-cycle over steps (2) to (4).

The label switching problem I

- Take simple MS model (no covariates, only an intercept term)

$$p(y_t | s_t, \boldsymbol{\theta}) = N(\mu_{s_t}, h_{s_t}^{-1})$$

$$\mathbf{P} = \{p(s_t = j | s_{t-1} = i, \boldsymbol{\theta})\}$$

- If permute labelling of the states (ie you call state 2 what you called state 1 before), the likelihood of the model does not change. So we have an identification problem.

The label switching problem II

- Numeric example

$$\begin{aligned}
 p(y_t | s_t = 1, \theta) &= N(1, 1) \\
 p(y_t | s_t = 2, \theta) &= N(-1, 10) \\
 \mathbf{P} &= \begin{bmatrix} .9 & .1 \\ .3 & .7 \end{bmatrix}
 \end{aligned}$$

the likelihood of this model is exactly the same as the likelihood of

$$\begin{aligned}
 p(y_t | s_t = 1, \theta) &= N(-1, 10) \\
 p(y_t | s_t = 2, \theta) &= N(1, 1) \\
 \mathbf{P} &= \begin{bmatrix} .7 & .3 \\ .1 & .9 \end{bmatrix}
 \end{aligned}$$

The label switching problem III

- Identification problem, related to the structural interpretation of the latent states, which can only be solved by imposing constraints.
- Examples of constraints (one of them is enough to achieve identification) are

$$\begin{aligned}\mu_1 &> \mu_2 \\ h_1 &> h_2 \\ p_{11} &> p_{22}\end{aligned}$$

The label switching problem IV

These constraints can be implemented in 2 different ways

- 1 draw from the relevant conditional posterior distribution that ignores the constraint until you finally satisfy it (works with any priors). For instance if constraint was on the mean, we draw from $p(\mu_1, \mu_2 | h_1, h_2, \mathbf{P}, \mathbf{y})$ until we get a draw satisfying the constraint.
- 2 Suppose we use a prior distribution which is symmetric across states, ie

$$\begin{aligned} p(h_1) &= p(h_2) \\ p(\mu_1) &= p(\mu_2) \\ p(p_{11}) &= p(p_{22}) \end{aligned}$$

we can draw from relevant conditional posteriors and then permute the order of the result to achieve the required ordering.

The label switching problem V

In other words, suppose the draw from the relevant conditional posteriors is

$$\begin{aligned}\mu_1 &= 0.1, \mu_2 = 0.2 \\ h_1 &= 1.5, h_2 = 2.5 \\ p_{11} &= .8, p_{22} = .9\end{aligned}$$

and we want to impose constraint on mean, then the draw would be permuted as

$$\begin{aligned}\mu_2 &= 0.1, \mu_1 = 0.2 \\ h_2 &= 1.5, h_1 = 2.5 \\ p_{22} &= .8, p_{11} = .9\end{aligned}$$

A MS Phillips Curve (I)

- Amisano and Giacomini, (2007).
- Phillips Curve (henceforth PC) =relationship in which inflation is related to some variable representing the level of real activity.
- In most cases this variable is the unemployment rate.
- Monthly data, period 1958:01-2004:07.
- CPI_t =CPIAUCSL, Consumer Price Index For All Urban Consumers: All Items.
- u_t =UNRATE, Civilian Unemployment Rate)
- Data were obtained from FRED[®] II (<http://research.stlouisfed.org/fred2/>) .

A MS Phillips Curve (II)

- Model:

$$\begin{aligned}\Delta\pi_t &= \alpha + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \\ &\quad + \beta_{12}\Delta\pi_{t-12} + \gamma_1 u_{t-1} + \\ +\sigma \cdot e_t &= \boldsymbol{\delta}' \mathbf{x}_t + \sigma \cdot e_t\end{aligned}$$

- \Rightarrow *NAIRU*

$$\Delta\pi = 0 \Rightarrow u = -\alpha/\gamma_1$$

A MS Phillips Curve (III)

Markov Switching (MS) specification:

$$\Delta\pi_t = \alpha^{s_t} + \beta_1^{s_t} \Delta\pi_{t-1} + \beta_2^{s_t} \Delta\pi_{t-2} + \beta_{12}^{s_t} \Delta\pi_{t-12} + \gamma_1^{s_t} u_{t-1} + \sigma^{s_t} \cdot e_t \quad (1)$$

$$= \delta^{s_t'} \mathbf{x}_t + \sigma^{s_t} \cdot e_t \quad (2)$$

$$s_t = \begin{cases} 1 \\ 2 \end{cases}, p(s_t = j | s_{t-1} = i) = p_{ij} \quad (3)$$

A MS Phillips Curve (IV)

We have estimated 2 variants of the MS model:

- Model MS1 is obtained by imposing on:

$$\sigma^{S_t} = \sigma$$

- MS2 is obtained by imposing same variance and:

$$\alpha^{S_t} = \alpha$$

$$\gamma_1^{S_t} = \gamma_1$$

Constancy of the NAIRU across states. In this way only different speeds of adjustment of the (changes of) inflation.

A MS Phillips Curve (V)

- MS2 model, as example
- M=55000 draws, discarded 10% of draws
- loose priors:

$$\underline{\mu}_{\beta} = \begin{bmatrix} 0.1 \\ -0.04 \\ \mathbf{0} \\ (3 \times 1) \\ \mathbf{0} \\ (3 \times 1) \end{bmatrix}, \underline{\mathbf{H}}_{\beta}^{-1} = (0.2)^2 \times \mathbf{I}_8$$

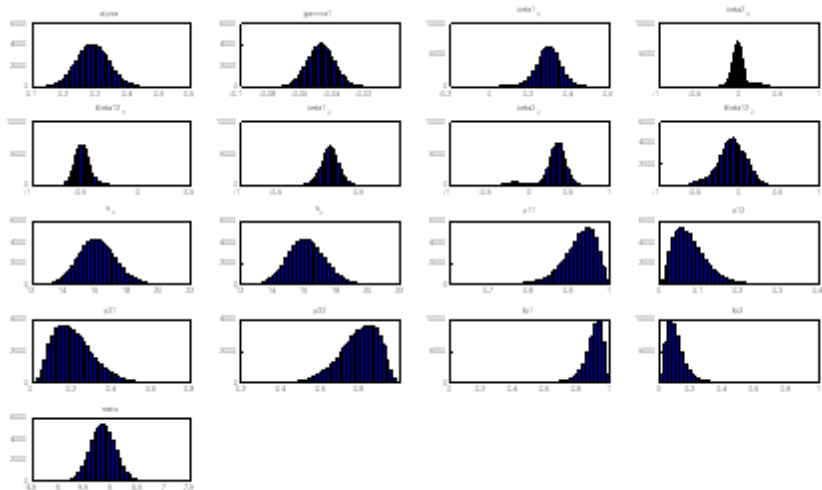
$$.3 \times h \sim \chi_{(3)}^2$$

$$r_{11} = 8, r_{12} = 2, r_{21} = 8, r_{22} = 2$$

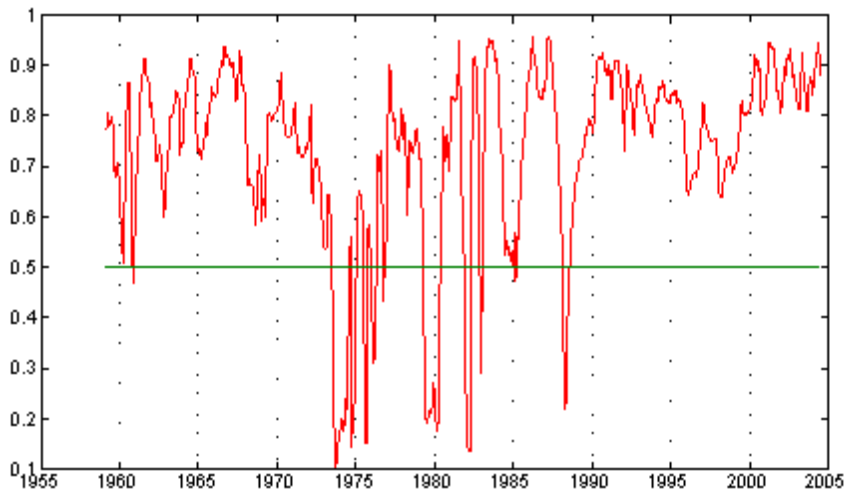
A MS Phillips Curve (VI)

	prior m.	prior sd	post m	post sd	low	up
α	.1	.2	0.29	0.05	0.19	0.39
γ	-.04	.2	-0.05	0.01	-0.07	-0.03
$\beta_{1,1}$	0	.2	0.29	0.07	0.14	0.41
$\beta_{2,1}$	0	.2	-0.01	0.10	-0.17	0.28
$\beta_{12,1}$	0	.2	-0.52	0.08	-0.65	-0.32
$\beta_{1,2}$	0	.2	0.12	0.12	-0.12	0.34
$\beta_{2,2}$	0	.2	0.30	0.18	-0.25	0.54
$\beta_{12,2}$	0	.2	-0.10	0.18	-0.50	0.23
p_{11}	.8	.12	0.92	0.04	0.82	0.98
p_{22}	.8	.12	0.79	0.10	0.56	0.94
nairu	-	-	5.84	0.24	5.36	6.32

A MS Phillips Curve (VII)



A MS Phillips Curve (VIII)



An Early Warning MS indicator for inflation, motivation

Amisano and Fagan, (2010).

- Role of money in predicting inflation dynamics
- nonlinear framework
- assess risk of transition to high inflation regime
- see whether money aggregate contain useful information
- highly stylised and simple model

EW-MS basics

- simple Markov Switching AR model for inflation
- two states, high and low inflation.
- transition probabilities are function of indicator variable (money growth)
- different specifications, univariate and multivariate
- investigate alternative parameterisations
- extend to panel data framework

EW-MS basics, I

- Money growth dynamics affect transition probabilities
- effect is sizeable and significant
- effects have signs corresponding to the intuition
- robust across specifications
- robust across countries (5 countries)
- robust to panel data extension

EW-MS basics, II

- MS-AR model for inflation

$$\pi_t = c_{s_t} + \phi_{s_t} \pi_{t-1} + \sigma_{s_t} e_t \quad (4)$$

$$e_t \sim NID(0, 1) \quad (5)$$

- s_t is MS with transition probabilities possibly depending on early indicator variables \mathbf{z}_{t-1} :

$$p(s_t = j | \mathbf{s}_{t-1}, \mathbf{y}_{t-1}, \boldsymbol{\theta}) = \quad (6)$$

$$= p(s_t = j | s_{t-1} = i, \mathbf{z}_{t-1}, \boldsymbol{\theta}) = p_{ij,t} \quad (7)$$

- alternatively

$$(1 - \phi_{s_t} L)(\pi_t - \mu_{s_t}) = \sigma_{s_t} e_t \quad (8)$$

EW-MS: Time varying transition probabilities (I)

- Assume (common slope): $\gamma_{21} = \gamma_{22} = \gamma_2$

$$p(s_t = 1 | s_{t-1} = i, I_{t-1}) = \Phi(\gamma_{1i} + \gamma_2 z_{t-1})$$

$$\Phi(\omega) = \int_{-\infty}^{\omega} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2}\right) d\mu$$

- Example with

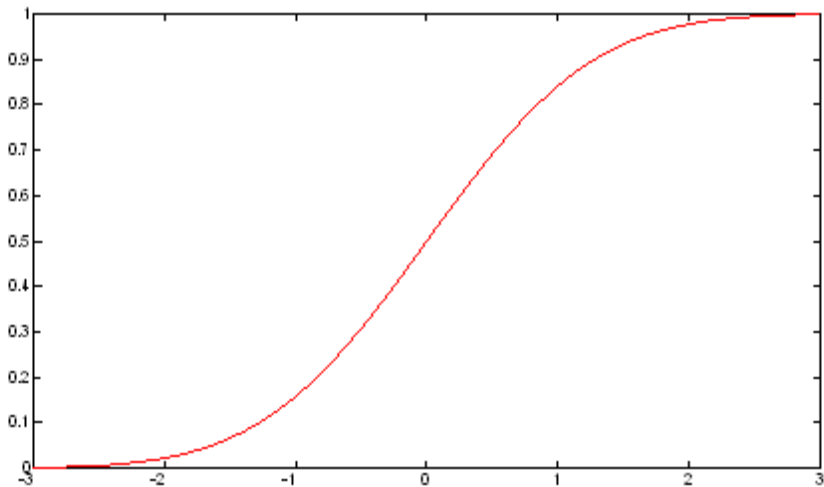
$$\gamma_{11} = .99, \gamma_{21} = \gamma_{22} = -.22, \gamma_{12} = -.49$$

- \Rightarrow :

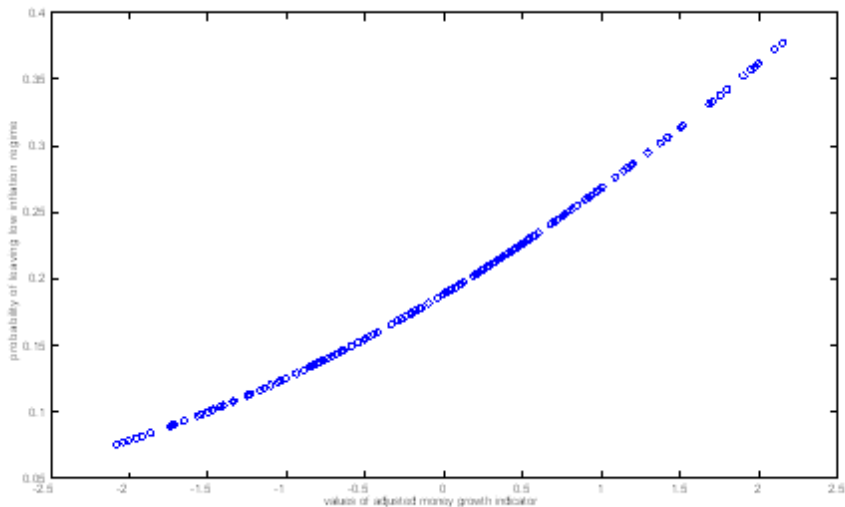
$$\mathbf{P}(z_{2t} = 0) = \begin{bmatrix} .81 & .19 \\ .33 & .67 \end{bmatrix}, \mathbf{P}(z_{2t} = -2.08) = \begin{bmatrix} .92 & .08 \\ .54 & .46 \end{bmatrix},$$

$$\mathbf{P}(z_{2t} = 2.15) = \begin{bmatrix} .62 & .38 \\ .16 & .84 \end{bmatrix}.$$

EW-MS: Time varying transition probabilities (II)



EW-MS: Time varying transition probabilities (III)



EW-MS: Bayesian estimation

- Data are observed on z_t, π_t
- Latent variable s_t
- parameters $\theta = [\phi, c_1, c_2, \sigma_1, \sigma_2, \text{vec}(\Gamma)']'$
- use Gibbs Sampling-data augmentation. See details in Amisano and Fagan, 2010.

EW-MS: results, I

- single country analysis: US, GE, UK, CA, EA from 1950s
- quarterly data on inflation and money indicator
- GE only up to inception of EMU
- EA: GE monetary aggregates up to 1998.

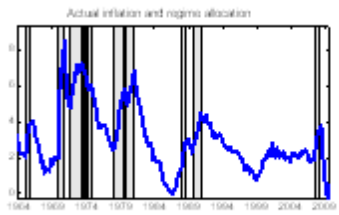
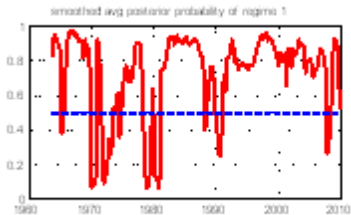
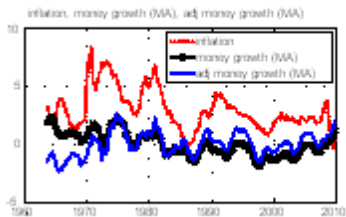
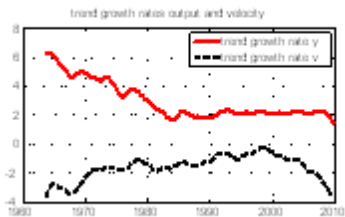
EW-MS: results, II

- slope coefficient is negative and significant
- elasticities of transition probabilities are sizeable
- most parameters estimated precisely with posterior std deviation lower than prior std dev.
- state allocations are fairly sensible

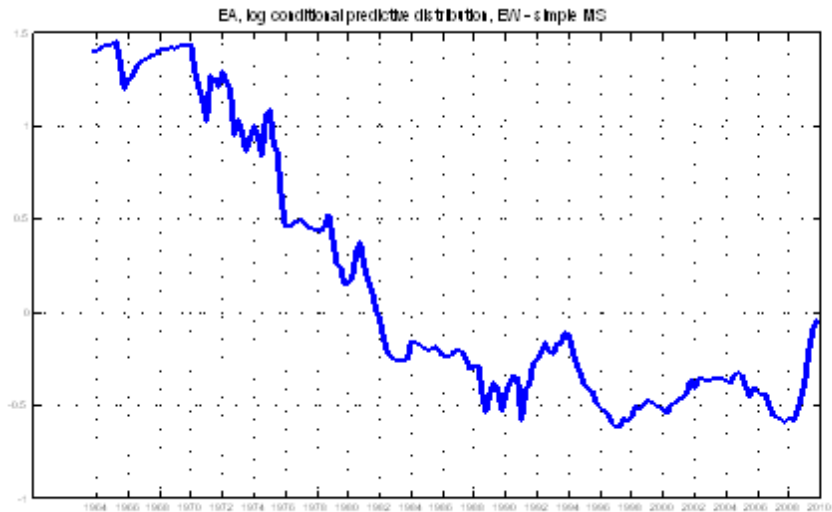
EW-MS: results for EA, I

	prior m	prior sd	post m.	post sd	post low	post up
ϕ	0.50	0.19	0.90	0.04	0.82	0.97
c_1	-0.84	1.24	0.15	0.10	-0.05	0.34
c_2	0.87	1.23	0.79	0.37	0.13	1.42
h_1	5.00	3.15	4.33	0.82	2.92	5.95
γ_{11}	1.50	0.05	0.93	0.10	0.72	1.08
γ_{21}	0.00	0.10	-0.22	0.17	-0.58	0.07
γ_{12}	-1.00	0.05	-0.47	0.09	-0.65	-0.32
γ_{22}	0.00	0.10	-0.22	0.17	-0.58	0.07
$\frac{c_1}{(1-\phi)}$	1.43					
$\frac{c_2}{(1-\phi)}$	7.65					

EW-MS: results for EA, II



EW-MS: results for EA, III



A MS-VAR model

- Following Sims and Zha, 2006.
- Amisano, (2010), in preparation
- US quarterly data 1956:Q1 to 2010:Q1 on Y, π, r, R
- flexible specification

$$\mathbf{A}(L)(\mathbf{y}_t - \boldsymbol{\mu}_{s_t}) = \mathbf{P}_{s_t} \mathbf{e}_t$$

- if only one lag, we have

$$(\mathbf{y}_t - \boldsymbol{\mu}_{s_t}) = \mathbf{A}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}) + \mathbf{P}_{s_t} \mathbf{e}_t$$

- EW version (financial indicators)

A DSGE model with MS in volatilities, motivation I

(Amisano and Tristani, (2010a, 2010b)

- A number of questions on the yield curve are related to its relationship with the macroeconomy:
 - Does the term structure reflect only (risk-neutral) expectations of future policy rates?
 - If risk premia are time-varying, what drives their variations?
 - How are term premia shaped by the monetary policy rule followed by the central bank?
- Problem: most macro-models don't speak to finance data. Risk premia are zero by construction when models are linearized completely

MS-DSGE, motivation II

- Some versions of consumption-based models are OK at fitting *unconditional* moments of yields (e.g. Piazzesi-Schneider, 2006; Wachter, 2006; Ravenna-Seppala, 2007; Gallmeyer-Hollifield-Palomino-Zin, 2007, Hördahl-Tristani-Vestin 2008; Rudebusch-Swanson, 2009)
- Key ingredients of these papers:
 - modelling inflation and the monetary policy rule is useful
 - "exotic preferences" (Backus-Routledge-Zin 2004) help
 - for structural models, higher (second or third) order approximation of the solution are used

MS-DSGE, main question in the paper

- Can we also fit *conditional* moments and produce time-variation in risk premia in a micro-founded, estimated general equilibrium model?
- Our route: allow for heteroskedastic shocks, rely on second order approximations and estimate the model using Bayesian methods
- On the basis of (preliminary) results, our tentative answer is yes

MS-DSGE, preferences

- Epstein-Zin-Weil preferences

$$U \left[u_t, \left(\mathbb{E}_t V_{t+1}^{1-\gamma} \right) \right] = \left\{ (1 - \beta) u_t^{1-\sigma} + \beta \left(\mathbb{E}_t V_{t+1}^{1-\gamma} \right)^{\frac{1-\sigma}{1-\gamma}} \right\}^{\frac{1}{1-\sigma}}$$

- Temporary utility

$$u_t = (C_t - hC_{t-1}) \cdot v(N_t)$$

MS-DSGE, the SDF

- Stochastic discount factor

$$Q_{t,t+1} = \beta \frac{\tilde{\Lambda}_{t+1}}{\tilde{\Lambda}_t} \frac{1}{\pi_{t+1}} \Phi_{t+1}$$

where

$$\begin{aligned} \tilde{\Lambda}_t &\equiv (C_t - hC_{t-1})^{-\sigma} [v(N_t)]^{1-\sigma} \\ &\quad - \beta h E_t (C_{t+1} - hC_t)^{-\sigma} [v(N_{t+1})]^{1-\sigma} \Phi_{t+1} \end{aligned}$$

$$\Phi_{t+1} \equiv \left(\frac{\left[E_t J_{t+1}^{1-\gamma} \right]^{\frac{1}{1-\gamma}}}{J_{t+1}} \right)^{\gamma-\sigma}$$

MS-DSGE: heteroskedastic shocks

- Shocks: productivity growth, government spending and inflation target (all serially correlated); policy (iid)

$$\hat{i}_t = \psi_{\Pi} (\pi_t - \pi_t^*) + \psi_Y (\tilde{y}_t - \tilde{y}) + \rho_I \hat{i}_{t-1} + \varepsilon_{\eta,t+1}$$

- Markov switching innovation variances:

$$\varepsilon_{\zeta,t+1} \sim N(0, \sigma_{\zeta, s_{\zeta,t}})$$

$$\varepsilon_{\eta,t+1} \sim N(0, \sigma_{\eta, s_{\eta,t}})$$

$$\varepsilon_{G,t+1} \sim N(0, \sigma_{G, s_{G,t}})$$

- Intuition: "Great moderation", "monetarist experiment", cyclical volatility

Solution

- We rely on second order perturbation methods
- The solution is

$$y_t = \kappa_{s_t} + F\hat{\mathbf{x}}_t + \frac{1}{2}\hat{\mathbf{x}}_t' E \hat{\mathbf{x}}_t$$

- Note: F and E are time-invariant. To second order, regime switching variances only affect the intercept κ_{s_t} of the policy functions

MS-DSGE, variability in risk premia

- Market prices of risk

$$\omega_t \equiv \Sigma'_t (F'_\pi - F'_\lambda)$$

- (1) prices of risk change across regimes, regime-switching risk is not priced (MS has no effect on 1-order vectors F_π, F_λ)
- (2) time-variability of prices of risk exogenous, but consistent with macro model:
 - changes in Σ_t must reflect conditional macro moments;
 - vectors F_π and F_λ are derived from microfoundations
- Expected excess holding period returns

$$\widehat{hpr}_{t,n} - \widehat{i}_t = F_{B_{n-1}} \Sigma_t \Sigma'_t (F'_\pi - F'_\lambda)$$

MS-DSGE, estimation method

- Problem: observation equation is nonlinear

$$\mathbf{y}_{t+1}^o = \mathbf{c}_j + \mathbf{C}_1 \mathbf{x}_{t+1} + \mathbf{C}_2 \overline{\text{vech}}(\mathbf{x}_{t+1} \mathbf{x}'_{t+1}) + \mathbf{D} \mathbf{v}_{t+1}$$

- Use as many observables as shocks to invert it for state variables. Multiple solutions. Scalar example

$$x_t = \begin{cases} \frac{-C_1 + \sqrt{C_1^2 - 4C_2(c_j - y_t^o)}}{2C_2} \\ \frac{-C_1 - \sqrt{C_1^2 - 4C_2(c_j - y_t^o)}}{2C_2} \end{cases}$$

- Pro: exact likelihood. Con: computing the likelihood is slow. In our model, 4 structural shocks and one observation error. At each t , up to 16 solutions \times 8 regimes.

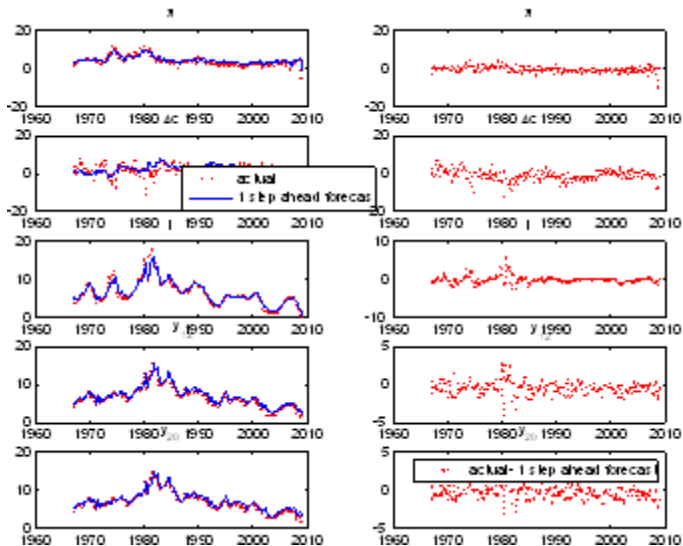
MS-DSGE, data

- Quarterly US data: 1966:q1 to 2009:q1
- Real personal per-capita consumption; consumption deflator; 3-month nominal rate; 3-year and 5-year zero-coupon yields
- "Measurement errors" on the 5-year yield

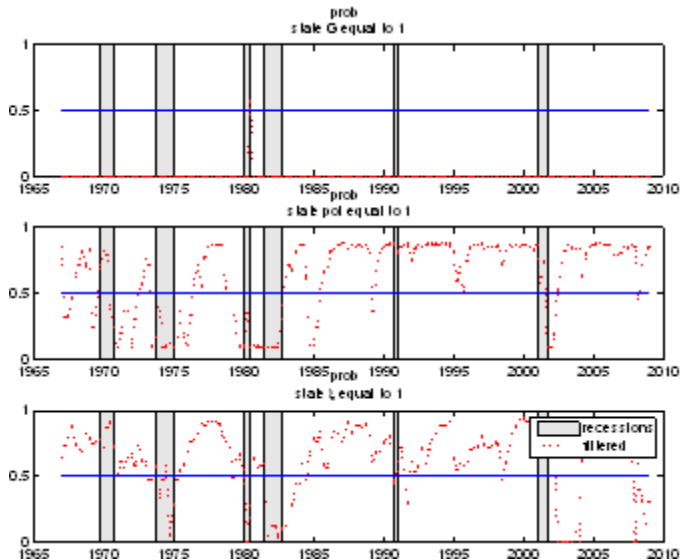
MS-DSGE, parameter estimates

	post mode	prior mean	prior sd
ψ_{π}	1.6998	0.5000	0.3535
ψ_y	0.0004	0.0308	0.0217
ρ_i	1.1887	0.7000	0.3000
ι	0.0135	0.6000	0.2000
ϕ	0.2391	2.0000	1.4142
γ	13.8494	5.0000	1.1547
σ	2.4739	2.0000	1.0000
ζ	19.6715	17.0000	2.0000
h	0.5309	0.7052	0.1074
θ	2.4758	8.0019	2.6457
β	0.9978	0.9900	0.0052
$\sigma_{me,20}$	0.0007	0.0014	0.0010

MS-DSGE, 1 step ahead forecast errors



MS-DSGE, regimes



MS-DSGE, excess holding period returns

