# Duration-Dependent Markov-Switching VAR Models with Applications to the Business Cycle Analysis

Matteo Pelagatti
Dipartimento di Statistica
Università di Milano-Bicocca
matteo.pelagatti@unimib.it

June 3, 2002

## 1 Introduction and motivation

Hamilton (1989) introduced the Markov-switching models (from now on MS) to the business cycle researchers. He applied the MS model to the real U.S. GDP and showed how good the probability of being in recession or expansion generated by his model matched the NBER dating. Since then, the studies on business cycles that exploited the MS model have grown exponentially. Nevertheless, the basic MS model, when used to model the business cycle, has at least two theoretical weaknesses: (i) it is univariate, (ii) it does not allow duration dependence. Since business cycles are fluctuations of aggregate economic activity, which express themselves through the comovements of many macro variables, point (i) is not a negligible limitation. The multivariate generalization of the MS model was carried out by Krolzig (1997), in his excellent work on the MS-VAR model. As far as point (ii) is concerned, some authors, such as Diebold, Rudebusch and Sichel (1993), Watson (1994), have found evidence of duration dependence in the U.S. business cycles, and therefore, as Diebold *et al.* (1993) point out, the MS model results miss-specified.

In the present paper a duration dependent MS-VAR model is presented and applied to the four macroeconomic time series that the NBER uses to date the U.S. business cycle. Bayesian inference on the unknown quantities of the model is carried out using Markov chain Monte Carlo methods, which outrun some limitation of the maximum-likelihood approach to such models.

A similar, although univariate, model was used by Kim and Nelson (1999, section 10.3.1) to test duration-dependence in the U.S. business cycle, but their inference on the state variable was carried out using a *single move Gibbs sampler*, which, in our experience, has a too slow convergence to the invariant distribution to be feasible in such a problem. In the next section it will be shown how to reformulate the inhomogeneous Markov chain used by Kim and Nelson in a homogeneous way, so that the fast converging *multimove Gibbs sampler* can be exploited.

# 2 The Model

The duration-dependent MS-VAR model is defined as follows:

$$\mathbf{y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 S_t + \mathbf{A}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 S_{t-1}) + \ldots + \mathbf{A}_p(\mathbf{y}_{t-p} - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 S_{t-p}) + \boldsymbol{\varepsilon}_t \quad (1)$$

where $\mathbf{y}_t$ is a vector of observable variables, $S_t$ is a binary (0-1) unobservable random variable following a Markov chain with varying transition probabilities, $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are coefficient matrices of a stable VAR process, and $\boldsymbol{\varepsilon}_t$ is a gaussian (vector) white noise with covariance matrix $\boldsymbol{\Sigma}$.

In order to achieve duration dependence for $S_t$, a Markov chain is built for the pair $(S_t, D_t)$, where $D_t$ is the duration variable defined as follows:

$$D_t = \begin{cases} D_{t-1} + 1 & \text{if } S_t = S_{t-1} \\ 1 & \text{if } S_t \neq S_{t-1} \end{cases} ; \quad (2)$$

an example of a possible sample path of the Markov chain $(S_t, D_t)$ is shown in table 1). A maximum value, $\tau$, for the duration variable $D_t$ must be fixed,

| $t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_t$ | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| $D_t$ | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 1 | 1 | 2 | 3 | 4 |

Table 1: A possible realization of processes $S_t$ and $D_t$.

so that the Markov chain $(S_t, D_t)$ is defined on the finite state space

$$\{(0,1), (1,1), (0,2), (1,2), \ldots, (0,\tau), (1,\tau)\},$$

with finite transition matrix[1]

$$\mathbf{P}=\begin{bmatrix} 0 & p_{0|1}(1) & 0 & p_{0|1}(2) & 0 & p_{0|1}(3) & \dots & 0 & p_{0|1}(\tau) \\ p_{1|0}(1) & 0 & p_{1|0}(2) & 0 & p_{1|0}(3) & 0 & \dots & p_{1|0}(\tau) & 0 \\ p_{0|0}(1) & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & p_{1|1}(1) & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & p_{0|0}(2) & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & p_{1|1}(2) & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & p_{0|0}(\tau) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & p_{1|1}(\tau) \end{bmatrix}$$

where $p_{i|j}(d) = \Pr(S_t = i|S_{t-1} = j, D_{t-1} = d)$.

When $D_t = \tau$, only four events are given non-zero probabilities:

$$\begin{array}{ll} (S_t = i, D_t = \tau)|(S_{t-1} = i, D_{t-1} = \tau) & i = 0, 1 \\ (S_t = i, D_t = 1)|(S_{t-1} = j, D_{t-1} = \tau) & i \neq j, \quad i, j = 0, 1 \end{array},$$

with the interpretation that, when the economy has been in state $i$ at least $\tau$ times, the additional periods in which the state remains $i$ influence no more the probability of transition.

As pointed out by Hamilton (1994, section 22.4), it is always possible to write the likelihood function of $\mathbf{y}_t$, depending only on the state variable at time $t$, even though in the model a $p$-order autoregression is present; this can be done using the extended state variable $S_t^* = (D_t, S_t, S_{t-1}, \dots, S_{t-p})$, which comprehends all the possible combinations of the states of the economy in the last $p$ periods. In table 2 the state space of non-negligible[2] $S_t^*$ when $p = 4$ and $\tau = 5$ is shown. If $\tau \geq p$ the maximum number of non-negligible states is given by $u = \sum_{i=1}^{p} 2^i + 2(\tau - p)$. The transition matrix $\mathbf{P}^*$ of the Markov chain $S_t^*$ will usually be a $(u \times u)$ matrix, although rather sparse, having a maximum number $2\tau$ of non-zero elements.

In order to reduce the number $(2\tau)$ of non-zero elements in $\mathbf{P}^*$ to be estimated, a more parsimonious probit specification is used. Consider the linear model

$$S_t^\bullet = [\beta_1 + \beta_2 D_{t-1}]S_{t-1} + [\beta_3 + \beta_4 D_{t-1}](1 - S_{t-1}) + \epsilon_t \tag{3}$$

[1] The transition matrix is here designed so that the rows, and not the columns, sum to ones.

[2] "Negligible states" stands here for 'states always associated with zero probability'. For example the state $(D_t = 5, S_t = 1, S_{t-1} = 0, S_{t-2} = s_2, S_{t-3} = s_3, S_{t-4} = s_4)$, where $s_2$, $s_3$ and $s_4$ can be either 0 or 1, is negligible as it is not possible for $S_t$ to have been 5 periods in the same state, if the state at time $t-1$ is different from the state at time $t$.

| | $D_t$ | $S_t$ | $S_{t-1}$ | $S_{t-2}$ | $S_{t-3}$ | $S_{t-4}$ | | $D_t$ | $S_t$ | $S_{t-1}$ | $S_{t-2}$ | $S_{t-3}$ | $S_{t-4}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 | 0 | 0 | 17 | 2 | 0 | 0 | 1 | 0 | 0 |
| 2 | 1 | 0 | 1 | 0 | 0 | 1 | 18 | 2 | 0 | 0 | 1 | 0 | 1 |
| 3 | 1 | 0 | 1 | 0 | 1 | 0 | 19 | 2 | 0 | 0 | 1 | 1 | 0 |
| 4 | 1 | 0 | 1 | 0 | 1 | 1 | 20 | 2 | 0 | 0 | 1 | 1 | 1 |
| 5 | 1 | 0 | 1 | 1 | 0 | 0 | 21 | 2 | 1 | 1 | 0 | 0 | 0 |
| 6 | 1 | 0 | 1 | 1 | 0 | 1 | 22 | 2 | 1 | 1 | 0 | 0 | 1 |
| 7 | 1 | 0 | 1 | 1 | 1 | 0 | 23 | 2 | 1 | 1 | 0 | 1 | 0 |
| 8 | 1 | 0 | 1 | 1 | 1 | 1 | 24 | 2 | 1 | 1 | 0 | 1 | 1 |
| 9 | 1 | 1 | 0 | 0 | 0 | 0 | 25 | 3 | 0 | 0 | 0 | 1 | 0 |
| 10 | 1 | 1 | 0 | 0 | 0 | 1 | 26 | 3 | 0 | 0 | 0 | 1 | 1 |
| 11 | 1 | 1 | 0 | 0 | 1 | 0 | 27 | 3 | 1 | 1 | 1 | 0 | 0 |
| 12 | 1 | 1 | 0 | 0 | 1 | 1 | 28 | 3 | 1 | 1 | 1 | 0 | 1 |
| 13 | 1 | 1 | 0 | 1 | 0 | 0 | 29 | 4 | 0 | 0 | 0 | 0 | 1 |
| 14 | 1 | 1 | 0 | 1 | 0 | 1 | 30 | 4 | 1 | 1 | 1 | 1 | 0 |
| 15 | 1 | 1 | 0 | 1 | 1 | 0 | 31 | 5 | 0 | 0 | 0 | 0 | 0 |
| 16 | 1 | 1 | 0 | 1 | 1 | 1 | 32 | 5 | 1 | 1 | 1 | 1 | 1 |

Table 2: State space of $S_t^* = (D_t, S_t, S_{t-1}, \ldots, S_{t-p})$ for $p = 4$, $\tau = 5$.

with $\epsilon_t \sim \mathcal{N}(0, 1)$, and $S_t^\bullet$ latent variable defined by

$$\Pr(S_t^\bullet \geq 0 | S_{t-1}, D_{t-1}) = \Pr(S_t = 1 | S_{t-1}, D_{t-1}) \quad (4)$$
$$\Pr(S_t^\bullet < 0 | S_{t-1}, D_{t-1}) = \Pr(S_t = 0 | S_{t-1}, D_{t-1}). \quad (5)$$

It's easy to show that it holds

$$p_{1|1}(d) = \Pr(S_t = 1 | S_{t-1} = 1, D_{t-1} = d) = \quad (6)$$
$$= 1 - \Phi(-\beta_1 - \beta_2 d)$$
$$p_{0|0}(d) = \Pr(S_t = 0 | S_{t-1} = 0, D_{t-1} = d) = \Phi(-\beta_3 - \beta_4 d) \quad (7)$$

where $d = 1, \ldots, \tau$, and $\Phi(.)$ is the standard normal cumulative distribution function. Now the four parameters $\beta$ completely define the transition matrix $\mathbf{P}^*$.

# 3 Bayesian Inference on the Model's Parameters

Bayesian inference on the model's unknowns $(\boldsymbol{\mu}, \mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \{(S_t, D_t)\}_{t=1}^T)$, where $\boldsymbol{\mu} = (\boldsymbol{\mu}_0', \boldsymbol{\mu}_1')'$ and $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_p)$, will be carried out using MCMC methods.

It will be used the prior distribution

$$p(\boldsymbol{\mu}, \mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, (S_0, D_0)) = p(\boldsymbol{\mu})p(\mathbf{A})p(\Sigma)p(\boldsymbol{\beta})p(S_0, D_0),$$

where

$$\begin{aligned} \boldsymbol{\mu} &\sim \mathcal{N}(\mathbf{m}_0, \mathbf{M}_0), & (8)\\ \text{vec}(\mathbf{A}) &\sim \mathcal{N}(\mathbf{a}_0, \mathbf{A}_0), & (9)\\ p(\boldsymbol{\Sigma}) &= |\boldsymbol{\Sigma}|^{-\frac{1}{2}(\text{rank}(\boldsymbol{\Sigma})+1)}, & (10)\\ \boldsymbol{\beta} &\sim \mathcal{N}(\mathbf{b}_0, \mathbf{B}_0), & (11)\\ & & (12) \end{aligned}$$

and $p(S_0, D_0)$ is the ergodic probability function of the Markov chain $\{S_t, D_t\}$.

Let $\boldsymbol{\theta}_i$, $i = 1, \ldots, I$, be a partition of the set $\boldsymbol{\theta}$ containing all the unknowns of the model, and $\boldsymbol{\theta}_{-i}$ represent the set $\boldsymbol{\theta}$ without the elements of $\boldsymbol{\theta}_i$. In order to implement a Gibbs sampler to sample from the joint posterior distribution of all the unknowns of the model, it is sufficient to find the full conditional posterior distribution $p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \mathbf{Y})$, with $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$ and $i = 1, \ldots, I$. A Gibbs sampler iteration is a generation of random numbers from $p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \mathbf{Y})$, $i = 1, \ldots, I$, where the elements of $\boldsymbol{\theta}_{-i}$ are substituted with the most recent generated values. Since, under mild conditions the Markov chain defined for $\boldsymbol{\theta}^{(i)}$, where $\boldsymbol{\theta}^{(i)}$ is the value of $\boldsymbol{\theta}$ generated at the $i^{\text{th}}$ iteration of the Gibbs sampler, converges to its stationary distribution, and that this stationary distribution is the "true" posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y})$, it is sufficient to fix an initial burn-in period of $M$ iterations, such that the Markov chain may "forget" the starting values $\boldsymbol{\theta}^{(0)}$ (furnished by the researcher), to sample form (an approximation of) the joint posterior distribution. The samples obtained for each element of $\boldsymbol{\theta}$ are samples from the marginal posterior distribution of each parameters.

## Gibbs Sampling Steps

### Step 1. Generation of $\{S_t^*\}_{t=1}^T$

We will use an implementation of the multi-move Gibbs sampler originally proposed by Carter and Kohn (1994), which, suppressing the conditioning on the other parameters from the notation, exploits the equality

$$p(S_1^*, \ldots, S_T^*|\mathbf{Y}_T) = p(S_T^*|\mathbf{Y}_T) \prod_{t=1}^{T-1} p(S_t^*|S_{t+1}^*, \mathbf{Y}_t), \tag{13}$$

with $\mathbf{Y}_t = (\mathbf{y}_1, \ldots, \mathbf{y}_t)$.

Let $\hat{\boldsymbol{\xi}}_{t|i}$ be the vector containing the probabilities of $S_t^*$ being in each state (the first element is the probability of being in state 1, the second element is the probability of being in state 2, and so on) given $\mathbf{Y}_i$ and the model's parameters. Let $\boldsymbol{\eta}_t$ be the vector containing the likelihood of each state given $\mathcal{Y}_t$ and the model's parameters, whose generic element is

$$(2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y}_t-\hat{\mathbf{y}}_t)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t-\hat{\mathbf{y}}_t)\right\},$$

where

$$\hat{\mathbf{y}}_t=\boldsymbol{\mu}_0+\boldsymbol{\mu}_1 S_t+\mathbf{A}_1(\mathbf{y}_{t-1}-\boldsymbol{\mu}_0-\boldsymbol{\mu}_1 S_{t-1})+\ldots+\mathbf{A}_p(\mathbf{y}_{t-p}-\boldsymbol{\mu}_0-\boldsymbol{\mu}_1 S_{t-p})$$

changes value according to the state of $S_t^*$.

The filtered probabilities of the states can be calculated using the Hamilton's (1994, section 22.4) filter

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

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

with the symbol $\odot$ indicating element by element multiplication. The filter is completed with the prior probabilities vector $\hat{\boldsymbol{\xi}}_{1|0}$, that, as already said, will be set equal to the vector of ergodic (or stationary) probabilities.

To sample from the distribution of $\{S_t^*\}_1^T$ given the full information set $\mathbf{Y}_T$, it can be shown that it holds

$$\Pr(S_t^*=i|S_{t+1}^*=k,\mathbf{Y}_t)=\frac{p_{i|k}\hat{\xi}_{t|t}^{(i)}}{\sum_{j=1}^m p_{j|k}\hat{\xi}_{t|t}^{(j)}},$$

where $p_{j|k}$ indicates the transition probability of moving to state $j$ from state $k$ (element $(j,k)$ in the transition matrix $\mathbf{P}^*$) and $\xi_{t|t}^{(j)}$ is the $j$-th element of vector $\boldsymbol{\xi}_{t|t}$. In matrix notation the same can be written as

$$\hat{\boldsymbol{\xi}}_{t|(S_{t+1}^*=k,\mathbf{Y}_T)}=\frac{\mathbf{P}_{[k,.]}'\odot\hat{\boldsymbol{\xi}}_{t|t}}{\mathbf{P}_{[k,.]}\hat{\boldsymbol{\xi}}_{t|t}} \tag{14}$$

where $\mathbf{P}_{[k,.]}$ identifies the $k$-th row of the transition matrix $\mathbf{P}$.

Now all the probability functions in equation (13) have been given a form, and the states can, thus, be generated starting from the filtered probability

$\hat{\boldsymbol{\xi}}_{T|T}$ and proceeding backward $(N-1,\ldots,1)$, using equation (14) where $k$ is to be substituted with the last generated value $(s_{t+1}^*)$.

Once as set of sampled $\{S_t^*\}$ has been generated, it is automatically available a sample for $\{S_t\}$ and $\{D_t\}$.

The advantage of using the described multi-move Gibbs sampler, compared to the single move Gibbs sampler that can be implemented as in Carlin, Polson and Stoffer (1992), or using the excellent software BUGS, is that the whole vector of states is sampled at once, improving significantly the speed of convergence of the Gibbs samper's chain to its ergodic distribution.

### Step 2. Generation of $(\mathbf{A},\boldsymbol{\Sigma})$

Conditionally on $\{S_t\}_{t=1}^T$ and $\boldsymbol{\mu}$ equation 1 is just a multivariate normal regression model for the variable $\mathbf{y}_t^* = \mathbf{y}_t - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 S_t$, whose parameters, given the discussed prior distribution, have the following posterior distributions, known in literature. Let $\mathbf{X}$ be the matrix, whose $t^{\text{th}}$ column is

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{y}_t^* \\ \mathbf{y}_{t-1}^* \\ \vdots \\ \mathbf{y}_{t-p}^* \end{pmatrix},$$

for $t = 1,\ldots,T$, and let $\mathbf{Y}^* = (\mathbf{y}_1^*,\ldots,\mathbf{y}_T^*)$.

The posterior for $(\text{vec}(\mathbf{A}),\boldsymbol{\Sigma})$ is, suppressing the conditioning on the other parameters, the normal-inverse Wishart distribution

$$\begin{aligned} &p(\text{vec}(\mathbf{A}),\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X}) = p(\text{vec}(\mathbf{A})|\boldsymbol{\Sigma},\mathbf{Y},\mathbf{X})p(\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X}) \\ &p(\boldsymbol{\Sigma}|\mathbf{Y},\mathbf{X}) \text{ density of a } \mathcal{IW}_k(\mathbf{V},n-m) \\ &p(\text{vec}(\mathbf{A})|\boldsymbol{\Sigma},\mathbf{Y},\mathbf{X}) \text{ density of a } \mathcal{N}(\mathbf{a}_1,\mathbf{A}_1), \end{aligned}$$

with

$$\begin{aligned} &\mathbf{V} = \mathbf{Y}^*\mathbf{Y}^{*\prime} - \mathbf{Y}^*\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{Y}^{*\prime} \\ &\mathbf{A}_1 = (\mathbf{A}_0^{-1} + \mathbf{X}\mathbf{X}'\boldsymbol{\Sigma}^{-1})^{-1} \\ &\mathbf{a}_1 = \mathbf{A}_1[\mathbf{A}_0^{-1}\mathbf{a}_0 + (\mathbf{X}\otimes\boldsymbol{\Sigma}^{-1})\text{vec}(\mathbf{Y})]. \end{aligned}$$

### Step 3. Generation of $\boldsymbol{\mu}$

Conditionally on $\mathbf{A}$ and $\boldsymbol{\Sigma}$, by multiplying both sides of equation (2) times

$$\mathbf{A}(L) = (\mathbf{I} - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p),$$

where $L$ is the lag or backward operator, we obtain

$$\mathbf{A}(B)\mathbf{y}_t = \boldsymbol{\mu}_0\mathbf{A}(1) + \boldsymbol{\mu}_1\mathbf{A}(B)S_t + \boldsymbol{\varepsilon}_t,$$

which is a multivariate normal linear regression model with known variance $\boldsymbol{\Sigma}$, and can be treated as shown at step 2., with respect to the specified prior for $\boldsymbol{\mu}$.

**Step 4. Generation of $\boldsymbol{\beta}$**

Conditionally on $\{S_t^*\}_{t=1}^T$, $\boldsymbol{\beta}$ contains the parameters of the porbit model described in section 2. Albert and Chib (1993) have proposed a method based on a data augmentation algorithm to simulate values for the parameters of a probit model. Given the parameter vector $\boldsymbol{\beta}$ of last iteration of the Gibbs sampler, generate the latent variables $\{S_t^\bullet\}$ from the respective truncated normal densities

$$\begin{aligned} S_t^\bullet|(S_t = 0, \mathbf{x}_t, \boldsymbol{\beta}) &\sim \mathcal{N}(\mathbf{x}_t'\boldsymbol{\beta}, 1)\mathbb{I}_{(-\infty,0)} & (15)\\ S_t^\bullet|(S_t = 1, \mathbf{x}_t, \boldsymbol{\beta}) &\sim \mathcal{N}(\mathbf{x}_t'\boldsymbol{\beta}, 1)\mathbb{I}_{[0,\infty)} & (16)\\ & & (17) \end{aligned}$$

with

$$\begin{aligned} &\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)' \\ &\mathbf{x}_t = (S_{t-1}, D_{t-1}, (1 - S_{t-1}), (1 - D_{t-1}))' \end{aligned}$$

and $\mathbb{I}_{(a,b)}$ indicator variable used to denote truncation.

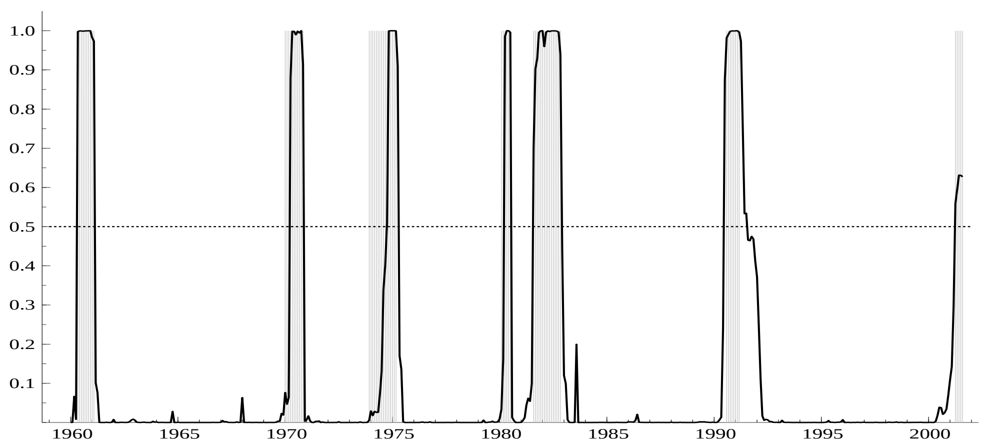
With the generated $S_t^\bullet$'s the probit regression equation (3) becomes, again, a normal linear model with known variance.

# 4 Application to the Analysis of the U.S. Business Cycle

The model proposed in the previous section has been applied to 100 times the log-difference of the four time series, on which the NBER relays to date the U.S. business cycle: industrial production (IP), total nonfarm-employment (EMP), total manufacturing and trade sales in million of 1996$ (TRADE), personal income less transfer payments in billions of 1996$ (INCOME).

The model[3], with $\tau = 60$ and $p = 2$ did not work too well, when the VAR component was present, while the results in absence of the VAR part ($\tau = 60$, $p = 0$) are rather encouraging[4]. Summaries of the marginal posterior distributions, based on a Gibbs sample of 11000 points, are shown in the table below, while figure 1 compares the probability of the U.S. economy being in recession resulting from the estimated model with the official NBER dating of the business cycles: the signal "probability of being in recession" extracted by the model here presented matches rather well the official dating, and is rather less noisy than the signal extracted by Hamilton (1989). Figure 2 shows how the duration of a state (recession or expansion) influences the transition probabilities: while the probability of moving from a recession into an expansion seems to be influenced by the duration of the recession, the probability of falling into a recession appears to be independent of the length of the expansion.

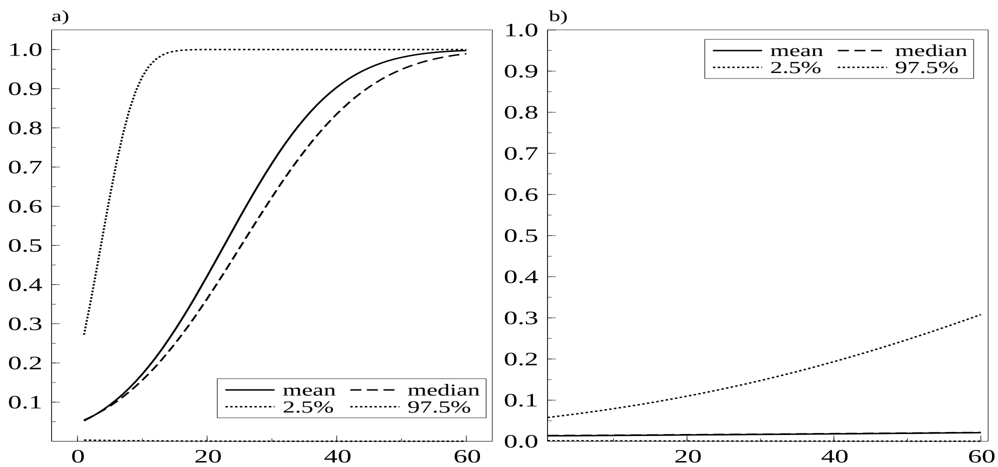
Figure 1: (Smoothed) probability of recession (line) and NBER dating (gray shade)




[3]All the calculations in this study have been done in Aptech Gauss, using the library MSARGIBB, written by the author to carry out Bayesian MCMC inference for a large class of dynamic models. The msargibb library and the programs used in the paper are available upon request.

[4]This is probably due to the fact that the duration dependent MS model is a stationary process, which, therefore, can be approximated with an autoregerssive model: so the duration dependent switching part and the VAR part try to "explain" almost the same features of the series, and the model is not too well identified.

Figure 2: Probability of moving a) from a recession into an expansion after $d$ months of recession b) from an expansion to a recession after $d$ months of expansion

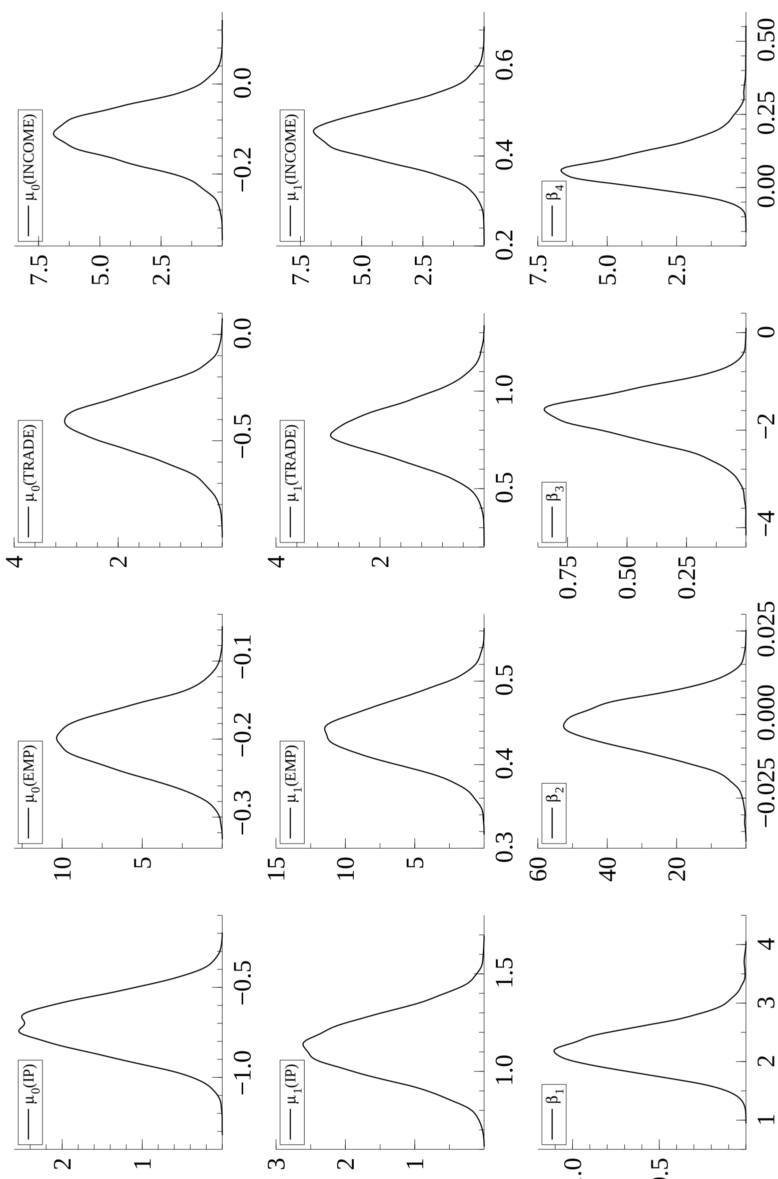



| Parameter | Prior mean | Prior var | Posterior mean | s.d. | 2.5% | 50% | 97.5% |
|---|---|---|---|---|---|---|---|
| $\mu_0$ IP | -0.300 | 1.000 | -0.708 | 0.145 | -0.987 | -0.707 | -0.435 |
| $\mu_0$ EMP | -0.300 | 1.000 | -0.200 | 0.036 | -0.269 | -0.200 | -0.132 |
| $\mu_0$ TRADE | -0.300 | 1.000 | -0.417 | 0.132 | -0.688 | -0.413 | -0.167 |
| $\mu_0$ INCOME | -0.300 | 1.000 | -0.111 | 0.056 | -0.224 | -0.111 | 0.004 |
| $\mu_1$ IP | 1.500 | 1.000 | 1.130 | 0.146 | 0.853 | 1.129 | 1.410 |
| $\mu_1$ EMP | 1.500 | 1.000 | 0.440 | 0.034 | 0.376 | 0.439 | 0.504 |
| $\mu_1$ TRADE | 1.500 | 1.000 | 0.795 | 0.139 | 0.527 | 0.792 | 1.076 |
| $\mu_1$ INCOME | 1.500 | 1.000 | 0.449 | 0.056 | 0.341 | 0.449 | 0.562 |
| $\beta_1$ | 0.000 | 5.000 | 2.224 | 0.358 | 1.591 | 2.203 | 2.982 |
| $\beta_2$ | 0.000 | 5.000 | -0.003 | 0.007 | -0.018 | 0.003 | 0.010 |
| $\beta_3$ | 0.000 | 5.000 | -1.698 | 0.478 | -2.708 | -1.672 | -0.839 |
| $\beta_4$ | 0.000 | 5.000 | 0.075 | 0.067 | -0.032 | 0.066 | 0.232 |

# 5 Conclusions

The model proved to have a good capability of discerning recessions and expansions, as the probabilities of recession tend to assume very low or very high values. Our probabilities of recession (between Jan-1960 and Aug-2001) have a correlation of 0.83 with the NBER classification, and using a 0.5-rule to determine the state of the economy, only 4.2% (21 out of 499) of our

Figure 3: Kernel density estimates of some parameters of the model.



states differ from the NBER states. It also interesting to notice how the last recession is picked up by our model, with a perfect synchronization to the NBER classification, with data until August 2001, while the official NBER announcement of the peak of March 2001 dates November 26, 2001[5].

The results on the duration-dependence of the business cycles are similar to those of Diebold and Rudebusch (1990), Diebold *et al.* (1993), Sichel (1991) and Durland and McCurdy (1994): recessions are duration dependent, while expansions seem to be not duration dependent. Some authors[6] have have argued that the value of the maximum cycle duration used in our application (60 months) is sufficient for a duration-dependence analysis of the recessions, which in the U.S. economy hardly reach two years of duration, but are too few for the same analysis on the expansions. Our failure to detected such a duration could, thus, be due also to a too low $\tau$ in our model. We will definitively try the same analysis with a higher value of $\tau$.

An analysis of the forecasting performance of our model will follow in future papers.

# References

Albert J.H. and Chib S. (1993) Bayesian analysis of binary and polychotomous responce data, *Journal of the American Statistical Association*, 88, 669–79.

Carlin B.P., Polson N.G. and Stoffer D.S. (1992) A monte carlo approach to nonnormal and nonlinear state-space modeling, *Journal of the American Statistical Association*, 87, 493–500.

Carter C.K. and Kohn R. (1994) On gibbs sampling for state space models, *Biometrika*, 81, 541–53.

Diebold F. and Rudebusch G. (1990) A nonparametric investigation of duration dependence in the american business cycle, *Journal of Political Economy*, 98, 596–616.

[5]Refer to the Internet page http://www.nber.org/cycles/november2001/. It's not possible to draw conclusions on the compared speed of the NBER and of our model to classify a new cycle because we are not aware of the delay and the precision, with which the NBER receives the first data on the four times series.

[6]We thank Prof. Estela Bee Dagum and Prof. Aride Mazzali for their helpful suggestions.

Diebold F., Rudebusch G. and Sichel D. (1993) Further evidence on business cycle duration dependence, in: Business Cycles, Indicators and Forcasting, Stock J. and Watson M., eds., The University of Chicago Press.

Durland J. and McCurdy T. (1994) Duration–dependent transitions in a markov model of u.s. gnp growth, *Journal of Business and Economic Statistics*, 12, 279–288.

Hamilton J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357–384.

Hamilton J.D. (1994) *Time Series Analysis*, Princeton University Press.

Kim C.J. and Nelson C.R. (1999) *State-space models with regime switching: classical and Gibbs-sampling approches with applications*, The MIT Press.

Krolzig H.M. (1997) *Markov-Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*, Springer-Verlag.

Sichel D.E. (1991) Business cycle duration dependence: A parametric approach, *Review of Economics and Statistics*, 73, 254–256.

Watson J. (1994) Business cycle durations and postwar stabilization of the u.s. economy, *American Economic Review*, 84, 24–46.