# Duration Dependent Markov-Switching Vector Autoregression: Properties, Bayesian Inference, Software and Application ⋆

November 2005

Matteo M. Pelagatti [1]

*Department of Statistics, Università di Milano–Bicocca, I-20126 Milan, Italy*

**Abstract**

Duration dependent Markov-switching VAR (DDMS-VAR) models are time series models with data generating process consisting in a mixture of two VAR processes. The switching between the two VAR processes is governed by a two state Markov chain with transition probabilities that depend on how long the chain has been in a state. In the present paper we analyze the second order properties of such models and propose a Markov chain Monte Carlo algorithm to carry out Bayesian inference on the model's unknowns. Furthermore, a freeware software written by the author for the analysis of time series by means of DDMS-VAR models is illustrated. The methodology and the software are applied to the analysis of the U.S. business cycle.

*Key words:* Markov-switching, business cycle, Gibbs sampler, duration dependence, vector autoregression.

## 1 Introduction and motivation

Since the path-breaking paper of Hamilton (1989), many applications of the Markov switching autoregressive model (MS-AR) to business cycle analysis have demonstrated its potential, particularly in dating the cycle in an "objective" way. The basic MS-AR model has, nevertheless, some limitations: (i) it is univariate, (ii) the probabilities of transition from one state to the other

⋆ An earlier version of this paper was presented at the 1st OxMetrics User Conference, London 2003. I would like to thank Prof. David Hendry for useful comments and his encouragement. This work was partially supported by a grant of the Italian Ministry of Education, University and Research (MIUR).
[1] Tel.: +39 02 64485834; fax: +39 02 6473312.
*E-mail address:* matteo.pelagatti@unimib.it.

(or to the other ones) are constant over time, iii) it is not capable of generating spectra with peaks in the business cycle frequencies. Since business cycles are fluctuations of the aggregate economic activity, involving many macroeconomic variables at the same time, point (i) is not a negligible weakness. The multivariate generalization of the MS model was carried out by Krolzig (1997), in his excellent monograph on the MS-VAR model.

As far as point (ii) is concerned, it is reasonable to believe that the probability of exiting a contraction is not the same at the very beginning of this phase as after several months. Some authors, such as Diebold and Rudebusch (1990), Diebold et al. (1993) and Watson (1994) have found evidence of duration dependence in the U.S. business cycles, and therefore, as Diebold et al. (1993) point out, the standard MS model results, in this framework, miss-specified. In order to face this limitation, Durland and McCurdy (1994) introduced the univariate duration-dependent Markov switching autoregression, designing an alternative filter for the unobservable state variable. In the present article the duration-dependent switching model is generalized in a multivariate manner, and it is shown how standard tools related to the MS-AR model, such as Hamilton's filter and Kim's smoother (Kim, 1994) can be used to model duration dependence. Indeed, the filter proposed by Durland and McCurdy (1994) may be shown to be equivalent to Hamilton's filter calculated for a more general Markov chain. While Durland and McCurdy (1994) carry out their inference on the model by exploiting maximum likelihood estimation, we relay on Bayesian inference using Markov chain Monte Carlo (MCMC) techniques. The advantages of this technique are at least threefold. It does not relay on asymptotics [2], and in latent variable models, where the unknowns are many, *asymptopia* may be far away. Inference on the latent variables is not conditional on the estimated parameters (like in MLE). Furthermore, since inference on MS models is notoriously rather sensitive to the presence of outliers, the possibility of using prior distributions on the parameters may limit their damages, making the estimates more robust.

As far as point iii) is concerned, the analysis of the second order properties of DDMS-VAR models carried out in this paper demonstrates that these processes may generate spectra with peaks in business cycle frequencies, similar to the typical spectral shapes of many (detrended) economic variables.

The work is organized as follows: the duration-dependent Markov switching VAR model (DDMS-VAR) is defined in section 2, its second order properties are derived in section 3, while the MCMC-based Bayesian inference is

[2] Actually MCMC techniques do relay on asymptotic results, but the size of the sample is under control of the researcher and some diagnostics on convergence are available. Here it is meant that the reliability of the inference does not depend on the sample size of the real-world data.

explained in section 4; section 5 briefly illustrates the features of the software DDMSVAR for Ox, written by the author for modelling with DDMS-VAR models, and an application of the model and of the software to the U.S. business cycle is carried out in section 6.

## 2 The model

The duration-dependent MS-VAR model[3] is defined by

$$\begin{aligned} \boldsymbol{y}_t = &\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 S_t + \boldsymbol{A}_1(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 S_{t-1}) + \dots \\ &+ \boldsymbol{A}_p(\boldsymbol{y}_{t-p} - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 S_{t-p}) + \boldsymbol{\varepsilon}_t \end{aligned} \tag{1}$$

where $\boldsymbol{y}_t$ is a vector of observable variables, $S_t$ is two state $\{0, 1\}$ Markov chain with time varying transition probabilities, $\boldsymbol{A}_1, \dots, \boldsymbol{A}_p$ 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 allow for duration dependence, the pair $(S_t, D_t)$ is considered, where $D_t$ is the duration variable defined by

$$D_t = \begin{cases} 1 & \text{if } S_t \neq S_{t-1} \\ D_{t-1} + 1 & \text{if } S_t = S_{t-1} \text{ and } D_{t-1} < \tau \\ D_{t-1} & \text{if } S_t = S_{t-1} \text{ and } D_{t-1} = \tau \end{cases} . \tag{2}$$

It easy to see that $(S_t, D_t)$ is also a Markov chain, since conditionally on $(S_{t-1}, D_{t-1})$, $(S_t, D_t)$ is independent of $(S_{t-k}, D_{t-k})$ with $k = 2, 3, \dots$. An example of a possible sample path of $(S_t, D_t)$ is shown in table 1. The value $\tau$

Table 1
A possible realization of the process $(S_t, D_t)$.

| $t$ | 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 |

is the maximum that the duration variable $D_t$ can reach and must be fixed *a priori* so that the Markov chain $(S_t, D_t)$ be defined on the finite state space

$$\{(0, 1), (1, 1), (0, 2), (1, 2), \dots, (0, \tau), (1, \tau)\}. \tag{3}$$

[3] Using Krolzig's terminology, we are defining a duration dependent MSM(2)-VAR, that is, Markov-Switching in Mean VAR with two states.

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

$$(S_t = i, D_t = \tau)|(S_{t-1} = i, D_{t-1} = \tau) \quad i = 0, 1$$
$$(S_t = i, D_t = 1)|(S_{t-1} = j, D_{t-1} = \tau) \quad i \neq j, \quad i, j = 0, 1.$$

with the following interpretation: 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 probabilities of transition. Thus, the transition matrix $\boldsymbol{P}$ has the form [4]

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

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

As pointed out by Hamilton (1994, section 22.4), it is always possible to write the likelihood function of $\boldsymbol{y}_t$, 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 states [5] $S_t^*$, with $p = 4$ and $\tau = 5$, is shown. If $\tau \geq p$ the number of non-negligible states is given by $u = 2(2^p + \tau - p - 1)$. The transition matrix $\boldsymbol{P}^*$ of the Markov chain $S_t^*$ is a rather sparse $(u \times u)$ matrix, having a maximum number $2\tau$ of independent non-zero elements.

[4] The transition matrix is designed so that the elements of each column sum to one. Our transition matrix is the transpose of the usual transition matrix in Markov chain literature.

[5] "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 $t$.

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

| | $D_t$ | $S_t$ | $S_{t-1}$ | $S_{t-2}$ | $S_{t-3}$ | $S_{t-4}$ | | $D_t$ | $S_t$ | $S_{t-1}$ | $S_{t-2}$ | $S_{t-3}$ | $S_{t-4}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 |

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

$$Z_t = [\beta_1 + \beta_2 D_{t-1}]S_{t-1} + [\beta_3 + \beta_4 D_{t-1}](1 - S_{t-1}) + \epsilon_t \tag{4}$$

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

$$\Pr(Z_t \geq 0 | S_{t-1}, D_{t-1}) = \Pr(S_t = 1 | S_{t-1}, D_{t-1}) \tag{5}$$
$$\Pr(Z_t < 0 | S_{t-1}, D_{t-1}) = \Pr(S_t = 0 | S_{t-1}, D_{t-1}). \tag{6}$$

It's easy to show that

$$p_{1|1}(d) = \Pr(S_t = 1 | S_{t-1} = 1, D_{t-1} = d) = \tag{7}$$
$$= 1 - \Phi(-\beta_1 - \beta_2 d)$$
$$p_{0|0}(d) = \Pr(S_t = 0 | S_{t-1} = 0, D_{t-1} = d) = \Phi(-\beta_3 - \beta_4 d) \tag{8}$$

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

## 3 Second order properties of the model

The second order properties of a non-linear, non-gaussian process are by no means exhaustive in describing its behavior, nevertheless there are good reasons for studying the cross- and auto-covariance structure and spectrum of such time series models. From a practical point of view, practitioners usually analyze the features of economic time series by means of sample second order moments; furthermore important concepts like business cycle, seasonality, etc. are (implicitly or explicitly) defined in the frequency domain.

For the purpose of this section, it is convenient to use the VAR representation of a Markov chain (Hamilton, 1994, p.679). Let $X_t$ be a Markov chain with state space $\{1, 2, \ldots, N\}$ and transition matrix $\boldsymbol{P}$. If we define the random vector

$$\boldsymbol{\xi}_t = \begin{cases} (1,0,0,\ldots,0,0)' & \text{for } X_t = 1 \\ (0,1,0,\ldots,0,0)' & \text{for } X_t = 2 \\ \quad\vdots & \quad\vdots \\ (0,0,0,\ldots,1,0)' & \text{for } X_t = N-1 \\ (0,0,0,\ldots,0,1)' & \text{for } X_t = N \end{cases}$$

it is straightforward to check that $\mathrm{E}[\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots] = \mathrm{E}[\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t] = \boldsymbol{P}\boldsymbol{\xi}_t$. This last consideration let us represent the Markov chain as

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{P}\boldsymbol{\xi}_t + \boldsymbol{v}_{t+1}, \tag{9}$$

with $\boldsymbol{v}_t$ martingale difference sequence with respect to the $\sigma$-algebra generated by $\{X_t, X_{t-1}, \ldots\}$. If we can observe a vector $\boldsymbol{y}_t$, which takes the value $\boldsymbol{z}_i$, $i = 1, 2, \ldots, N$ when $X_t$ is in its $i$-th state, $\boldsymbol{y}_t$ has the representation

$$\boldsymbol{y}_t = \boldsymbol{Z}\boldsymbol{\xi}_t$$

with $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N]$.

The following proposition that holds in this more general setting will be useful in determining the properties of the DDMS-VAR model.

**Proposition 1** *Let* $\{X_t\}$ *be an ergodic Markov chain with state space* $1, 2, \ldots, N$, *let* $\boldsymbol{P} = \{\Pr(X_{t+1} = i|X_t = j)\}$ *be its transition matrix and* $\boldsymbol{\pi}$ *the vector of ergodic probabilities. Then*

$$\mathrm{E}[\boldsymbol{y}_t] = \boldsymbol{Z}\boldsymbol{\pi} \tag{10}$$

$$\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t-k}] = \boldsymbol{Z}[\boldsymbol{P}^k \mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}']\boldsymbol{Z}' \tag{11}$$

**PROOF.** Using the VAR representation of the Markov chain the expectation of $\boldsymbol{y}_t$ is just

$$\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{y}_t] = \boldsymbol{Z}\mathrm{E}[\boldsymbol{\xi}_t] = \boldsymbol{Z\pi}.$$

For the cross-covariance function we have

$$\begin{aligned}
\mathrm{E}[(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_{t-k} - \boldsymbol{\mu})'] &= \mathrm{E}[(\boldsymbol{Z\xi}_t - \boldsymbol{Z\pi})(\boldsymbol{Z\xi}_{t-k} - \boldsymbol{Z\pi})'] \\
&= \boldsymbol{Z}\mathrm{E}[(\boldsymbol{\xi}_t - \boldsymbol{\pi})(\boldsymbol{\xi}'_{t-k} - \boldsymbol{\pi}')]\boldsymbol{Z}' \\
&= \boldsymbol{Z}\mathrm{E}[(\boldsymbol{\xi}_t\boldsymbol{\xi}'_{t-k}) - \boldsymbol{\pi\pi}']\boldsymbol{Z}' \\
&= \boldsymbol{Z}[\boldsymbol{P}^k\mathrm{E}(\boldsymbol{\xi}_{t-k}\boldsymbol{\xi}'_{t-k}) - \boldsymbol{\pi\pi}']\boldsymbol{Z}' \\
&= \boldsymbol{Z}[\boldsymbol{P}^k\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi\pi}']\boldsymbol{Z}'
\end{aligned}$$

□

The DDMS-VAR model has the representation

$$\boldsymbol{y}_t = \boldsymbol{Z\xi}_t + \boldsymbol{w}_t \tag{12}$$

where $\boldsymbol{w}_t$ is a stable VAR($p$) process. The Markov chain driving $\boldsymbol{\xi}_t$ is here $(S_t, D_t)$ defined in the previous section and the matrix $\boldsymbol{Z}$ has the form

$$\boldsymbol{Z} = \mathbf{1}'_\tau \otimes \Big[\boldsymbol{\mu}_0 \,|\, \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1\Big] \tag{13}$$

with $\mathbf{1}_\tau$ vector of ones of dimension $\tau$. The matrix $\boldsymbol{Z}$ associates the mean vector $\boldsymbol{\mu}_0$ to the states for which $S_t = 0$ (odd states in Table 2) and $\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1$ to the sates for which $S_t = 1$ (even states in Table 2).

Since $\boldsymbol{\xi}_t$ and $\boldsymbol{w}_t$ are independent processes, the cross-covariance function of $\boldsymbol{y}_t$ is just the sum of the cross-covariance functions of $\boldsymbol{\xi}_t$ and of $\boldsymbol{w}_t$. Since the latter is well known, we concentrate on the former and suppose that $\boldsymbol{w}_t$ in (12) is identically zero. Thus, in the following we assume

$$\boldsymbol{y}_t = \boldsymbol{Z\xi}_t.$$

The correlation structure of $\boldsymbol{y}_t$ is given by the following proposition.

**Proposition 2 (Cross-correlation function of a DDMS process)** *Under the hypotheses of proposition 1, the correlation of any element of $\boldsymbol{y}_t$ with any element of $\boldsymbol{y}_{t-k}$, with $\boldsymbol{Z}$ as in (13), is given by*

$$\mathrm{Corr}(y_{i,t}, y_{j,t}) = \frac{\boldsymbol{\zeta}'\Big[\boldsymbol{P}^k\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi\pi}'\Big]\boldsymbol{\zeta}}{\boldsymbol{\zeta}'\Big[\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi\pi}'\Big]\boldsymbol{\zeta}} \qquad \forall i, j = 1, 2, \ldots, K \tag{14}$$

*where $\boldsymbol{\zeta}$ is a $2\tau$-vector of one of the two following forms*

$$\boldsymbol{\zeta} = (1, 0, 1, 0, \ldots 1, 0)' \;\; \textit{or} \;\; \boldsymbol{\zeta} = (0, 1, 0, 1, \ldots 0, 1)'.$$

*Thus, all the auto-correlation and cross-correlation functions are equal and independent of the choice of* $(\mu_{i,0}, \mu_{i,1})$, $i = 1, \ldots, K$.

**PROOF.** Since correlations are invariant with respect to translations of the random variables, let's consider the variables

$$\tilde{y}_{i,t} = y_{i,t} - \mu_{i,0} = (\mu_{i,0}, \mu_{i,0} + \mu_{i,1}, \ldots, \mu_{i,0}, \mu_{i,0} + \mu_{i,1})\boldsymbol{\xi}_t - \mu_{i,0} = \mu_{i,1}\boldsymbol{\zeta}'\boldsymbol{\xi}_t$$

with $\boldsymbol{\zeta}' = (0, 1, 0, 1, \ldots 0, 1)$. Using proposition 1, we have

$$\begin{aligned}\mathrm{Corr}(\tilde{y}_{i,t}, \tilde{y}_{j,t-k}) &= \frac{\mu_{i,1}\mu_{j,1}\boldsymbol{\zeta}'\Big[\boldsymbol{P}^k \mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'\Big]\boldsymbol{\zeta}}{\sqrt{\mu_{i,1}^2\boldsymbol{\zeta}'\Big[\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'\Big]\boldsymbol{\zeta} \cdot \mu_{j,1}^2\boldsymbol{\zeta}'\Big[\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'\Big]\boldsymbol{\zeta}}} \\ &= \frac{\boldsymbol{\zeta}'\Big[\boldsymbol{P}^k \mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'\Big]\boldsymbol{\zeta}}{\boldsymbol{\zeta}'\Big[\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'\Big]\boldsymbol{\zeta}}.\end{aligned}$$

The proof still holds if we take $\tilde{y}_{i,t} = y_{i,t} - \mu_{i,0} - \mu_{i,1} = -\mu_{i,1}\boldsymbol{\zeta}'\boldsymbol{\xi}_t$ with $\boldsymbol{\zeta}' = (1, 0, 1, 0 \ldots 1, 0)$. □

Since the autocorrelation function of the DDMS process is a complicated function of the elements of $\boldsymbol{P}$, which in the Probit specification are functions of the parameters $\beta_i$, $i = \{1, 2, 3, 4\}$, we will rely on numerical computations to study the behavior of the relative spectral density [6].

Figure 1 shows the spectra of some symmetric DDMS models. The effect of $\beta_1$ $(= -\beta_3)$ on the spectrum may be seen in the first panel of the figure, while the consequences of changing $\beta_2$ $(= -\beta_4)$ are evident in the second panel. It is interesting to notice that the DDMS model is capable of a wide range of cyclical behaviors.

Even more interesting is the behavior of asymmetric DDMS's. As figure 2 illustrates, asymmetric DDMS's can have multi-modal spectra. This feature seems particularly useful, since (detrended) economic time series having estimated spectra with most of the power concentrated around frequency zero and a local maximum at business cycle frequencies are not rare [7].

---

[6] The existence of the spectral density is guaranteed by the geometric convergence of the Markov chain.

[7] This feature may be clearly seen, for example, in the spectrum (here not reported) of the U.S. employment data used later in this paper.

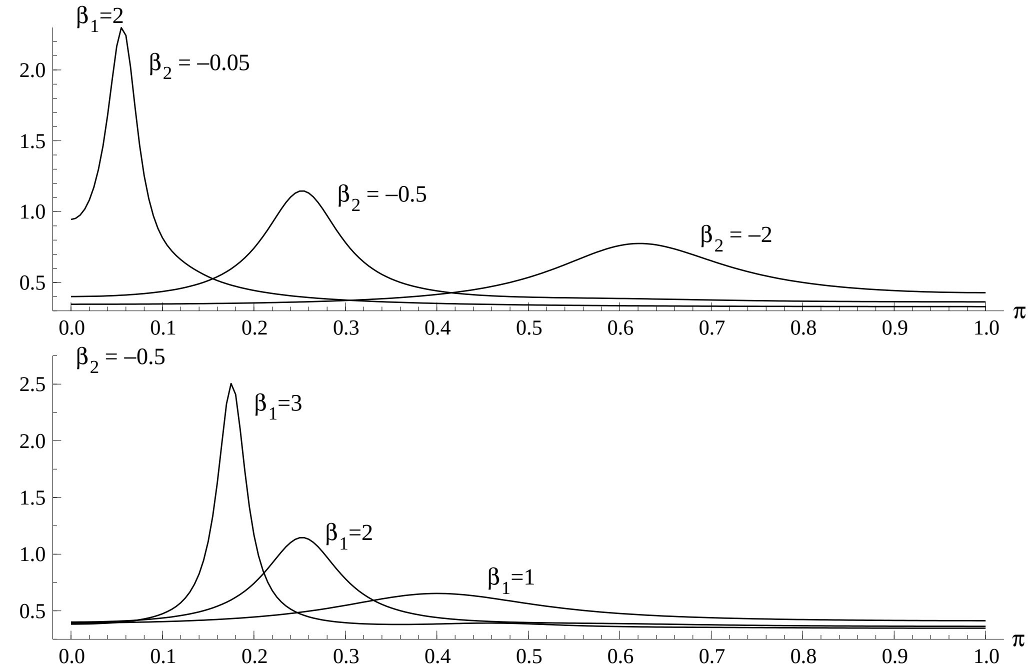


Fig. 1. Spectra of symmetrical DDMS: $\beta_1 = -\beta_3$ and $\beta_2 = -\beta_4$.

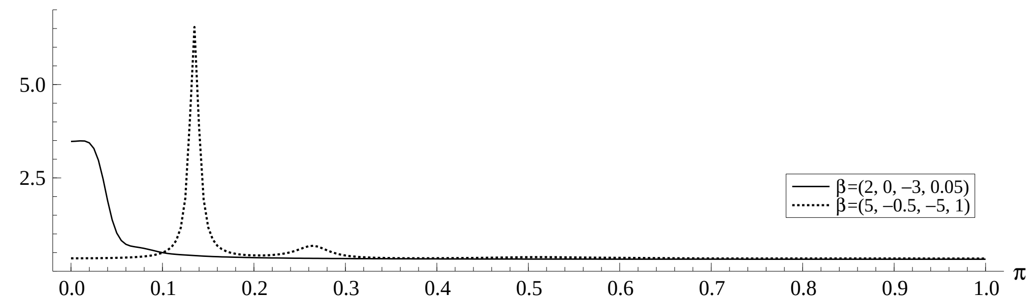


Fig. 2. Spectra of asymmetrical DDMS.

## 4 Bayesian inference on the model's unknowns

In this section it is shown how to carry out Bayesian inference on the model's unknowns

$$\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{A}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \{(S_t, D_t)\}_{t=1}^{T}),$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_0', \boldsymbol{\mu}_1')'$ and $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$, using MCMC methods.

### 4.1 Priors

In order to exploit conditional conjugacy, we use the prior joint distribution[8]

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

where

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{M}_0), \tag{15}$$

$$\text{vec}(\boldsymbol{A}) \sim \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{A}_0), \tag{16}$$

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(\text{rank}(\boldsymbol{\Sigma})+1)}, \tag{17}$$

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{b}_0, \boldsymbol{B}_0), \tag{18}$$

and $p(S_0, D_0)$ is a probability function that assigns a prior probability to every element of the state-space of $(S_0, D_0)$. Alternatively it is possible to let $p(S_0, D_0)$ be the ergodic probability function of the Markov chain $\{(S_t, D_t)\}$.

### 4.2 Gibbs sampling in short

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 in $\boldsymbol{\theta}_i$. 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 $p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \boldsymbol{Y})$, with $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ and $i = 1, \ldots, I$. A Gibbs sampler step is the generation of a random variate from $p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \boldsymbol{Y})$, $i = 1, \ldots, I$, where the elements of $\boldsymbol{\theta}_{-i}$ are substituted with the most recent sampled values of the relative variates. Since, under mild regularity 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 this stationary distribution is the "true" posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y})$, it is sufficient to fix an initial burn-in period of $M$ iterations, such that the Markov chain may virtually "forget" the arbitrary starting values $\boldsymbol{\theta}^{(0)}$, to sample from (an approximation of) the joint posterior distribution. The values obtained for each element of $\boldsymbol{\theta}$ are samples from the marginal posterior distribution of each parameters.

[8] $p(\bullet)$ denotes a generic density or probability function.

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

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

$$p(S_1^*, \ldots, S_T^* | \boldsymbol{Y}_T) = p(S_T^* | \boldsymbol{Y}_T) \prod_{t=1}^{T-1} p(S_t^* | S_{t+1}^*, \boldsymbol{Y}_t), \tag{19}$$

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

Let $\hat{\boldsymbol{\xi}}_{t|r}$ 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 $\boldsymbol{Y}_r$ and the model's parameters. Let $\boldsymbol{\eta}_t$ be the vector containing the likelihood of each state given $\boldsymbol{Y}_t$ and the model's parameters, whose generic element is

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

where

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

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

The filtered probabilities of the states can be calculated using Hamilton's filter

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

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

with the symbol $\odot$ indicating elementwise multiplication. The filter is completed with the prior probabilities vector $\hat{\boldsymbol{\xi}}_{1|0}$, that, as already remarked, can be set equal to the vector of ergodic probabilities of the Markov chain $\{S_t^*\}$.

In order to sample from the distribution of $\{S_t^*\}_1^T$ given the full information set $\boldsymbol{Y}_T$, we exploit the result

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

where $p_{i|j}$ is the transition probability of moving to state $i$ from state $j$ (element $(i, j)$ of the transition matrix $\boldsymbol{P}^*$) 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

$$\hat{\boldsymbol{\xi}}_{t|(S_{t+1}^*=i,\boldsymbol{Y}_T)} = \frac{\boldsymbol{p}_{i\bullet} \odot \hat{\boldsymbol{\xi}}_{t|t}}{\boldsymbol{p}_{i\bullet}' \, \hat{\boldsymbol{\xi}}_{t|t}} \tag{20}$$

where $\boldsymbol{p}_{i\bullet}'$ denotes the $i$-th row of the transition matrix $\boldsymbol{P}^*$.

Now all the probability functions in equation (19) have been given a form, and the states can be generated starting from the filtered probability $\hat{\boldsymbol{\xi}}_{T|T}$ and proceeding backward $(T-1, \ldots, 1)$, using equation (20) where $i$ is to be substituted with the last generated value $s_{t+1}^*$.

Once a set of sampled $\{S_t^*\}_{t=1}^T$ has been generated, it is automatically available a sample for $\{S_t\}_{t=1}^T$ and $\{D_t\}_{t=1}^T$.

The advantage of using the described multi-move Gibbs sampler, compared to the single move Gibbs sampler that can be implemented as in Carlin et al. (1992), or using the software BUGS[9], 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. Kim and Nelson (1999, section 10.3), in their outstanding monograph on state-space models with regime switching, use a single-move Gibbs sampler (12000 sample points) to achieve (almost) the same goal as in this paper, but the slow convergence properties of the single-move sampler do not give evidence in favour of the reliability of their estimates.

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

Conditionally on $\{S_t\}_{t=1}^T$ and $\boldsymbol{\mu}$ equation (1) is just a multivariate normal (auto-)regression model for the variable $\boldsymbol{y}_t^* = \boldsymbol{y}_t - \boldsymbol{\mu}_0 - \boldsymbol{\mu}_1 S_t$, whose parameters, given the discussed prior distribution, have the following posterior

[9] *http://www.mrc-bsu.cam.ac.uk/bugs/*

distributions, known in literature. Let $\boldsymbol{X}$ be the matrix, whose $t^{\text{th}}$ column is

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

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

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

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

with

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

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

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

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

where L is the lag operator, we obtain

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

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

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

Conditionally on $\{S_t^*\}_{t=1}^T$, consider the probit model described in section 2. Albert and Chib (1993) have proposed a method based on a data augmentation algorithm to draw from the posterior of the parameters of a probit model.

Given the parameter vector $\boldsymbol{\beta}$ of last Gibbs sampler iteration, generate the latent variables $\{S_t^\bullet\}$ from the respective truncated normal densities

$$Z_t|(S_t = 0, \boldsymbol{x}_t, \boldsymbol{\beta}) \sim \mathcal{N}(\boldsymbol{x}_t'\boldsymbol{\beta}, 1)\mathbb{I}_{(-\infty,0)}$$
$$Z_t|(S_t = 1, \boldsymbol{x}_t, \boldsymbol{\beta}) \sim \mathcal{N}(\boldsymbol{x}_t'\boldsymbol{\beta}, 1)\mathbb{I}_{[0,\infty)}$$

with

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)'$$
$$\boldsymbol{x}_t = (S_{t-1}, D_{t-1}, (1 - S_{t-1}), (1 - S_{t-1})D_{t-1})' \tag{21}$$

and $\mathbb{I}_{\{\bullet\}}$ indicator function used to denote truncation.

With the generated $Z_t$'s the Probit regression equation (4) becomes, again, a normal linear model with known variance.

The former Gibbs sampler steps were numbered from 1 to 4, but any ordering of the steps would eventually bring to the same ergodic distribution.

## 5 The software

DDMSVAR for Ox[10] is a software for time series modeling with DDMS-VAR processes that can be used in three different ways: (i) as a menu driven package[11], (ii) as an Ox object class, (iii) as a software library for Ox. The DDMSVAR software is freely available[12] at the author's internet site[13]. In this section I give a brief description of the software and in next section I illustrate its use with a real-world application.

### *5.1 OxPack version*

The easiest way to use DDMSVAR is adding the package to OxPack giving DDMSVAR as class name. The following steps must be followed to load the

[10] Ox (Doornik, 2001) is an object-oriented matrix programming language freely available for the academic community in its console version.
[11] If run with the commercial version of Ox (OxProfessional).
[12] The software is freely available and usable (at your own risk): the only condition is that the present article should be cited in any work in which the DDMSVAR software is used.
[13] *www.statistica.unimib.it/utenti/p_matteo/*

data, specify the model and estimate it.

### Formulate

Open a database, choose the time series to be modelled and give them the Y variable status. If you wish to specify an initial series of state variables, this series has to be included in the database and, once selected in the model variables' list, give it the State variable init status; otherwise DDMSVAR assigns the state variable's initial values automatically.

### Model settings

Chose the order of the VAR model (p), the maximal duration (tau), which must be at least[14] 2, and write a comma separated list of percentiles of the marginal posterior distributions, that you want to read in the output (default is 2.5,50,97.5).

### Estimate/Options

At the moment only the illustrated Gibbs sampler is implemented. Choose the data sample and press Options.... The options window is divided in three areas.

ITERATIONS
Here you choose the number of iteration of the Gibbs sampler, and the number of burn in iteration, that is, the amounts of start iterations that will not be used for estimation, because too much influenced by the arbitrary starting values. Of course the latter must be smaller than the former.

PRIORS & INITIAL VALUES
If you want to specify prior means and variances of the parameters to be estimated, do it in a .in7 or .xls database following these rules: prior means and variances for the vectorization of the autoregressive matrix $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_p]$ must be in fields with names mean_a and var_a; prior means and variances for the mean vectors $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ must be in fields with names mean_mu0, var_mu0,

[14] If you wish to estimate a classical MS-VAR model, choose tau $= 2$ and use priors for the parameters $\beta_2$ and $\beta_4$ that put an enormous mass of probability around 0. This will prevent the duration variable from having influence in the probit regression. The maximal value for tau depends only on the power of your computer, but have care that the dimensions of the transition matrix $u \times u$ don't grow too much, or the waiting time may become unbearable.

mean_mu1 and var_mu1; the fields for the vector $\boldsymbol{\beta}$ are to be named mean_beta and var_beta. The file name is to be specified with extension. If you don't specify the file, DDMSVAR uses priors that are vague for typical applications.

The file containing the initial values for the Gibbs sampler needs also to be a database in .in7 or .xls format, with fields a for vec($\boldsymbol{A}$), mu0 for $\boldsymbol{\mu}_0$, mu1 for $\boldsymbol{\mu}_1$, sigma for vech($\boldsymbol{\Sigma}$) and beta for $\boldsymbol{\beta}$. If no file is specified, DDMSVAR assigns initial values automatically.

SAVING OPTIONS

In order to save the Gibbs sample generated by DDMSVAR, specify a file name (you don't need to write the extension, at the moment the only format available is .in7) and check Save also state series if the specified file should contain also the samples of the state variables. Check Probabilities of state 0 in *filename.ext* to save the smoothed probabilities $\{\Pr(S_t = 0|\boldsymbol{Y}_T)\}_{t=1}^T$ in the database from which the time series are taken.

### Program's Output

Since Gibbs sampling may take a long time, after five iterations the program prints an estimate of the waiting time. The user is informed of the progress of the process every 100 iterations.

At the end of the iteration process, the estimated means, standard deviations (in the output named standard errors), percentiles of the marginal posterior distributions are given.

The output consists also of a number of graphs:

(1) probabilities of $S_t$ being in state 0 and 1,
(2) mean and percentiles of the transition probabilities distributions with respect to the duration,
(3) autocorrelation function of every sampled parameter (the faster it dies out, the higher the speed of the Gibbs sampler in exploring the posterior distribution's support, and the smaller the number of iteration needed to achieve the same estimate's precision),
(4) kernel density estimates of the marginal posterior distributions,
(5) Gibbs sample graphs (to check if the burn in period is long enough to ensure that the initial values have been "forgot"),
(6) running means, to visually check the convergence of the Gibbs sample means.

### 5.2 *The DDMSVAR() object class*

The second simplest way to use the software is creating an instance of the object DDMSVAR and using its member functions. The best way to illustrate the most relevant member functions of the class DDMSVAR is showing a sample program and commenting it.

```
#include "DDMSVAR.ox"
main() {
    decl dd = new DDMSVAR();

    dd->LoadIn7("USA4.in7");
    dd->Select(Y_VAR, {"DLIP", 0, 0, "DLEMP", 0, 0,
                       "DLTRADE", 0, 0, "DLINCOME",0 ,0});
    dd->Select(S_VAR,{"NBER", 0, 0});
    dd->SetSelSample(1960, 1, 2001, 8);

    dd->SetVAROrder(0);
    dd->SetMaxDuration(60);
    dd->SetIteration(21000);
    dd->SetBurnIn(1000);
    dd->SetPosteriorPercentiles(<0.05,50,99.5>);
    dd->SetPriorFileName("prior.in7");
    dd->SetInitFileName("init.in7");
    dd->SetSampleFileName("prova.in7",TRUE);

    dd->Estimate();

    dd->StatesGraph("states.eps");
    dd->DurationGraph("duration.eps");
    dd->Correlograms("acf.eps", 100);
    dd->Densities("density.eps");
    dd->SampleGraphs("sample.eps");
    dd->RunningMeans("means.eps");
}
```

`dd` is declared as instance of the object DDMSVAR. The first four member functions are an inheritance of the class Database and will not be commented here [15] . Notice only that the variable selected in the `S_VAR` group must contain the initial values for the state variable time series. Nevertheless, if no series is selected as `S_VAR`, DDMSVAR calculates initial values for the state variables automatically.

[15] See Doornik (2001).

SetVAROrder(const iP) sets the order of the VAR model to the integer value iP.

SetMaxDuration(const iTau) sets the maximal duration to the integer value iTau.

SetIteration(const iIter) sets the number of Gibbs sampling iterations to the integer value iIter.

SetBurnIn(const iBurn) sets the number of burn in iterations to the integer value iBurn.

SetPosteriorPercentiles(const vPerc) sets the percentiles of the posterior distributions that have to be printed in the output. vPerc is a row vector containing the percentiles (in %).

SetPriorFileName(const sFileName),
SetInitFileName(const sFileName) are optional; they are used to specify respectively the file containing the prior means and variances of the parameters and the file with the initial values for the Gibbs sampler (see the previous subsection for the format that the two files need to have). If they are not used, priors are vague and initial values are automatically calculated.

SetSampleFileName(const sFileName, const bSaveS) is optional; if used it sets the file name for saving the Gibbs sample and if bSaveS is FALSE the state variables are not saved, otherwise they are saved in the same file sFileName. sFileName does not need the extension, since the only available format is .in7.

Estimate() carries out the iteration process and generates the textual output (if run within GiveWin-OxRun it does also the graphs). After 5 iteration the user is informed of the expected waiting time and every 100 iterations also about the progress of the Gibbs sampler.

StatesGraph(const sFileName),
DurationGraph(const sFileName),

`Correlograms(const sFileName, const iMaxLag)`,
`Densities(const sFileName)`,
`SampleGraphs(const sFileName)`,
`RunningMeans(const sFileName)` are optional and used to save the graphs described in the last subsection. `sFileName` is a string containing the file name with extension (.emf, .wmf, .gwg, .eps, .ps) and `iMaxLag` is the maximum lag for which the autocorrelation function should be calculated.

### *5.3 DDMSVAR software library*

The last and most complicated (but also flexible) way to use the software is as library of functions. The DDMS-VAR library consists in 25 functions, but the user need to know only the following 10. Throughout the function list, it is used the notation below.

| | | |
|---|---|---|
| p | scalar | order of vector autoregression (VAR($p$)) |
| tau | scalar | maximal duration ($\tau$) |
| $k$ | scalar | number of time series in the model |
| $T$ | scalar | number of observations of the $k$ time series |
| $u$ | scalar | dimension of the state space of $\{S_t^*\}$ ($u = 2(2^p + \tau - p - 1)$) |
| Y | $(k \times T)$ | matrix of observation vectors ($\boldsymbol{Y}_T$) |
| s | $(T \times 1)$ | vector of current state variable ($S_t$) |
| mu0 | $(k \times 1)$ | vector of means when the state is 0 ($\boldsymbol{\mu}_0$) |
| mu1 | $(k \times 1)$ | vector of mean-increments when the state is 1 ($\boldsymbol{\mu}_1$) |
| A | $(k \times pk)$ | VAR matrices side by side ($[\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p]$) |
| Sig | $(k \times k)$ | covariance matrix of VAR error ($\boldsymbol{\Sigma}$) |
| SS | $(u \times p{+}2)$ | state space of the complete Markov chain $\{S^*\}$ (tab. 2) |
| pd | $(\text{tau} \times 4)$ | matrix of the probabilities $[p_{00}(d), p_{01}(d), p_{10}(d), p_{11}(d)]$ |
| P | $(u \times u)$ | transition matrix relative to SS ($\boldsymbol{P}^*$) |
| xi_flt | $(u \times T{-}p)$ | filtered probabilities ($[\hat{\boldsymbol{\xi}}_{t\|t}]$) |
| eta | $(u \times T{-}p)$ | matrix of likelihoods ($[\eta_t]$) |

`ddss(p,tau)`
Returns the state space SS (see table 2).

`A_sampler(Y,s,mu0,mu1,p,a0,pA0)`
Carry out step 2. of the Gibbs sampler, returning a sample point from the posterior of vec($\boldsymbol{A}$) with `a0` and `pA0` being respectively the prior mean vector and the prior precision matrix (inverse of covariance matrix) of vec($\boldsymbol{A}$).

`mu_sampler(Y,s,p,A,Sig,m0,pM0)`
Carry out step 3. of the Gibbs sampler, returning a sample point from the posterior of $[\boldsymbol{\mu}_0', \boldsymbol{\mu}_1']'$ with `m0` and `pM0` being respectively the prior mean vector and the prior precision matrix (inverse of covariance matrix) of $[\boldsymbol{\mu}_0', \boldsymbol{\mu}_1']'$.

`probitdur(beta,tau)`
Returns the matrix `pd` containing the transition probabilities for every duration $d = 1, 2, \ldots, \tau$.

$$\text{pd} = \begin{pmatrix} p_{0|0}(1) & p_{0|1}(1) & p_{1|0}(1) & p_{1|1}(1) \\ p_{0|0}(2) & p_{0|1}(2) & p_{1|0}(2) & p_{1|1}(2) \\ \vdots & \vdots & \vdots & \vdots \\ p_{0|0}(\tau) & p_{0|1}(\tau) & p_{1|0}(\tau) & p_{1|1}(\tau) \end{pmatrix}.$$

`ddtm(SS,pd)`
Puts the transition probabilities `pd` into the transition matrix relative to the chain with state space `SS`.

`ergodic(P)`
Returns the vector `xi0` of ergodic probabilities of the chain with transition matrix `P`.

`msvarlik(Y,mu0,mu1,Sig,A,SS)`
Returns `eta`, matrix of $T - p$ columns of likelihood contributions for every possible state in `SS`.

`ham_flt(xi0,P,eta)`
Returns `xi_flt`, matrix of $T - p$ columns of filtered probabilities of being in each state in `SS`.

`state_sampler(xi_flt,P)`
Carry out step 1. of the Gibbs sampler. It returns a sample time series of values drawn from the chain with state space `SS`, transition matrix `P` and filtered probabilities `xi_flt`.

`new_beta(s,X,lastbeta,diffuse,b,B0)`
Carry out step 4. of the Gibbs sampler. It returns a new sample point from

the posterior of the vector $\boldsymbol{\beta}$, given the dependent variables in `X`, where the generic row is given by (21). If `diffuse`$\neq 0$, a diffuse prior is used.

The functions of this library may be used also to carry out maximum likelihood estimation of the parameter of the DDMS-VAR model with minimum effort: an example program is available from the author.

## 6 Duration dependence in the U.S. business cycle

The model and the software illustrated in the previous sections have been applied to 100 times the difference of the logarithm of the four time series, on which the NBER mostly relays to date the U.S. business cycle, dating from January 1960 to August 2001: i) industrial production (IP), ii) total nonfarm-employment (EMP), iii) total manufacturing and trade sales in million of 1996$ (TRADE), iv) personal income less transfer payments in billions of 1996$ (INCOME).

The model estimated on these data is a DDMS-VAR(1) with diagonal autoregressive matrix and $\tau = 60$ (5 years). The choice of using the DDMS alone as the only common dynamic factor is justified by the fact that the estimates of the cospectral densities for each pair of time series have very similar behaviors.

The inference on the model unknowns is based on a Gibbs sample of 11000 points, the first 1000 of which were discarded. The autocorrelations and the kernel density estimates for each parameter are available from the author at request. All the correlations die out before the $100^{\text{th}}$ lag, thus the choice of a burn-in sample of 1000 points seems quite reasonable.

The results of an earlier experiment with $\tau = 120$ (10yrs) and $p = 0$ has not been reported: the results were quite similar to the ones reported below and the conclusions the same.

Summaries of the marginal posterior distributions are shown in table 3, while figure 3 compares the probability of the U.S. economy being in recession resulting from the model with the official NBER dating: the signal "probability of being in recession" extracted by the model here presented matches the official dating rather well, and is less noisy than the signal extracted by Hamilton (1989), based on the IP series only. The NBER dating seems to be best matched if, every time the model's probability of being in recession exceeds 0.5, the peak date is set equal the time the line crosses a low probability level (say 0.3) from below and the trough date is set equal the time the probability line crosses a high probability level (say 0.8) from above. NBER trough dates seem to be matched more frequently by the model than the peaks.

Figure 4 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.

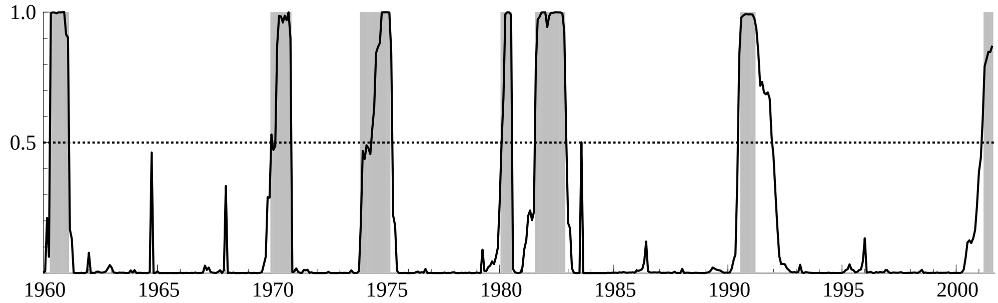


Fig. 3. (Smoothed) probability of recession (line) and NBER dating (gray shade)

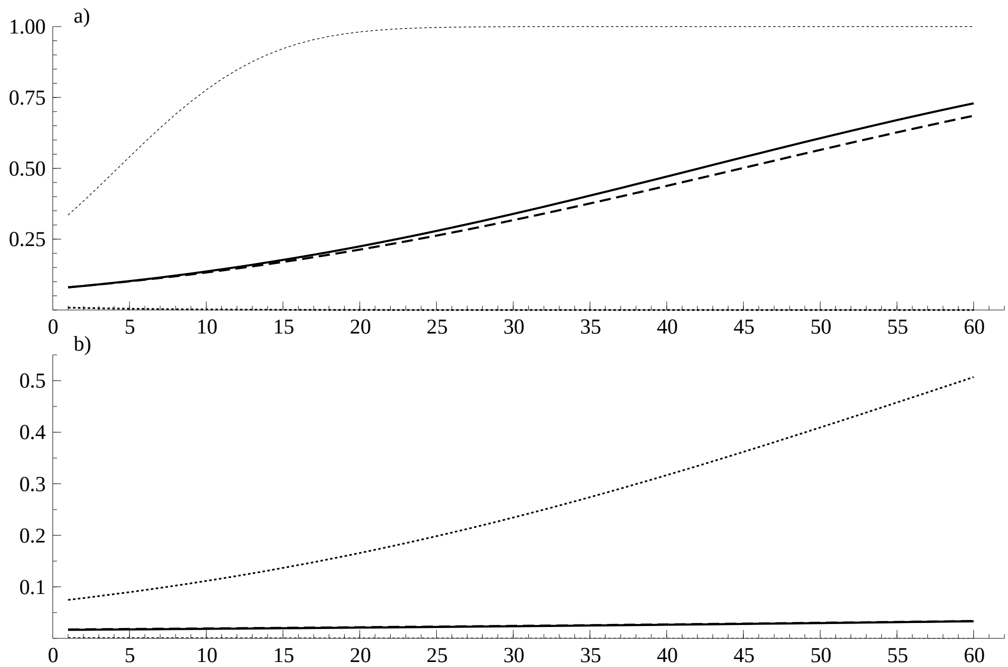


Fig. 4. Mean (solid), median (dash) and 95% credible interval (dots) of the posterior distribution of the probability of moving a) from a recession into an expansion after $d$ months of recession b) from an expansion to a recession after $d$ months of expansion

Table 3
Description of the prior and posterior distributions of the model's parameters.

| | | Prior | | Posterior | | | | |
|---|---|---|---|---|---|---|---|---|
| | Parameter | mean | var | mean | s.d. | 2.5% | 50% | 97.5% |
| $\boldsymbol{\mu}_0$ | IP | 0 | 4 | -0.584 | 0.136 | -0.873 | -0.578 | -0.335 |
| | EMP | 0 | 4 | -0.153 | 0.040 | -0.237 | -0.151 | -0.082 |
| | TRADE | 0 | 4 | -0.407 | 0.109 | -0.636 | -0.401 | -0.210 |
| | INCOME | 0 | 4 | -0.094 | 0.055 | -0.205 | -0.092 | 0.009 |
| $\boldsymbol{\mu}_1$ | IP | 0 | 4 | 1.027 | 0.139 | 0.772 | 1.022 | 1.313 |
| | EMP | 0 | 4 | 0.400 | 0.037 | 0.333 | 0.398 | 0.478 |
| | TRADE | 0 | 4 | 0.817 | 0.114 | 0.600 | 0.813 | 1.053 |
| | INCOME | 0 | 4 | 0.446 | 0.058 | 0.334 | 0.446 | 0.561 |
| $\boldsymbol{A}_1$ | IP | 0 | 1 | 0.078 | 0.040 | 0.002 | 0.077 | 0.159 |
| | EMP | 0 | 1 | 0.088 | 0.054 | -0.010 | 0.086 | 0.199 |
| | TRADE | 0 | 1 | 0.000 | 0.001 | -0.002 | 0.000 | 0.002 |
| | INCOME | 0 | 1 | -0.094 | 0.055 | -0.205 | -0.092 | 0.009 |
| $\boldsymbol{\beta}$ | $\text{Const}_0$ | 1 | 5 | 2.137 | 0.371 | 1.466 | 2.118 | 2.925 |
| | $\text{Dur}_0$ | 0 | 5 | -0.005 | 0.010 | -0.025 | -0.005 | 0.013 |
| | $\text{Const}_1$ | -1 | 5 | -1.441 | 0.437 | -2.329 | -1.435 | -0.558 |
| | $\text{Dur}_1$ | 0 | 5 | 0.034 | 0.047 | -0.054 | 0.032 | 0.132 |

## 7 Conclusions

We have analyzed the second order properties of the class of DDMS-VAR processes and proposed a Gibbs sampler and a free software for the Bayesian estimation of the unknowns.

The second order properties of the model seem to be flexible and well fit the empirical features and co-features of many (detrended) macroeconomic time series.

Once applied to four time series rather important for the dating of the U.S. business cycle, the model has proved to have a good capability of discerning recessions and expansions, as the probabilities of recession tend to assume extremely low or high values and, the resulting dating of the U.S. business cycle is very close to the official one.

As far as duration-dependence is concerned, my results are similar to those of Diebold and Rudebusch (1990), Diebold et al. (1993), Sichel (1991) and Durland and McCurdy (1994): U.S. recessions are duration dependent, while expansions seem to be not duration dependent. This could be simply due to the fact that governments are interested in exiting contractions, while the opposite is not true, and the policies they put in practice in order to achieve this goal seem effective.

The DDMSVAR software has demonstrated to work fine, even though it must be recognized that it is far from being fully optimized. Future versions will be more efficient.

The Gibbs sampling approach has many advantages but also a big disadvantage: the former are that (i) it allows prior information to be exploited, (ii) it avoids the computational problems pointed out by Hamilton (1994, p. 689) that can arise with maximum likelihood estimation, (iii) it does not relay on asymptotic inference (see note 2), (iv) the inference on the state variables is not conditional on the set of estimated parameters. The big disadvantage is a long computation time, and sometimes some numerical instability.

## References


Albert J.H., Chib S., 1993. Bayesian analysis of binary and polychotomous responce data. Journal of the American Statistical Association 88, 669–679.

Carlin B.P., Polson N.G., 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., Kohn R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.

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

Diebold F., Rudebusch G., Sichel D., 1993. Further evidence on business cycle duration dependence. In: Stock J., Watson M. (Eds), Business Cycles, Indicators and Forcasting. The University of Chicago Press: Chicago, 255–280.

Doornik J.A., 2001. Ox. An object-oriented matrix programming language. Timberlake Consultants Ltd: London.

Durland J., McCurdy T., 1994. Duration-dependent transitions in a Markov model of U.S. GNP growth. Journal of Business and Economic Statistics 12, 279–288.

Fruwirth-Schnatter S., 1994. Data Augmentation and Dynamic Linear Models. Journal of Time Series Analysis 15, 183–202.

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: Princeton.

Kim C.J., 1994. Dynamic Linear Models with Markov-Switching. Journal of Econometrics 60, 1–22.

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

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

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.