MSVARlib: a new Gauss library to estimate multivariate Hidden Markov Models

Benoît BELLONE

June 15, 2004

The views expressed in this paper are those of the author[A] and do not necessarily represent those of the French Forecasting and Economic Analysis Directorate. This working paper describes research in progress by the author and is published to elicit comments and to further debate.

[A] At the time of designing this program, the author was working at the French Forecasting and Economic Analysis Directorate, e-mail: benoit.bellone@ensae.org

(This draft June 2004)[1]

Abstract:

This paper introduces a new open source Gauss library to estimate Multivariate Hidden Markov Models (HMM) in their simpler specification. These new programs are based upon the works of Hamilton (1994) and Krolzig (1998) and allow assessment of models with 2, 3 or 4 states through classical optimization of the maximum likelihood method. The modular architecture of the program is presented in a first part. It has been designed to allow new improvements (generalized non linear MS models or enhancement to a Bayesian framework). A second part, gives some illustration through a three state model based on the American Industrial production and a new stochastic coincident indicator of a recession for the US economy, following the papers of Ferrara (2003), Bellone and Saint-Martin (2003) and Bellone (2004).

Résumé:

Cet article présente une nouvelle librairie Gauss pour estimer des modèles multivariés à changement de régime markovien dans leur version la plus simple. Elle s'inspire grandement des travaux d'Hamilton (1994) et Krolzig (1998) et permet notamment d'estimer des modèles à chaîne de markov cachée à 2, 3 ou 4 états par procédure classique de maximum de vraisemblance. Cet article explicite l'architecture modulaire de la librairie qui lui permet d'être enrichie et améliorée à l'avenir (modèles markoviens non linéaires généralisés et extension au cadre bayésien). Il illustre son utilisation à travers un modèle univarié à trois états de l'industrie américaine et un modèle multivarié à deux états basé sur les travaux de Ferrara (2003), Bellone et Saint-Martin (2003) et Bellone (2004).

Key Word: Multivariate Hidden Markov Models, Business cycle, EM algorithm, "Kitagawa-Hamilton" Filtering, Open source, Gauss library

JEL Classification: C32, E32, E44

[1] I am grateful to Vladimir Borgy for his material help during all the stages of this research, to Eric Dubois and Emmanuel Michaux for their support and their "Scilab-ism". A special thank to Fabrice Lenglart who stimulated this work and gave me access to lots of resources, for his remarks and advice. A special thank to David Saint-Martin, who helped me to develop the first univariate versions of these programs. The views expressed in this working paper are those of the author and do not necessarily represent those of the French Forecasting and Economic Analysis Directorate. This note describes research in progress by the author and is published to elicit comments, find eventual mistakes, and call for potential material collaboration in the development of this library.

Table of Contents

1. Introduction

MSVARlib 1.0 “Markov switching Vector Autoregression library” is an open-source basic package designed to model univariate or multivariate time series subject to shifts in regime. This work is fully related to the MSVAR library developed by Krolzig[2] (1998) which provides statistical tools for the maximum likelihood estimation and model evaluation of Markov-Switching Vector Autoregressions. MSVAR is a compiled program designed through Ox and not an open source package. Yet, econometricians or statisticians may be eager to control their estimate results and to master all the assumptions and implied choices related to the numerical procedures inside a program, of which algorithms and initial values. Among the available open resources in matrix language, most of programs have spread thanks to the work of Kim and Nelson (1999) and Hamilton (1989)[3], however most of their routines were not designed in a generic multivariate framework, but rather as specific univariate programs not so easy to be extended or generalized.

This is the reason why MSVARlib has been designed: it offers a full open source set of procedures dedicated to model a basic framework inspired by Hamilton (1994) and to be extended to the general framework introduced by Krolzig (1997). At last, this library was designed to propose a flexible way for applied economist to estimate Hidden Markov Models in a well-known language matrix such as Gauss. For the time being, MSVARlib does not provide a wide array of model specification regarding autoregressive, exogenous and co-integrated non-linear systems, but it could be easily extended to this framework, with some add-ins. Contrary to Ox, the library is not designed in an oriented-object environment, but still fully uses the power of a matrix language. If it has been first developed in the Gauss environment, the using of the specific Gauss subroutines and matrix operators has been minimized or reconstructed so as the library to be easily translated in other matrix languages. Note that this package should next be written to run on Scilab[4].

2. Copyright and Conditions of Use

The GAUSS programs available on the website http://bellone.ensae.net are written and Copyright © 2004 by Benoît BELLONE, all rights reserved. No part of these programs may be reproduced in any form or by any means, except for the sole purpose of academic research and of modifications or improvement for specific applications. Distribution of these programs, including in modified form, for the purpose of sale or otherwise is strictly prohibited.

If this package is open source, it should be cited whenever it is used: “any use of these programs, modified or otherwise, in published work must be acknowledged by citation of the textbook”. For example, “The authors acknowledge use of computer routines described in Bellone (2004)”. If you plan to use these programs in published work or would like to distribute these programs for research or teaching purposes, please contact Bellone via e-mail at benoit.bellone@ensae.org to seek permission in writing.

3. Disclaimer

No representation is made or implied as to the accuracy or completeness of the programs, which may indeed contain bugs or errors unknown to the author. Benoît Bellone takes no responsibility for results produced by the MSVARlib programs, which are used entirely at the reader's risk.

---

[2] Because it has fully inspired this work, careful readings of this paper should be all the more useful: http://www.economics.ox.ac.uk/research/hendry/krolzig/.
[3] See http://www.econ.washington.edu/user/cnelson for Kim and Nelson (1999) readers only, and the Hamilton’s web site: http://weber.ucsd.edu/~jhamilto/.
[4] See www.scilab.org .

This package is, by no means, finished and a comprehensive enhancement “to do list” remains to be done. If you plan to extend the library, find any problems or have suggestions for improvement please contact the author (email: benoit.bellone@ensae.org).

4. Installation and Main structures

Before installing the package and execute the applications, you need to be ale to run a GAUSS program.

4.1 Version and contributions

This package has been developed thanks to the programs of Roncalli (1995), the theoretical work of James Hamliton (1994) and Krolzig (1997). For the time being, MSVARlib is written in the GAUSS ® matrix language. It requires the GAUSS® 3.2 or a later version. It has been tested on the 3.2, 5.0 and 6.0 versions.

4.2 Installation instructions

To install and run these programs, **you must absolutely create a directory C:\GAUSS\MSVAR** where you should unzip the MSVARlib package. This package includes two directories, a readme file and this paper:

- the **MSLIB directory** includes programs and saved output results,
- the **DATA directory** includes input data files and data sample spreadsheets with convenient templates.
- Readme.txt exhibits basic installation instructions

Once unzipped, you should put the first two directories in the following subdirectories:

C:\GAUSS\MSVAR\MSLIB

C:\GAUSS\MSVAR\DATA

To start estimation, open the Gauss® program, **select C:\GAUSS\MSVAR\MSLIB as your working library**. Three programs are available: “MSVAR.prg” which is the main program used to run a generic multivariate Markov switching estimate and two specific programs: Rec.prg and IPI.prg, written on this template to provide examples.

4.3 Main files and data organization

In the DATA directory, some sample ASCII files[5] are available. These data, saved in the “.txt” format with tab-spaces separators, are based upon American survey and quantitative time series starting from January 1960 to February 2004[6]. The first two columns of an input file should start with a month series (m format: 1 to 12) and a year series (yyyy format), the following columns should refer to data. Note that no label should be referenced[7] and that a missing value is represented by a dot **“.”**.

Here is a list of included sample files in the DATA directory:

- MSVARrec4.txt and MSVARAnas.txt deal with the four series used in Bellone (2004) and in Bellone, Saint - Martin (2003) as a US recession Hidden Markov Models: they are related to the “March 2003 vintage”. Those four series are: the reversed unemployment

[5] USDatation.txt includes the NBER dates (dummy =1 during recession periods), and lagged NBER dates, the last series corresponds to the ECR growth cycle dates. This file is facultative.
[6] The template is extended with missing values to December 2004.
[7] We preferred the following solution: write a buffer file in a spreadsheet allowing a matching between series, label and dates. cf Excel spreadsheets.

rate, the help wanted advertising index, the Industrial Production index, 1000/Jobs hard to Get component.

- MSVARrec3.txt includes the three first indicators for the "March 2003 vintage",

- MSVARrec4_2.txt and MSVARrec3_2.txt include the same series but refer to an updated "April 2004 vintage", which allows to assess the possible impact of revisions between March 2003 and February 2004.

- MSVARUN.txt, MSVARHW.txt, MSVARIPI.txt, include the first three series related to the "February 2004 vintage".

- On top of these files a spreadsheets MSVARlib_data.xls include the whole data set and a utility to print output which are saved in the DATA directory, of which the final parameters (parameters.txt) and unconstrained parameters (paramfin.txt), the filtered and smoothed probabilities (filtprob.txt, smooprob.txt, statistics and standard deviations (Res1.txt, Res2.txt) can be imported with a macro (see the "import" sheet).

In the MSLIB directory, you will find the whole set of program of which:

- MSVAR.prg – main program
- MSVAR_call.prg – general program calls to subroutines

Subroutines dedicated to Input / Output relations and "EM algorithm" estimation programs:

- MSVAR_load.prg: generates input files,
- MSVAR_Setsample.prg and MSVAR_Setdatation.prg: perform transformation of input series such as differences, dlog(), level, standardization,
- MSVAR_moment.prg: computes descriptive statistics, sorts series and prepares automatic initializations,
- MSVAR_Init: present automatic initialization of parameters subroutines,
- MSVAR_FiltHmm.prg and MSVAR_SmoHmm.prg: generate Kitagawa-Hamilton Markov Switching filter and smoother,
- MSVAR_MaxHmm.prg defines loglikelihood and maximum likelihood routines,
- MSVAR_stderr.prg: computes non-linear "Wald" tests and standard deviations, with associated statistics (t-stats, p value…).

5. The basic multivariate Makov-switching framework

For a comprehensive presentation of Markov-switching vector auto-regression models, the reader should report to Krolzig (1997) and for a synthetic and full presentation of family of regime switching models with Ox to Krolzig (1998). For a basic presentation of Hidden Markov Models and for a more comprehensive study of gaussian distribution mixtures one should report to Hamilton (1994).

Please, note that this library covers a more limited range of models than Krolzig's compiled programs but introduce a quite generic open source framework, that should be easy to extend to more complex specifications such as MSI-M(H)-VAR(P) or pure non linear MS Models.

Be $Y_t^* = \left(y_{1,t}^*,..., y_{K,t}^*\right)'$ a vector[8] of size (K, 1), with K the indices designating the number of variables upon which the indicator is based.

We define $S_t = \{1,.., M\}$ as a "M state unobserved variable"[9], following a first order Markov Chain. When $S_t = 1$ ($S_t = M$), the time series are said to be in the "lowest" ("highest") regime. The matrix P of transition probabilities is defined very classically:

(1) $P = (p_{kj})$, $\forall k, j \in \{1 \cdots M\}$, with:

$$p_{kj} = P(S_t = j / S_t = k,..., S_1 = l, I_{t-1}) = P(S_t = j / S_{t-1} = k) \text{ and } \sum_{k=1}^{M} p_{kj} = 1 \text{ such that}$$

$I_{t-1} = (Y_{t-1}^*, \ldots, Y_1^*)$ is the information set available in t-1. Following Hamilton's notations, we define the "shadow variables":

(2). $\xi_t^j = 1\{S_t = j\}$ with $\forall j \in \{1 \cdots M\}$ and the associated vector $\xi_t' = (\xi_t^1, \ \ldots \ \xi_t^M)'$ such that

(3). $P(S_t = j / I_t) = E(\xi_t^j | I_t)$ and $P(S_t / I_t) = \hat{\xi}_{t/t} = E(\xi_t | I_t) = (E(\xi_t^1 | I_t), \ \ldots, \ E(\xi_t^M | I_t))'$

Following the definition of Krolzig (1998), MSVARlib allows then to estimate the multiple time series MSI(M)AH-VAR(0), with both possibilities of heteroscedasticity / homoscedasticity and switching variances.

- **The "heteroscedastic specification":**

(4) $Y_t^* = \mu_{S_t} + \eta_{S_t} = \begin{bmatrix} \mu_{S_t}^1 \\ \vdots \\ \mu_{S_t}^K \end{bmatrix} + \begin{bmatrix} \eta_{S_t}^1 \\ \vdots \\ \eta_{S_t}^K \end{bmatrix}$, where $\eta_{S_t} \sim N(0, \Sigma_{S_t})$ such that:

$\Sigma_{S_t}^{-1/2} \eta_{S_t} = u_t$ is a conditionnal gaussian uncorrelated white noise.

(5) $\mu_{S_t}^i = \mu_M^i \xi_t^M + \mu_{M-1}^i \xi_t^{M-1} + ... + \mu_1^i \xi_t^1$ where $\mu_1^i < ... < \mu_M^i$, $\forall i = 1 \cdots K$, and

(6) $\Sigma_{S_t} = \begin{pmatrix} \sigma_1^2(S_t) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_K^2(S_t) \end{pmatrix}$ with

(7) $\sigma_i^2(S_t) = \sigma_{i,M}^2 \xi_t^M + \sigma_{i,M-1}^2 \xi_t^{M-1} + ... + \sigma_{i,1}^2 \xi_t^1$, $\forall i = 1 \cdots K$.

---

[8] The series may be standardized thanks to some option in MSVAR_Setsample.

[9] For the time being automatic specification are designed to estimate 2, 3 and 4 state models because of economic and pragmatic use. However, it is really easy to extend those simple limitations.

- **The “homoscedastic specification”:**

$$(8)\ \Sigma_{S_t} = \begin{pmatrix} \sigma^2(S_t) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2(S_t) \end{pmatrix} = \sigma^2(S_t).I_K .$$

Of course, the library allows a more parcimonious specification such as the covariance invariant option:

(9) $Y_t^* = \mu_{S_t} + \eta_t = \begin{bmatrix} \mu_{S_t}^1 \\ \vdots \\ \mu_{S_t}^p \end{bmatrix} + \begin{bmatrix} \eta_t^1 \\ \vdots \\ \eta_t^p \end{bmatrix}$ with $\eta_t \sim N(0,\Sigma)$ and Σ specified as diagonal, heteroscedastic or homoscedastic, but as an invariant covariance matrix.

- **A few words about “Filters and smoothers”:**

In the most general specification, all parameters are conditioned by S_t such that M different (Kx1) vectors can be considered. Thus, for a given regime $S_t = s_t = j$,

(10) $Y_t^* / s_t = \mu_j + \eta_t / s_t$ and $\eta_t / s_t \sim N(0,\Sigma_j)$.

If we call $\theta = ((p_{i,j}), \mu_1,..., \mu_M, \Sigma_1,..., \Sigma_M)'$ the (M x (M-1) +2 x K x M, 1) [10] vector of parameters to be estimated, the conditional probability density function of Y_t^* is then denoted by:

$$(11)\ f\left(Y_t^* / S_t = j, I_{t-1}, \theta\right) = (2\pi)^{-K/2} \det(\Sigma_j^{-1/2}) \exp\left(-\frac{\left(Y_t^* - \mu_j\right)' \Sigma_j^{-1} \left(Y_t^* - \mu_j\right)}{2} \right)$$

To a certain extent, the models can then be represented as a M “stacked” vector of conditional gaussian densities estimated by maximum likelihood with the following properties linked to the Kitagawa-Hamilton filter.

To summarize the logic of the MSVAR main program:

- First, initial values of the vector θ of parameters are computed[11]. By default, an automatic procedure implement initial conditions: the series are sorted, split in M parts on which initial conditional moments are computed to run the Maximum likelihood descent.

[10] There are M x (M-1) parameters associated to the transition matrix (M identification constraint are used to identify the last line of the matrix), K x M means and K x M variances to estimate.

- Second, the model is recursively estimated from the density built as a mixed-conditional distribution:

$$(12)\quad \begin{cases} f\left(Y_t^* / I_{t-1}, \theta\right) = \sum_{j=0}^{M} P\left(S_t = j, Y_t^* / I_{t-1}, \theta\right) \\ \qquad = \sum_{j=0}^{M} P\left(S_t = j / I_{t-1}, \theta\right) \bullet f\left(Y_t^* / S_t = j, I_{t-1}, \theta\right) \end{cases}$$

where ● (/●) represents the element by element product (the element by element division) (**.*** and **./** in the Gauss matrix language).

Thus the filtered probability is computed such as:

$$(13)\quad \begin{cases} P\left(S_t = j / I_t, \theta_i\right) = P\left(S_t = j / Y_t^*, I_{t-1}, \theta\right) = \dfrac{P\left(S_t^i = j, Y_t^* / I_{t-1}, \theta\right)1}{f\left(Y_t^* / I_{t-1}, \theta\right)} \\ \qquad = \dfrac{P\left(S = j / I_{t-1}, \theta\right) \bullet f\left(Y_t^* / S_t = j, I_{t-1}, \theta\right)}{\sum_{k=1}^{M} P\left(S_t = k / I_{t-1}, \theta\right) \bullet f\left(Y_t^* / S_t = k, I_{t-1}, \theta\right)} \end{cases}$$

Following the notations of Krolzig or Hamilton, the algorithm replicates the generic following formula to extract a M x T matrix of filtered probabilities, the components of which are:

$$(14)\quad P\left(S_t / I_t\right) = \hat{\xi}_{t/t} = \frac{f\left(Y_t^* / S_t\right) \bullet \hat{\xi}_{t/t-1}}{1'_M \left(f\left(Y_t^* / S_t\right) \bullet \hat{\xi}_{t/t-1}\right)}$$ [12]

A novel aspect of the MSVARlib package is also to define a more concise formulation of the smoothing algorithm in a multivariate specification, that of course applies in an univariate one through the recursive formula:

$$(15).\quad P\left(S_t / I_T\right) = \left(P\left(S_t / I_t\right)' \bullet \left\{P'\left[P\left(S_{t+1} / I_T\right)' / \bullet P\left(S_{t+1} / I_t\right)'\right]\right\}\right)', \quad \forall t = 1 \cdots T-1 .$$

For a comprehensive treatment of Baum-Lindgren-Hamilton-Kim (BLHK) filters and smoothers in a multivariate framework, the reader should report on Krolzig (1997), chap 5, and Hamilton (1994), chap22.

6. Two illustrations of MSVARlib use.

For convenience, the default option to set the sample is to establish a pre-standardization of data, however this option can easily be turned out in the MSVAR_setsample subroutine.

[11] This part of the programs is really important in the framework of the EM algorithm. MSVARlib allows manual initialization or automatic proceedings. The later one is a little more complicated but should be comprehensively studied.
[12] Note that equation (14) is a simple rewriting of stacked equation (13).

6. 1 A three-state univariate Markov switching model on the US Industrial Production.

In the "IPI.prg" program file, we assume that the US industrial production is subject to three regimes: contraction, above trend and upper trend growths, with possible heteroscedasticity and switch in variances between regimes. Starting with the MSVARIPI.txt file, the code is reported in IPI.prg. Which modifications were then implemented to reproduce this program with MSVAR.prg ?

- **Dealing with series and sample files:**

_ncol=3; /* *2 (month* and year) + the number of series, here equal to1 */
Data*_file="MSVARIPI";*

- **Model specification:**

First, we have to choose the number of states (here 3):
_M=3;

then, are there switch in variances ?

_M_V=_M; /* *_M_V=1 no switch in variance if _MV_=_M (switch in variance) */*

The heteroscedastic or homoscedastic option has then to be specified[13]:

_Var_opt=1*; (heteroscedastic case)*

Then, one has to choose lags (k=2) and transformation (here dlog(.)):

F_option=1;
k_lag=2;

So, in this case, starting from the $Y_i = (Y_{i,1}, \ldots, Y_{i,T})'$ series, final standardized input data $y^*{}_i = (y^*{}_{i,1}, \ldots, y^*{}_{i,T})'$ are computed by this way:

$$y^*{}_{i,t} = \frac{(y_{i,t} - m_{y_i})}{\sigma_{y_i}},$$

with $y_{i,t} = T_k(L)F(Y_{i,t}) \times 100 = \left(\log(Y_{i,t}) - \log(Y_{i,t-k})\right) \times 100$ and m_{y_i} and σ_{y_i} the corresponding empirical mean and standard deviation of growth rates of the series i.

- **Periods of estimate and/or out of sample filtering:**

Follow the initialization of periods, we recall that 1 corresponds to Jan 1960, 86 to February 1967[14],

deb=86; ie (1967-2)
fin=480; ie (1999-12)[15].

---

[13] It does not matter in this case, but rather in a multivariate framework, see the next example below for the multivariate specification).
[14] See the Excel® spreadsheet MSVAR_data.xls for correspondence with templates and different vintages.
[15] Deb (fin), for starting (ending) period of estimates or filtering.

- **Estimation or Out-of sample filtering with imported parameters:**

At last, the option estimation (estim =1)[16] has to be chosen. Output and graphics are presented in Appendix A[17]. The estimated periods are the same (1967-2 / 1999-12, that is deb=86 /fin= 480). In a second round one may filter with previous estimates on an extended range ending in March 2003 (fin=517). Two options must be modified: extending the end of sample for instance from 480 to 517 and turn on the option "estim" from 1 to 0. By running then a second time the program, a set of parameters[18] is imported.

At this stage, some points must be underlined:
- as expected, variances in the "low" regime are bigger than in the "growth" regime,
- NBER recessions are well detected with the "low regime" phases,
- this simple model tracks quite faithfully acceleration and slowdown phases and remains a simple but efficient tool of qualitative short term analysis of the US growth cycle,
- three-state models are extremely sensitive to the sample period, which denotes a like of robustness and an instability between the middle and high regime[19]. The reader should experience varying estimation periods and judge the changing of significance tests.

6. 2 A two-state multivariate Markov switching model of US recessions

The Rec.prg file, elicited from Bellone (2004), gives another example of a multivariate HM model implementation. It is a "two state model" rather than a "three-state HMM' because of higher stability and robustness to changes in sample periods. One should pay attention to the number of series (here 4 +2 =6), to the new file name: "MSVARrec4" and options linked to the model. The estimated periods are the same (1967-2 / 1999-12) and the out of sample filtering ends in March 2003 (fin=517). We don't give much information, comments inside the program should be judged sufficient now. Note that one should pay attention to the difference between the switch in variance option (_M=_M_V) and the heteroscedastic / homoscedastic specification (_varopt=1 / 2)[20]. Both are complementary as in the Rec.prg specification. Outputs are presented in appendix B. Here we also present both smoothed probabilities and filtered probabilities, generated through an out-of sample filtering (estim = 0 in the next step).

7. Conclusion and "to do list".

MSVARlib is a far more modest package than Krolzig's "MSVAR for Ox" compiled programs. It was not designed to offer a better tool, but rather to provide convenient and open source programs to estimate, through an "EM algorithm", basic and generic Multivariate Hidden Markov Models. As underlined in Albert and Chib (1993), Lahiri and Wang (1994), Chauvet and Pigger (2003 and Ferrara (2003), those basic models with no autoregressive components remain more robust in term of detection and more parcimonious than the sophisticated classes of models such as MSM/I(H)-VAR(p). Of course, there is no denying that the assumption of pure gaussian white noise residuals is often

[16] If there is no parameter available in library the first time you launch a model, estim=0 will send an error message.
[17] Correlation matrices of parameters are omitted but can be printed out.
[18] Each parameters are saved in C:\GAUSS\MSVAR\DATA\parameter.txt
[19] To that extent, there are signs of a structural break at the beginning of the 80's, which is quite common in the literature, and may justify to start the estimation around 1984-1985 as in Ferrara (2003) for three-state models.
[20] Being in a multivariate framework, variances differ from one another in two dimensions: the regime and the series.

wrong (a Jarque and Berra subroutine is implemented in MSVARlib to check these results). Nevermind, we prefer a better and timely detection than noisy , distorted or lagged results.

Nonetheless, the MSVARlib package has been designed to evolve and is prepare to enhance its family models framework. If I had to draw "a to do list", here could be the following adds and improvement:
-a clearing and optimization of algorithms (there are undoubtedly some gains to make),
-an extension to MSIH(M)-VAR or MSH(M)-VAR models such as :

$$A_{S_{t_t}}(L)Y_t^* = \mu_{S_t} + \eta_{S_{t_t}} \text{ or } A_{S_{t_t}}(L)\left(Y_t^* - \mu_{S_t}\right) = \eta_{S_t} ,$$

-a development of a Multivariate Bayesian framework with a Multimove Gibbs sampler defined in Kim and Nelson (1999) or "Particle Filters" such as designed in Chopin (2001) should be useful to improve the estimation process,°

-an extension to pure linear (which encompass the former model) or non linear Markov Switching models such as :

$$A_{S_{t_t}}(L)Y_t^* = X_t \beta_{S_t} + \eta_{S_t} \text{ and } A_{S_{t_t}}(L)Y_t^* = \psi\left(X_t, \beta_{S_t}\right) + \eta_{S_t} ,$$

-a development of augmented diagnosis tools on distributions (correlograms, spectral densities,…)
-an implementation of tests for breaking and co-breaking, and to a lesser extent Forecasting and co-integration analysis tools[21].

[21] However I prefer MS models for their ability to deliver qualitative message whereas I am dubious on their global performance as significant better predictors than linear models.

8. References

Albert, J H. et Chib, S. (1993), "Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts", *Journal of Business and Economic Statistics*, 11, pp. 1-15.

Anas, J. and Ferrara, L. (2002), "Un Indicateur d'Entrée et de Sortie de Récession: Application aux Etats-Unis", Centre d'Observation Economique, Paris, Working paper.

Andersson, E. et Bock, D., Frisen, M. (2002), "Detecting of Turning Points In Business Cycles", 26th CIRET Conference, Taipei , Working paper.

Bellone, B., et Saint-Martin, D. (2003), "Detecting turning points with many predictors through hidden markov models", Working paper presented in Séminaire Fourgeaud December, 3rd 2003: "Etudes pour la conjoncture".

Bellone, B. (2004), "Une lecture probabiliste du cycle d'affaires américain", Working paper.
Chauvet, M. et Pigger, J M. (2003), "Identifying Business Cycle Turning Points in Real Time", *Review of the Federal Reserve Bank of St. Louis*, pp. 47-62.

Chopin, N. (2001), "Sequential Inference and State Number Determination for Discrete State-Space Models through Particle Filtering ", *Document de travail du CREST,* 2001-34.

Ferrara, L. (2003), "A Three-Regime Real-Time Indicator for the US Economy", *Economics Letters*, (81), Issue 3, December 2003, pp 373-378.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle", *Econometrica*, 57 pp. 357-84.

Hamilton, J. D. (1994), "Time Series Analysis", Princeton University Press, Princeton, Chap 22.

Kim, C. J., and Nelson, C. R. (1999), "State-space Models with Regime Switching", MIT Press ed.

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

Krolzig, H. M. (1998), "Econometric Modelling of Markov-Switching Vector Autoregressions using MSVAR for Ox", Working paper.

Krolzig, H. M. (2003), "Constructing Turning Point Chronologies with Markov-Switching Vector Autoregressive Models: the Euro-Zone Business Cycle",Working paper.

Lahiri, K., et Whang J. G. (1994) "Predicting Cyclical Turning Points with leading index in the Markov Switching model.", *Journal of Forecasting*, vol. 13, pp. 245-263.

Roncalli, T., (1995) "Introduction à la programmation sous GAUSS, Applications à la Finance et à l'Econométrie", vol 2, Ritme informatique ed.

9. Appendix A:
a three-state univariate model on Industrial production

```
============== MSVAR estimation with MSVARlib - Copyright (C) 2004 by Benoit BELLONE =============

Initial statistical moments and  parameters


mu_out:
    -1.01256799      0.07832529      0.93424269

cov_out':

     0.83585450
     0.02895431
     0.22331941

var_out:
     0.83585450      0.02895431      0.22331941



Initial real and transformed parameters, para_m and param_init:

     4.00000000      0.95373160
     0.50000000      0.02880020
    -1.00000000      0.02714903
     2.50000000      0.89905227
    -4.50000000      0.01016196
    -2.50000000      0.07508731
    -1.01256799     -1.01256799
     0.07832529      0.07832529
     0.93424269      0.93424269
     0.91425079      0.83585450
     0.17015966      0.02895431
     0.47256683      0.22331941

Initial matrix of transition markovian probabilities, ptrans_init:

     0.95373160      0.02714903      0.01016196
     0.02880020      0.89905227      0.07508731
     0.01746820      0.07379870      0.91475073

Initial conditional state intercepts, mu_init:
    -1.01256799      0.07832529      0.93424269

Initial ergodic state probabilities, prob_st_init:

     0.28585572
     0.35138388
     0.36276040

Initial state covariance, var_init:
     0.83585450      0.02895431      0.22331941


det_var_init:
     1.19638047     34.53717237      4.47789105


 Optimization step, be patient ...

……….


================================================================================
   iteration:  63
   algorithm: BFGS         step method: STEPBT
   function:  424.21786    step length:  1.00000      backsteps:  0
-------------------------------------------------------------------------------
   param.      param. value     relative grad.
      1           2.0146            0.0000
```

2	-4.7860	0.0000
3	-0.6977	0.0000
4	2.7685	0.0000
5	-7.6903	0.0000
6	-1.7384	0.0000
7	-1.5858	0.0000
8	-0.0835	0.0000
9	0.8424	0.0000
10	1.1674	0.0000
11	0.4765	0.0000
12	0.5871	0.0000

=================Final Results of the Maximum likelihood optimization ==================

Log likelihood -424.21785895
Convergence code : 0.00000000
The EM algorithm has converged.

Degree of freedom 381.00000000

Number of observations: 393
Number of parameters: 12

	Estimates	Gradient	Standard-errors	T-student	Pvalue.
X01	0.881444	-0.000083	0.046751	18.853866	0.000000
X02	0.001000	0.000001	0.000000	+INF	0.000000
X03	0.028550	0.000051	0.012999	2.196339	0.028670
X04	0.914066	-0.000055	0.031257	29.243218	0.000000
X05	0.001000	-0.000000	0.000000	+INF	0.000000
X06	0.149457	0.000048	0.049531	3.017438	0.002720
X07	-1.585836	0.000053	0.175325	-9.045112	0.000000
X08	-0.083606	-0.000778	0.082414	-1.014457	0.311009
X09	0.842364	0.000334	0.126555	6.656109	0.000000
X10	1.362744	-0.000061	0.271729	5.015083	0.000001
X11	0.226984	-0.002646	0.030986	7.325424	0.000000
X12	0.344645	0.000971	0.051444	6.699365	0.000000

====================Final parameters========================

Final matrix of markovian transition probabilities, ptrans_res:

0.88144414	0.02855006	0.00100000
0.00100000	0.91406642	0.14945735
0.11755586	0.05738352	0.84954265

Final conditional intercepts, mu_res :
-1.58583555 -0.08360583 0.84236358

Final ergodic probabilities, prob_st_res :

0.13504900
0.54975922
0.31519178

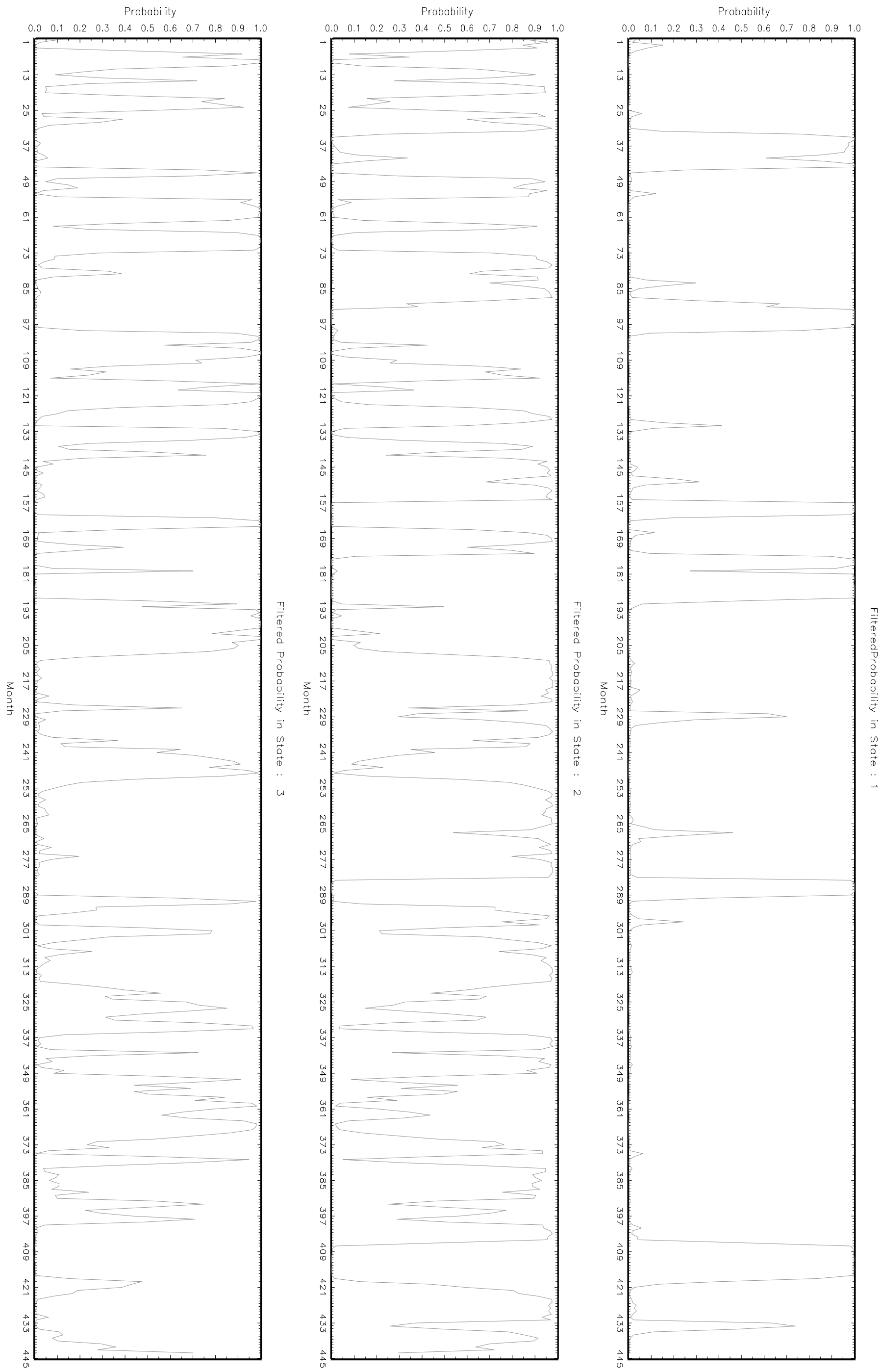
Final transposed conditional covariances, var_res' :

1.36274407
0.22698404
0.34464482

det_var_res':

0.73381350
4.40559614
2.90153789

Figure 1
A three state US IPI HMM- Filtered probabilities of low, middle and high regime
In sample estimates1967-2 / 1999-12 - Out of sample filtering 1967-2 / 2004-2



10. Appendix B:
a two-state multivariate Markov switching Recession model

============= MSVAR estimation with MSVARlib - Copyright (C) 2004 by Benoit BELLONE ============

Initial statistical moments and parameters

mu_out:

-0.73527688	0.73154451
-0.75954335	0.75568780
-0.69673890	0.69320215
-0.75803309	0.75418521

cov_out':

0.71912897	0.64376261	0.73145154	0.64859232
0.64376261	0.58609999	0.65922350	0.58857086
0.73145154	0.65922350	0.77142793	0.66617547
0.64859232	0.58857086	0.66617547	0.59802366
0.20132401	0.22289195	0.22358295	0.22322686
0.22289195	0.26168115	0.25767551	0.25510213
0.22358295	0.25767551	0.25882553	0.25513604
0.22322686	0.25510213	0.25513604	0.25436703

var_out:

0.71912897	0.20132401
0.58609999	0.26168115
0.77142793	0.25882553
0.59802366	0.25436703

Initial real and transformed parameters, para_m and param_init:

3.50000000	0.97068777
2.50000000	0.92414182
-0.73527688	-0.73527688
-0.75954335	-0.75954335
-0.69673890	-0.69673890
-0.75803309	-0.75803309
0.73154451	0.73154451
0.75568780	0.75568780
0.69320215	0.69320215
0.75418521	0.75418521
0.84801472	0.71912897
0.76557167	0.58609999
0.87830970	0.77142793
0.77331989	0.59802366
0.44869144	0.20132401
0.51154780	0.26168115
0.50874899	0.25882553
0.50434812	0.25436703

Initial matrix of transition markovian probabilities, ptrans_init:

0.97068777	0.07585818
0.02931223	0.92414182

Initial conditional state intercepts, mu_init:

-0.73527688	0.73154451
-0.75954335	0.75568780
-0.69673890	0.69320215
-0.75803309	0.75418521

Initial ergodic state probabilities, prob_st_init:

0.72128824
0.27871176

Initial state covariance, var_init:

0.71912897	0.00000000	0.00000000	0.00000000	0.20132401	0.00000000	0.00000000	0.00000000
0.00000000	0.58609999	0.00000000	0.00000000	0.00000000	0.26168115	0.00000000	0.00000000
0.00000000	0.00000000	0.77142793	0.00000000	0.00000000	0.00000000	0.25882553	0.00000000
0.00000000	0.00000000	0.00000000	0.59802366	0.00000000	0.00000000	0.00000000	0.25436703

det_var_init:
5.14289647 288.31289466

Optimization step, be patient ...
….

==
iteration: 74
algorithm: BFGS step method: STEPBT
function: 1857.94871 step length: 1.00000 backsteps: 0
----

param.	param. value	relative grad.
1	1.9723	0.0000
2	3.7963	0.0000
3	-1.5512	0.0000
4	-1.4193	0.0000
5	-1.3772	0.0000
6	-1.3726	0.0000
7	0.2889	0.0000
8	0.2643	0.0000
9	0.2565	0.0000
10	0.2556	0.0000
11	0.9795	0.0000
12	0.9208	0.0000
13	1.2615	0.0000
14	0.9737	0.0000
15	0.6877	0.0000
16	0.7618	0.0000
17	0.6840	0.0000
18	0.7684	0.0000

=================Final Results of the Maximum likelihood optimization ==================

Log likelihood -1857.94870592
Convergence code : 0.00000000
The EM algorithm has converged.

Degree of freedom 375.00000000

Number of observations: 393
Number of parameters: 18

	Estimates	Gradient	Standard-errors	T-student	Pvalue.
X01	0.877855	0.000062	0.043585	20.141237	0.000000
X02	0.978041	0.000036	0.008800	111.144725	0.000000
X03	-1.551234	-0.000434	0.137491	-11.282437	0.000000
X04	-1.419245	0.001388	0.133002	-10.670835	0.000000
X05	-1.377209	-0.000259	0.169229	-8.138117	0.000000
X06	-1.372624	-0.000145	0.129314	-10.614676	0.000000
X07	0.288877	-0.000729	0.039196	7.370145	0.000000
X08	0.264303	-0.000177	0.042605	6.203627	0.000000
X09	0.256470	-0.000161	0.039061	6.565884	0.000000
X10	0.255618	0.000128	0.043935	5.818042	0.000000
X11	0.959404	-0.000064	0.182350	5.261322	0.000000
X12	0.847892	-0.000061	0.162634	5.213509	0.000000
X13	1.591501	0.000001	0.304801	5.221444	0.000000
X14	0.948199	0.000342	0.173764	5.456812	0.000000
X15	0.472970	-0.004824	0.038068	12.424419	0.000000
X16	0.580346	0.007985	0.045509	12.752372	0.000000
X17	0.467836	-0.004815	0.037901	12.343543	0.000000
X18	0.590417	-0.004219	0.047889	12.328837	0.000000

====================Final parameters========================

Final matrix of markovian transition probabilities, ptrans_res:

0.87785543	0.02195929
0.12214457	0.97804071

Final conditional intercepts, mu_res :

-1.55123400	0.28887732
-1.41924505	0.26430312
-1.37720859	0.25647014
-1.37262440	0.25561773

Final ergodic probabilities, prob_st_res :

0.15238519
0.84761481

Final transposed conditional covariances, var_res' :

0.95940373	0.00000000	0.00000000	0.00000000
0.00000000	0.84789188	0.00000000	0.00000000
0.00000000	0.00000000	1.59150137	0.00000000
0.00000000	0.00000000	0.00000000	0.94819916
0.47297008	0.00000000	0.00000000	0.00000000
0.00000000	0.58034588	0.00000000	0.00000000
0.00000000	0.00000000	0.46783570	0.00000000
0.00000000	0.00000000	0.00000000	0.59041688

det_var_res':

0.81461338
13.18946864

Figure 2
A two state Recession model- Filtered probabilities of low and high regime
In sample estimation 1967-2 / 1999-12 - Out of sample filtering 1967-2 / 2003-3

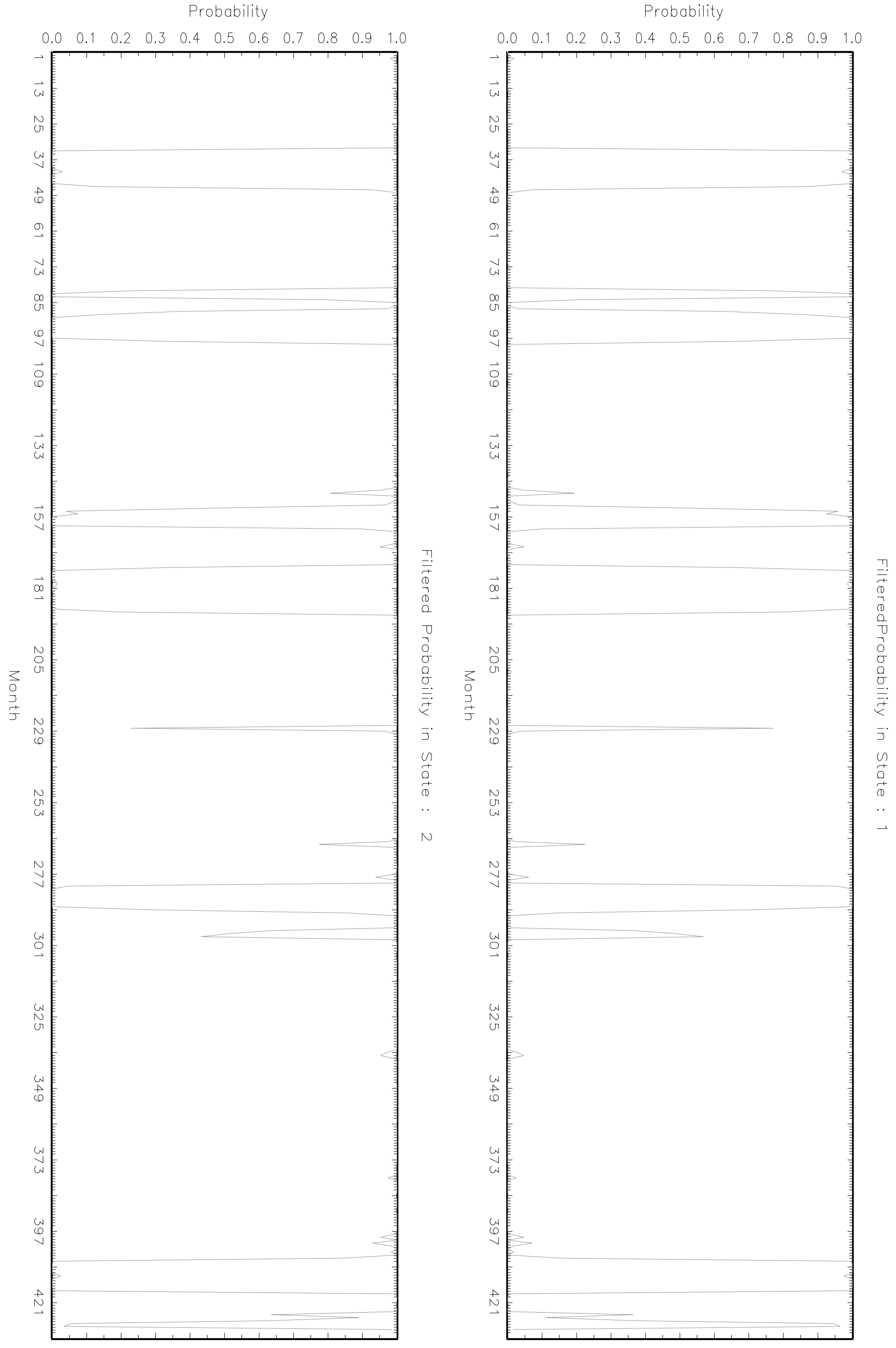


Figure 3
A two state Recession model- Smoothed probabilities of low and high regime
In sample estimation 1967-2 / 1999-12 - Out of sample filtering 1967-2 / 2003-3