Multivariate Generalizations of the Markov-switching Model

BY MOHAMAD KHALED

University of Paris I Panthéon-Sorbonne

February 2006

Preliminary and Very Incomplete
Do not quote please

Abstract

We present a multivariate generalization of the simple markov-switching model. We allow for the introduction of several latent processes that have a simple parametric distribution. The matrix-variate bernoulli distribution yields a flexible yet parsimonious pattern of dependence between the different latent processes while preserving the markovian property. We also show how to estimate the model in the bayesian framework and give several examples.

Keywords: Bayesian statistics, markov-switching, matrix-variate bernoulli distribution, multivariate generalizations.

JEL Classification: C11, C32, C51.

1 Introduction

The Markov-switching regression model has proved to be a useful tool in econometrics over the past two decades. Generalizations of the model to the multivariate case have been undertaken in several papers (see for instance [Krolzig, 1997] and [Sims and Zha, 2004b]. [Khaled, 2004] is a recent survey with a lot of references.) However, all the generalizations assume the existence of a single latent process underlying the model.

A powerful and simple justification behind the use of a single Markov chain as the unique latent process underlying the system is that a Markov chain is already a very general model that can encompass several processes by augmenting the dimension of the system. As a simple example, suppose that the system is driven by two dependent Markov chains, then we can always construct a "super"-Markov chain that contains both chains and that is easily capable of describing the system.

However, one criticism arises in the case of the super-chain and that is the explosion of the number of parameters. Moreover, the great number of parameters in the super-chain increases the possibility of zero occurrences for certain parameters inside the matrix of transition probabilities.

One solution to the problem is to try the simplify the possible patterns of dependence between the different chains. It is posible to think of two polar situatios. At one end, there is the case of completely independent[1] chains. At the other end, there is the case where we allow complete flexibility by permitting complex patterns of dependence through the construction of the "super"-chain. The objective of this paper is to construct a model that lies between those two extremes andt that is flexible enough to allow complex pattern of dependence while staying parsimonious. We show in section 3 how this can be accomplished by showing a flexible model where the dimensionality was drastically reduced (we present an example wherein the number of 50 parameters is reduced to just three!).

Briefly speaking, we build on the matrix-variate bernoulli distribution introduced in [Lovison, 2006]. We are going to show that this distribution is capable of modeling parameters of associations between different markov chains in a very compact way.

We shall introduce the matrix-variate bernoulli distribution in section 2. Section 3 presents the general model and describes how to estimate it in the Bayesian framework. Several examples are presented in section 4. Section 5 concludes.

2 The matrix-variate bernoulli distribution

This section quickly reviews the paper of [Lovison, 2006] in which the matrix-variate bernoulli distribution was introduced. We then present those aspects of the distribution that we will require for our modeling of the markov-switching multivariate regression. Throughout the section, we will keep [Lovison, 2006]'s initial notation.

Let $\boldsymbol{z}_j$, $j = 1, \ldots, m$ be a vector of binary variables containing T observations. In our case, each $\boldsymbol{z}_j$ represents a latent Markov chain. The ith entry in $\boldsymbol{z}_j$ represents the ith time period. The matrix-variate bernoulli distribution describes the joint modeling of the matrix $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m)$ constructed by the concatenation of all the $\boldsymbol{z}_j$ vectors. Therefore $\boldsymbol{Z}$ has $T \times m$ entries. The matrix-variate bernoulli distribution allows for different patterns of dependence. Those include dependence between different variables (simple column-wise dependence between, say, $\boldsymbol{z}_j$ and $\boldsymbol{z}_{j'}$ for $j' \neq j$), dependence between different periods in time (or, put differently, observational unit dependence between, say, $z_{i,j}$ and $z_{i',j}$ for $i' \neq i$) and finally mixed variable-unit dependence (say between $z_{i,j}$ and $z_{i',j'}$). However, the distribution allows for pairwise interactions only. That might prove a formidable restriction in certain applications, but in our case, the distribution offers exactly what we need. The parameters describing association can therefore be classified in three types. We are going to describe three different notations accordingly

- The parameters describing pure variable association

$$\theta_i^j = \theta^j, \forall i$$

$$\theta_i^{j,j'} = \theta^{j,j'}, \forall i$$

1. [Khaled, 2005] introduces a model where all the chains are statistically independent. [Khaled, 2005] nevertheless attempts at tackling the dependence in a multivariate system by adding correlations between the residuals of several sub-blocks of equations.

- The parameters describing pure unit association

$$\lambda_{i,i'}^{j}$$

- The parameters describing mixed unit-variable association

$$\phi_{i,i'}^{j,j'}$$

The way each of those parameters is described is through the use of odds ratios. For instance $\phi_{i,i'}^{j,j'}$ can be defined as

$$\phi_{i,i'}^{j,j'} = \log\left\{\frac{P\{z_i^j = 1, z_{i'}^{j'} = 1\}.P\left\{z_i^j = 0, z_{i'}^{j'} = 0\right\}}{P\left\{z_i^j = 1, z_{i'}^{j'} = 0\right\}.P\left\{z_i^j = 0, z_{i'}^{j'} = 1\right\}}\right\}$$

The probabilities in the log-odds ratios are conditional on the rest being zero.

Now we are going to put the parameters. Let us introduce the following symmetric matrices

$$\boldsymbol{\Theta} = \begin{pmatrix} \theta^1 & \theta^{1,2} & \cdots & \theta^{1,m} \\ & \theta^2 & & \\ & & \ddots & \\ & & & \theta^m \end{pmatrix}$$

$$\boldsymbol{\Lambda}_{i,i'} = \boldsymbol{\Lambda}'_{i',i} = \begin{pmatrix} \lambda_{i,i'}^1 & \phi_{i,i'}^{1,2} & \cdots & \phi_{i,i'}^{1,m} \\ & \lambda_{i,i'}^2 & & \vdots \\ & & \ddots & \vdots \\ & & & \lambda_{i,i'}^m \end{pmatrix}$$

We can put all those into one big matrix

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Theta} & \boldsymbol{\Lambda}_{1,2} & \cdots & \boldsymbol{\Lambda}_{1,T-1} & \boldsymbol{\Lambda}_{1,T} \\ \boldsymbol{\Lambda}_{2,1} & \boldsymbol{\Theta} & & \vdots & \boldsymbol{\Lambda}_{2,T} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ \boldsymbol{\Lambda}_{T-1,1} & \vdots & & \boldsymbol{\Theta} & \boldsymbol{\Lambda}_{T-1,T} \\ \boldsymbol{\Lambda}_{T,1} & \boldsymbol{\Lambda}_{T,2} & \cdots & \boldsymbol{\Lambda}_{T,T-1} & \boldsymbol{\Theta} \end{pmatrix} = \boldsymbol{I}_T \otimes \boldsymbol{\Theta} + \sum_{i=1}^{T} \sum_{i' \neq i} \boldsymbol{E}_{i,i'} \otimes \boldsymbol{\Lambda}_{i,i'}$$

$\boldsymbol{E}_{i,i'}$ is such that all of its entries are zero except for the (i, i')th entry which is equal to one.

The joint likelihood of $\boldsymbol{Z}$ is equal to

$$C(\boldsymbol{\Psi}).\exp\left\{\mathrm{vec}(\boldsymbol{Z}')'.\boldsymbol{\Psi}.\mathrm{vec}(\boldsymbol{Z}')\right\}$$

where $C(\mathbf{\Psi})$ is the integration constant that depends on $\mathbf{\Psi}$.

$$C(\mathbf{\Psi}) = \left(\sum_{k=1}^{2^{Tm}} \exp\{\mathrm{vec}(\boldsymbol{Z}_k')'.\mathbf{\Psi}.\mathrm{vec}(\boldsymbol{Z}_k')\} \right)^{-1}$$

We can factor the likelihood so as write it in a form proportional to

$$\exp\{\mathrm{tr}[\boldsymbol{Z}'.\boldsymbol{Z}.\mathbf{\Theta}]\} \prod_{i=1}^{T} \prod_{i' \neq i} \exp\{\mathrm{tr}[\boldsymbol{Z}'.\boldsymbol{E}_{i,i'}.\boldsymbol{Z}.\mathbf{\Lambda}_{i,i'}]\}$$

After writing the likelihood in that form, we can immediately see that the quantities

$$\boldsymbol{Z}'.\boldsymbol{Z}$$

and each one of

$$\boldsymbol{Z}'.\boldsymbol{E}_{i,i'}.\boldsymbol{Z}$$

are sufficient statistics for $\mathbf{\Theta}$ and each one of the $\mathbf{\Lambda}_{i,i'}$ respectively.

2.1 Markov Chains

Now we are going to illustrate the special case of Markov chains. The problem simplifies greatly since we need temporal dependence of the first order only. The contemporaneous dependence between chain j and chain j' will be captured by $\theta^{j,j'}$.

The markovian dependence between units in each chain is captured by

$$\lambda_{i,i'}^{j} = \begin{cases} \lambda_1^j & \text{if } i' = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Mixed dependence of the first order can also be allowed (i.e. dependence in one chain on past values of other chains)

$$\phi_{i,i'}^{j,j'} = \begin{cases} \phi_1^{j,j'} & \text{if } i' = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$\mathbf{\Lambda}_{i,i+1} = \mathbf{\Lambda}_1 = \begin{pmatrix} \lambda_1^1 & \phi_1^{1,2} & \cdots & \phi_1^{1,m} \\ & \lambda_1^2 & & \vdots \\ & & \ddots & \vdots \\ & & & \lambda_1^m \end{pmatrix}$$

The easiest case is, of course, when there is no temporal dependence across different chains (i.e. the current value of a given chain does not depend on one-period lagged values of other chains). In that case $\mathbf{\Lambda}_1$ is diagonal

$$\mathbf{\Lambda}_1 = \begin{pmatrix} \lambda_1^1 & 0 & \cdots & 0 \\ & \lambda_1^2 & & \vdots \\ & & \ddots & \vdots \\ & & & \lambda_1^m \end{pmatrix}$$

The case of diagonal $\mathbf{\Lambda}_1$ will be enough in most circumstances.

The matrix containing all parameters can, as a result, been written as

$$\mathbf{\Psi} = \boldsymbol{I}_T \otimes \mathbf{\Theta} + \boldsymbol{L}_1 \otimes \mathbf{\Lambda}_1$$

$$\boldsymbol{L}_1 = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

and the likelihood can be written as

$$C(\mathbf{\Psi}).\exp\{\mathrm{tr}[\boldsymbol{Z}'.\boldsymbol{Z}.\mathbf{\Theta}] + \mathrm{tr}[\boldsymbol{Z}'.\boldsymbol{L}_1.\boldsymbol{Z}.\mathbf{\Lambda}_1]\}$$

where $\boldsymbol{Z}'.\boldsymbol{Z}$ and $\boldsymbol{Z}'.\boldsymbol{L}_1.\boldsymbol{Z}$ are the sufficient statistics for that case.

3 The general model

The usual markov switching multivariate regression model is

$$\begin{cases} \boldsymbol{y}_t = \boldsymbol{x}_t.\boldsymbol{\beta}_{s_t} + \boldsymbol{u}_t \\ \\ \boldsymbol{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{s_t}) \end{cases}$$

where $\boldsymbol{y}_t$ is a $1 \times k$ vector of independent variables, $\boldsymbol{z}_t$ is a $1 \times q$ vector of covariates and s_t is a discrete variable that follows a finite Markov chain. As described earlier, we will restrict ourselves to the case where s_t has only two states (i.e., we can as well model it as a binary variable $s_t \in \{0, 1\}$). The equation represents a multivariate regression where the parameters $\mathbf{\Gamma}_{s_t}$ and $\mathbf{\Sigma}_{s_t}$ are state-varying (having a discrete number of possible values). That is, conditionally on each value of the unobserved s_t, we simply have a multivariate regression.

One drawback that one can think of is the fact that s_t describes the behavior of all the parameters at the same time. This is highly restrictive because there is no reason why $q.k + \frac{k.(k+1)}{2}$ different parameters should stochastically change in the same way. Nevertheless, the model is less restrictive than it might seems at the beginning because one can always expand the different chains describing the behavior of different parameters into one big chain. However, the number of additional parameters to be estimated can easily explode. Imagine one introduces three different chains with two states each. A super-chain of eight states will be necessary and that chain will contain 56 parameters to be estimated. The original three different chains contained only six parameters to be estimated. Which in turn means that the increase in the number of parameters necessary to model the association between the chains is 50. If we decide to model the dependence structure between the three chains through the use of the matrix-variate bernoulli distribution, then we will only need six additional parameters. And if we decide to ignore the depdence of a given chain on lagged values of other chains, then the number of additional parameters is going to be three. Three parameters is much less than 50 and this illustrates the highly parsimonious nature of modeling with the matrix-variate bernoulli distribution instead of the super-chain. Moreover, the estimation with the super-chain might not be possible with all datasets, especially in the case where the time series is not very long.

[Khaled, 2005] considers in detail several of those aspects. It also considers some encompassing issues that originate from the introduction of several processes. However, in that paper, the author assumes that the chains are completely independent and attempts to recover some correlations between different sub-blocks of the dependent variables by assuming correlations between the residuals arising from those different sub-blocks. In a sense, our paper is an extension to [Khaled, 2005].

Writing the model in the multiple dependent latent chains case is straightforward. We only need to change the subscript of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ from s_t (the scalar chain at date t) to $\boldsymbol{z}_t$ (the row vector at date t in the matrix $\boldsymbol{Z}$). Estimation procedures will vary from model to model because this general framework encompasses an enormous number of different possible parametrizations. Due to the complex nature of the model, it would be convenient to think of the estimation problem in a modular way, in the sense of decomposing the problem into a multitude of simple distinct estimation procedures. We will show that this is straightforward to undertake in a bayesian framework.

We shall describe the estimation problem in the next section.

3.1 Bayesian Estimation

The bayesian estimation of Markov switching models was introduced by [Albert and Chib, 1993]. Some of the later extensions were given by, among others, [Chib, 1996], [Frühwirth-Schnatter, 2001] and [Sims and Zha, 2004a]. [Khaled, 2004] gives a a survey on Bayesian approaches to Markov-switching multivariate regressions.

Let $\boldsymbol{\phi}$ represent the parameters in the model, i.e.,

$$\boldsymbol{\phi} = \{\boldsymbol{\beta}_{\boldsymbol{z}_t}, \boldsymbol{\Sigma}_{\boldsymbol{z}_t}, \boldsymbol{\Psi}, \boldsymbol{Z}\}$$

We need to put a prior on $\boldsymbol{\phi}$. The prior distributions on $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ will be of the matrix-variate normal - inverse wishart distributions. One must note that different distributions need to be given on different sub-blocks corresponding to different $z_{t,j}$s.

We can specify any type of distribution for $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}_1$ (for instance, matrix-variate normal), beacause, as we shall see, we are going to need a Metropolis-Hastings step in the algorithm.

The objectivie of the estimation procedure is to draw a sample of $\boldsymbol{\phi}$ from the posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{y}) \propto p(\boldsymbol{\phi}).p(\boldsymbol{y}|\boldsymbol{\phi})$ where $p(\boldsymbol{\phi})$ is the prior and $p(\boldsymbol{y}|\boldsymbol{\phi})$ is the likelihood. Monte Carlo Markov chains can be used to get that sample. In particular, a hybrid gibbs sampler with a metropolis-hastings step can be used.

This is how a standard estimation procedure can be made

1. We begin from an initial value for $\boldsymbol{\phi}$, say $\boldsymbol{\phi}^0$.
2. From the iteration m to iteration $m+1$ (i.e. we already in possession of a draw $\boldsymbol{\phi}^m$)
   a. Conditional on the latent chains $\boldsymbol{Z}^m$, we can classify the dataset in several clusters and then undertake usual ways of drawing different sub-blocks of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ for those clusters.
   b. Conditional on $\boldsymbol{Z}^m$, we can undertake a metropolis-hastings step for the estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}_1$.
      We use the product of matrix-variate bernoulli likelihood with the matrix-variate normal prior on $\boldsymbol{\Psi}$.
   c. Conditional on $\boldsymbol{\Psi}^m$, $\boldsymbol{\beta}^m_{\boldsymbol{z}_t}$ and $\boldsymbol{\Sigma}^m_{\boldsymbol{z}_t}$, we can compute the proability of occuence of each state conditional on the observables and the values of the parameters $p(\boldsymbol{z}_t|\boldsymbol{y}_t, \boldsymbol{\Psi}^m, \boldsymbol{\beta}^m_{\boldsymbol{z}_t}, \boldsymbol{\Sigma}^m_{\boldsymbol{z}_t})$ through a recursive filtering algorithm. (Extensive details on how to run the recursive filtering algorithm can be found in [Kim and Nelson, 1999] or [Krolzig, 1997].)
      From the density $p(\boldsymbol{z}_t|\boldsymbol{y}_t, \boldsymbol{\Psi}^m, \boldsymbol{\beta}^m_{\boldsymbol{z}_t}, \boldsymbol{\Sigma}^m_{\boldsymbol{z}_t})$, we can draw $\boldsymbol{Z}^{m+1}$.

Our presentation of the algorithm is too abstract. The examples and simulations of the next section are going to illustrate in detail how to apply the algorithm.

4 Some examples

This section is very incomplete. It is going to be extensively revised in future versions of the paper. In particular, we shall add detailed comments on the estimation methodology and some simulations.

4.1 A simple example

We begin by a very simple example just to illustrate the model. Consider

$$y_t = \boldsymbol{x}_t.\boldsymbol{\beta}_{z_{1,t}} + \sigma_{z_{2,t}}.\varepsilon_t$$

$$\varepsilon_t \sim \mathcal{N}(0,1)$$

$\boldsymbol{\beta}$ and σ depend on two different latent chains $z_{1,t}$ and $z_{2,t}$ that can be dependent.

4.2 A multivariate regression model with a different chain underlying each equation

$$\begin{pmatrix} y_{1,t} & y_{2,t} & y_{3,t} \end{pmatrix} = \boldsymbol{x}_{t\cdot} \begin{pmatrix} \boldsymbol{\beta}_{1,z_{1,t}} & \boldsymbol{\beta}_{2,z_{2,t}} & \boldsymbol{\beta}_{3,z_{3,t}} \end{pmatrix} + \boldsymbol{u}_t$$

$$\boldsymbol{u}_t \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \sigma^2_{1,z_{1,t}} & & \\ \sigma_{21} & \sigma^2_{2,z_{2,t}} & \\ \sigma_{31} & \sigma_{32} & \sigma^2_{3,z_{3,t}} \end{pmatrix}\right)$$

5 Conclusion

One critique that can be leveled against the paper is the restriction to the two-state model. What is needed is a matrix-variate multinomial distribution for a more general use of the methodology in this paper. This a topic for future research. Since a natural representation of a vector of multinomail variables can be a matrix of indicator variables coding the states, the matrix-variate bernoulli distribution can in fact be used to model multinomial variables. The ideas are going to be further explored by the author in future work.

Moreover, due to the big number of possible alternative modelings, automatic selection procedures are of great interest. This might prove a quite intriguing and definitely challenging problem for future research.

Bibliography

[Albert and Chib, 1993] Albert, J. and 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*, 11:1–15.

[Chib, 1996] Chib, S. (1996). Calculating posterior distributions and model estimates in markov mixture models. *Journal of Econometrics*, 75:79–97.

[Frühwirth-Schnatter, 2001] Frühwirth-Schnatter, S. (2001). Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96:194–209.

[Khaled, 2004] Khaled, M. (2004). Bayesian estimation of markov switching multivariate regressions. Master's thesis, Université Paris I Panthéon-Sorbonne.

[Khaled, 2005] Khaled, M. (2005). A bayesian solution to the concomitance problem in markov switching multivariate regressions. Working Paper Université Paris I Panthéon-Sorbonne.

[Kim and Nelson, 1999] Kim, C.-J. and Nelson, C. (1999). *State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press.

[Krolzig, 1997] Krolzig, H. (1997). *Markov-Switching Vector Autoregressions*. Springer.

[Lovison, 2006] Lovison, G. (2006). A matrix-valued bernoulli distribution. *Journal of Multivariate Analysis*.

[Sims and Zha, 2004a] Sims, C. and Zha, T. (2004a). MCMC method for markov mixture simultaneous equation models: A note. Manuscript. Princeton University and Federal Reserve Bank of Atlanta.

[Sims and Zha, 2004b] Sims, C. and Zha, T. (2004b). Were there regime switches in US monetary policy? Manuscript. Princeton University and Federal Reserve Bank of Atlanta.