Indirect estimation of Markov switching models with endogenous switching

Otranto, Edoardo and Calzolari, Giorgio and Di Iorio, Francesca

Universita' di Sassari, Italy, Universita' di Firenze, Italy, Universita' di Napoli Federico II, Italy

2005

INDIRECT ESTIMATION OF MARKOV SWITCHING MODELS WITH ENDOGENOUS SWITCHING

Edoardo Otranto[1], Giorgio Calzolari[2] and Francesca Di Iorio[3]

[1] Dipartimento di Economia, Impresa e Regolamentazione
Università di Sassari
Via Torre Tonda, 34-07100 Sassari
(e-mail: eotranto@uniss.it)
[2] Dipartimento di Statistica "G. Parenti"
Università di Firenze
Viale Morgagni, 59-50134 Firenze
(e-mail: calzolar@ds.unifi.it)
[3] Dipartimento di Scienze Statistiche
Università di Napoli Federico II
Via Rodinò, 22-80138 Napoli
(e-mail: fdiiorio@unina.it)

ABSTRACT. Markov Switching models have been successfully applied to many economic problems. The most popular version of these models implies that the change in the state is driven by a Markov Chain and that the state is an exogenous discrete unobserved variable. This hypothesis seems to be too restrictive. Earlier literature has often been concerned with endogenous switching, hypothesizing a correlation structure between the observed variable and the unobserved state variable. However, in this case the classical likelihood-based methods provide biased estimators. In this paper we propose a simple "estimation by simulation" procedure, based on indirect inference. Its great advantage is in the treatment of the endogenous switching, which is about the same as for the exogenous switching case, without involving any additional difficulty. A set of Monte Carlo experiments is presented to show the interesting performances of the procedure.

1 INTRODUCTION

Markov Switching (MS hereafter) models have been successful in the econometric literature, since they are able to consider different states in the analyzed variable, and to provide an inference on the regime. The most popular version of these models (Hamilton, 1990) supposes that the change in the state is driven by a Markov chain with constant transition probability matrix and with an exogenous latent state variable controlling the regime change. In some applications the hypothesis of exogenous switching seems too strong; for example, in the analysis of business cycle, the state variable would represent recession and growth periods, and would be naturally correlated with the observable dependent variable.

The motivation for which the state variable has been considered exogenous with respect to the dependent variable is that the estimation method used for MS models, based on the nonlinear filter of Hamilton (1990), provides biased estimation in the case of endogenous switching. The researchers have dealt with this problem only recently. In particular, Kim *et al.* (2003) propose a bias correction in the likelihood; their approach is very interesting, because it is able to explicit the exact likelihood function, hypothesizing a particular Normal model for the state variable; however when the model underlying the state variable is not Normal, their approach provides only an approximation of the likelihood.

In this work, we propose to estimate the MS model with endogenous switching via indirect estimation procedures (Gouriéroux *et al.* 1993, Gallant and Tauchen, 1996), that seem robust under different model specifications for the state variable. Using "computational" care in dealing with the discontinuous response implied by the switching regime, the implementation of indirect inference is quite straightforward. In particular, we show, via Monte Carlo experiments, that the auxiliary model we propose provides good performance of indirect estimation both for exogenous and endogenous switching and both for the Normal and non Normal cases.

In the next section we briefly describe the MS model we deal with; in section 3 the indirect estimation method is introduced and in section 4 we show some simulation experiments to evaluate this approach, comparing it with the procedures proposed by Kim *et al.* (2003) and Hamilton (1990).

2 THE MODEL

For the sake of simplicity, let us consider the simple MS model:

$$\begin{aligned} & y_t = \alpha_0 + \alpha_1 s_t + \varepsilon_t \qquad\qquad t = 1, ..., T \\ & \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_{s_t}\right) \end{aligned} \tag{1}$$

where the intercept and the variance of the error terms depend upon the state variable s_t, which can assume two values (name them 0 and 1), representing the regimes. The state variable is not observable and its distribution is unknown; we suppose that s_t follows an ergodic Markov chain, in which the probability of a particular realization of s_t depends only on the realization of s_{t-1}. The transition probabilities are in the matrix $\mathbf{P}$, where the generic element $p_{ij} = Pr(s_t = j|s_{t-1} = i),\ i, j = 0, 1$ represents the probability of being in state j at time t given that the state at time $t-1$ was i. Of course, model (1) can be generalized, including autoregressive terms or exogenous variables, depending or not on the variable s_t, and considering a generic number k of regimes or more complex probability structures (for instance as in Otranto, 2005).

Hamilton (1990) estimates this model developing a non linear Kalman filter to integrate the effect of the lagged states, simplifying the likelihood function. Such a model considers the state variable s_t exogenous with respect to y_t.

To introduce the MS model with endogenous switching we need to specify a certain degree of correlation between s_t and y_t; as in Kim *et al.* (2003), we suppose a probit representation of the state process:

$$s_t = \begin{cases} 0 \ if \ s^*_t < 0 \\ 1 \ if \ s^*_t > 0 \end{cases}$$

where:

$$\begin{aligned} & s^*_t = \vartheta_0 + \vartheta_1 s_{t-1} + \eta_t \qquad\qquad t = 1, ..., T \\ & \eta_t \sim \mathcal{N}(0, 1) \end{aligned} \tag{2}$$

From this representation, it follows that the transition probabilities are obtained as:

$$\begin{aligned} p_{00} &= \Phi(-\vartheta_0) \\ p_{11} &= 1-\Phi(-\vartheta_0-\vartheta_1) \end{aligned}$$

where Φ is the standard Normal cumulative distribution function. In addition, we suppose that ε_t in (1) and η_t in (1) are correlated with correlation parameter ρ. Of course, to obtain unbiased (consistent) estimators of the unknown parameters in (1), we have to take into account, in the likelihood specification, the expected value and the variance of ε_t conditional on s_t and s_{t-1}. Kim *et al.* (2003) propose a bias correction, which provides the exact likelihood in the case of Normal η_t, whereas it is only an approximation in the other cases. Its use in other cases may lead to biased (inconsistent) estimators. In the next section we propose an indirect inference methodology to obtain consistent estimators in more generale cases.

3 INDIRECT ESTIMATION

Among the many interesting simulation based estimation procedures proposed in the literature, Indirect Inference (Gouriéroux et al., 1993) and the Efficient Method of Moments (Gallant and Tauchen, 1996) deserve a special attention from the computational point of view for their simplicity of implementation.

The underlying idea of these methods is that the estimation of a model of interest $y_t = f(y_{t-1}, x_t, \varepsilon_t, \alpha)$, where x_t are exogenous variables, $\alpha \in A$ is the vector of parameters of interest, ε_t are the errors with *known parametric distribution*, y_t are endogenous variables and y_{t-1} lagged endogenous variables, can be conducted using a misspecified model (called *auxiliary*) $y_t = g(y_{t-1}, x_t, \nu_t, \beta)$, where $\beta \in B$ is a parameter vector, ν_t are errors.

The Indirect Inference of α proposed by Gouriéroux et al. (1993) is given by:

$$\hat{\alpha} = argmin[\hat{\beta} - \tilde{\beta}(\alpha)]' \Omega_1^{-1} [\hat{\beta} - \tilde{\beta}(\alpha)] \tag{3}$$

where $\hat{\beta}$ are the parameter estimates of the auxiliary model using the observed data y_t generated by the model of interest, $\tilde{\beta}(\alpha)$ are the parameter estimates of the auxiliary model using data $\tilde{y}(\alpha)$, simulated from the model of interest conditional on x_t, ε_t and α, and Ω_1 is a positive definite matrix.

Gallant and Tauchen (1996) proposed a different version of (3), also called Efficient Method of Moments (*EMM* hereafter), given by:

$$\hat{\alpha} = argmin \frac{\partial L}{\partial \beta'}(\tilde{y}(\alpha); \hat{\beta}) \Omega_2^{-1} \frac{\partial L}{\partial \beta}(\tilde{y}(\alpha); \hat{\beta}) \tag{4}$$

where $\frac{\partial L}{\partial \beta}(\tilde{y}(\alpha); \hat{\beta})$ is the score function of the auxiliary model on the simulated data $\tilde{y}(\alpha)$ evaluated in $\hat{\beta}$, and Ω_2 is a positive definite matrix.

Estimators (3) and (4) associated with the optimal choice of the weighting matrix Ω_1 and Ω_2 have the same asymptotic efficiency and are both consistent and asymptotically Normal under general conditions (Gouriéroux and Monfort, 1996).

The simplicity of the indirect estimation "costs" in finite sample a larger variance of the estimated parameters. The same variance can be reduced considering the average of h (=100 in our experiments) estimates of the auxiliary model obtained on h independent simulated data sets from the model of interest: $\tilde{\beta}(\alpha) = \frac{1}{h}\sum_{j=1}^{h}\tilde{\beta}_j(\alpha)$, since the expression of the variance includes a multiplying factor $(1+\frac{1}{h})$. Gouriéroux *et al.* (1993) showed that, without exogenous variables, this variance reduction can be obtained with a single $h*T$ size simulated series from the model of interest. The same result can be found for (4). Moreover, Gallant and Tauchen (1999) proved that if the auxiliary model encompasses the true data generating process, than the Quasi-Likelihood becomes a sufficient statistic and the EMM becomes fully efficient, while if the auxiliary model is a close approximation to the data generating process, the EMM efficiency can be expected to be close to the Maximum Likelihood estimator.

One of the remarkable features of indirect estimation methods is that they can work (and sometimes they can work "well") even with surprisingly simple auxiliary models, the only requirement being the existence of a regular and well behaved binding function $b(\alpha)$ for $\alpha \in A$ (see Gouriéroux et al., 1993): in practice, some well behaved relationship must exist between parameters of the model of interest (α) and parameters (or score) of the auxiliary model $\tilde{\beta}(\alpha)$.

For our problem we suggest to use as auxiliary model a Normal bivariate model very similar to (1)-(1), avoiding the unobservable variables. The model is:

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1\zeta_t + \varepsilon_t \\ \zeta_t &= \vartheta_0 + \vartheta_1\zeta_{t-1} + \eta_t \\ \mathbf{e}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{H}_{s_t}) \end{aligned} \tag{5}$$

where $\mathbf{e}_t$=$[\varepsilon_t, \eta_t]'$ and

$$\mathbf{H}_{s_t} = \begin{bmatrix} \sigma_0^2(1-\zeta_t)+\sigma_1^2\zeta_t & \rho\sigma_0(1-\zeta_t)+\sigma_1\zeta_t \\ \rho\sigma_0(1-\zeta_t)+\sigma_1\zeta_t & 1 \end{bmatrix}$$

The variable ζ_t is obtained by the logistic function:

$$\zeta_t = \frac{\exp[cy_t]}{1+\exp[cy_t]}$$

and substitutes the latent variables s_t^* and s_t in (1)-(1). The constant c is chosen experimentally (values around 5–10 have given good results in the experiments). Supposing $s_t = 1$ the "high" regime and $s_t = 0$ the "low" regime, the logistic transformation pushes large values towards 1 (likely related to $s_t = 1$), and pushes towards 0 all small values of y_t (likely related to $s_t = 0$). It has, however, the advantage of being a continuous transformation (discontinuities cause great, sometimes overwhelming, computational difficulties in indirect estimation, often making minimization difficult or impossible; see Di Iorio and Calzolari, 2005).

What must be done to obtain an estimate of the auxiliary model parameters ($\hat{\beta}$), is simply to maximize the pseudo-likelihood of model (**??**):

$$\prod_{t=1}^{T} |2\pi\mathbf{H}_{s_t}|^{-0.5} \exp\left\{-0.5\begin{bmatrix} y_t-\alpha_0-\alpha_1\zeta_t \\ \zeta_t-\vartheta_0-\vartheta_1\zeta_{t-1}\end{bmatrix}' \mathbf{H}_{s_t}^{-1} \begin{bmatrix} y_t-\alpha_0-\alpha_1\zeta_t \\ \zeta_t-\vartheta_0-\vartheta_1\zeta_{t-1}\end{bmatrix}\right\}$$

Its score, which has a simple closed form expression, is then used in the EMM procedure, as in equation (4).

Table 1. *Simulation results: means and variances (in parentheses) of the estimates. Number of replications=1000; T=1000*

Normal case							
Par.	α_0	α_1	ϑ_0	ϑ_1	σ_0^2	σ_1^2	ρ
True	-2	4	-1.3	2.6	1	1	0
EMM	-1.9951	3.9868	-1.2951	2.6039	0.9981	1.0109	-0.0018
	(0.0049)	(0.0127)	(0.0086)	(0.0216)	(0.0096)	(0.0097)	(0.0097)
KPS	-1.9966	4.0036	-1.3028	2.5811	1.0067	0.9848	-0.0040
	(0.0027)	(0.0056)	(0.0068)	(0.0142)	(0.0069)	(0.0056)	(0.0073)
H	-1.9950	3.9999	-1.3037	2.5828	1.0059	0.9844	
	(0.0023)	(0.0043)	(0.0067)	(0.0140)	(0.0068)	(0.0056)	
True	-2	4	-1.3	2.6	1	1	0.9
EMM	-1.9936	3.9872	-1.2932	2.5986	1.0050	1.0097	0.9061
	(0.0028)	(0.0057)	(0.0058)	(0.0150)	(0.0049)	(0.0059)	(0.0014)
KPS	-2.0008	3.9932	-1.2976	2.6007	0.9975	1.0113	0.9021
	(0.0023)	(0.0047)	(0.0064)	(0.0125)	(0.0061)	(0.0040)	(0.0006)
H	-2.3120	4.6173	-1.2980	2.5935	0.9023	0.9165	
	(0.0018)	(0.0049)	(0.0073)	(0.0115)	(0.0045)	(0.0032)	
True	-2	4	-1.3	2.6	0.25	1	0.9
EMM	-1.9974	3.9960	-1.3188	2.6570	0.2487	1.0000	0.9171
	(0.0008)	(0.0111)	(0.0129)	(0.0802)	(0.0004)	(0.0059)	(0.0015)
KPS	-2.0004	3.9928	-1.2976	2.6007	0.2494	1.0112	0.9021
	(0.0006)	(0.0030)	(0.0064)	(0.0125)	(0.0004)	(0.0040)	(0.0006)
H	-2.1561	4.4619	-1.2991	2.5955	0.2252	0.9146	
	(0.0005)	(0.0030)	(0.0072)	(0.0113)	(0.0003)	(0.0031)	
Mixed case							
True	-2	4	-1.3	2.6	1	1	0.9
EMM	-1.9980	3.9997	-1.3092	2.6274	0.9705	0.9554	0.9048
	(0.0024)	(0.0053)	(0.0164)	(0.0652)	(0.0192)	(0.0116)	(0.0090)
KPS	-2.0675	4.1361	-1.5420	3.0768	0.9656	0.9752	0.7820
	(0.0018)	(0.0042)	(0.0069)	(0.0147)	(0.0164)	(0.0148)	(0.0033)
H	-2.2544	4.5102	-1.4772	2.9604	0.9277	0.9313	
	(0.0019)	(0.0039)	(0.0069)	(0.0168)	(0.0138)	(0.0121)	
True	-2	4	-1.3	2.6	0.25	1	0.9
EMM	-1.9967	4.0098	-1.3427	2.7119	0.2483	0.9316	0.9254
	(0.0005)	(0.0049)	(0.0171)	(0.1027)	(0.0021)	(0.0185)	(0.0036)
KPS	-2.0321	4.0915	-1.5384	3.0728	0.2347	0.9994	0.7925
	(0.0004)	(0.0022)	(0.0069)	(0.0151)	(0.0006)	(0.0220)	(0.0021)
H	-2.1264	4.3802	-1.4776	2.9616	0.2254	0.9507	
	(0.0005)	(0.0023)	(0.0069)	(0.0169)	(0.0005)	(0.0191)	

4 MONTE CARLO EXPERIMENT

In this work we follow the Gallant and Tauchen (1996) EMM approach that presents, in this framework, some computational advantages. We propose some simulation experiments to evaluate the performance of the indirect method in the estimation of a model (1)-(1). The length of each simulated time series is $T = 1000$ and the number of Monte Carlo replications is 1000.

We have performed several simulation experiments, considering the case of high correlation and the case of no correlation, as well as the case of switching variance and fixed variance. In Table 1 we compare the results obtained by the EMM procedure with the method of Kim *et al.* (KPS) and the classical estimation via Hamilton (1990) non linear filter (H), which is biased when $\rho \neq 0$. In addition, we consider two cases: the case named "Normal" in which both ε_t and η_t are Normal, and the case named "Mixed", in which η_t is generated as a Student's-*t* random variable with 4 degrees of freedom (successively standardized), and ε_t as a weighted sum of η_t and a standard Normal variable, with weights calibrated to guarantee a correlation equal to ρ and a variance of ε_t equal to $\sigma^2_{s_t}$. The values of ϑ_0 and ϑ_1 correspond to p_{00}=p_{11}=0.9032, representing a situation of strong persistence in the same regime (typical situation in many real cases, e.g. Hamilton, 1990).

From Table 1 we can note the performance of EMM. It shows a loss in efficiency with respect to *KPS* in the Normal case, where *KPS* considers the true likelihood function; the loss is particularly evident in the estimation of ρ. However, the EMM method is the only robust with respect to the Mixed case, in which *KPS* and *H* perform poorly, in particular in the estimation of ρ, ϑ_0 and ϑ_1.

REFERENCES

DI IORIO, F., CALZOLARI, G. (2005 forthcoming): Discontinuities in Indirect Estimation: an Application to EAR Models, *Computational Statistics and Data Analysis*

GALLANT, A.R., TAUCHEN, G. (1996): Which Moments to Match?, *Econometric Theory*, *12*, 657-681.

GALLANT, A.R., TAUCHEN, G. (1999): The Relative Efficiency of the Method of Moments Estimator, *Journal of Econometrics* , 92, 149-172.

GOURIÉROUX, C., MONFORT, A., (1996): *Simulation-Based Econometric Methods*, Oxford University Press.

GOURIÉROUX, C., MONFORT, A., RENAULT, E. (1993): Indirect Inference, *Journal of Applied Econometrics*, *8*, S65-S118.

HAMILTON, J.D. (1990): Analysis of Time Series Subject to Changes in Regime, *Journal of Econometrics*, *45*, 39-70.

KIM, C.-J., PIGER, J., STARTZ, R. (2003): Estimation of Markov Regime-Switching Regression Models with Endogenous Switching, *Federal Reserve Bank of St. Louis, Working Paper*, *2003-015*.

OTRANTO, E. (2005 forthcoming): The Multi-Chain Markov Switching Model, *Journal of Forecasting*.