STATISTICAL ESTIMATION AND TESTING OF TURNING POINTS IN MULTIVARIATE REGIME-SWITCHING MODELS

V. I. Malugin[1], A. Yu. Novopoltsev[2]
Belarusian State University
Minsk, BELARUS
e-mail: [1]`malugin@bsu.by`, [2]`novopsacha@gmail.com`

Abstract

For vector autoregressive models with Markov switching states (MS-VARX) we propose the algorithms of classification of states based on classified and non-classified learning samples. We also suggest the procedure to exclude short-term (acyclic) fluctuations in system states. It is based on successive application of algorithms implementing the Bayesian plug-in decision rule of point-wise classification and a statistical test for expected probability of misclassification. Accuracy of the algorithms is examined by means of computer simulation experiments.

1 Models and tasks of the research

Regime-switching models (RS-Models) are convenient for analyzing complex systems with cyclic changes of state [1]. Most studies are devoted to *Markov-switching* vector autoregressive model (MS-VAR) [2]. In the case of independent states *independent regime-switching* autoregressive and regressive models (IS-Models) should be used. These models are also preferable under the Markov dependence condition when there are high uncertainty about the future state of a system. The models of this type were thoroughly studied in [3, 4]. In this paper, the object of study is the vector autoregressive model with Markov-switching states including exogenous variables (MS-VARX), thus allowing a multivariate linear regressive model (MS-MLR) as its special case.

Let a complex system at time t is characterized by a random observation vector defined on the probability space $(\Omega, \mathfrak{F}, \mathbf{P})$, where Ω — space of elementary objects $\omega \in \Omega$; $\mathbf{P}$ — probability measure: $\mathbf{P}(A) = \mathbf{P}\{\omega \in A\}$, $A \in \mathfrak{F}$. Let $\{\Omega_0, \dots, \Omega_{L-1}\}$ – decomposition of Ω into a finite number of non-empty disjoint subsets, such that $\Omega_l \in \mathfrak{F}$, $\mathbf{P}\{\Omega_l\} = \mathbf{P}(\{\omega \in \Omega_l\}) > 0$, $\bigcup_{l \in S(L)} \Omega_l = \Omega$, $S(L) = \{0, \dots, L-1\}$. These subsets are the classes of states of a complex system, the number of which is L.

A random vector of observation $y_t = ({x'}_t,\ {z'}_t)' \in \Re^n$ can be partitioned into subsectors of endogenous variables $x_t = (x_{tj}) \in \Re^N$ and exogenous variables $z_t = (z_{tk}) \in \mathfrak{X} \subset \Re^M$. It is assumed that, in general, the time series is described by a model RS-VARX(p)($p \geq 1$):

$$x_t = \sum_{i=1}^{p} A_{d(t),i} x_{t-i} + B_{d(t)} z_t + \ {\eta_{d(t)}}_t, \quad t = 1, \dots, T, \tag{1}$$

where $x_{1-p}, \dots, x_0 \in \Re^N$ — the set of the given initial values; $\eta_{d(t),t} \in \Re^N$ — random disturbances or innovation process; $d(t) \in S(L) = \{0, \dots, L-1\}$ — the class of state number.

Model (1) should satisfy with the following conditions:

M.1. Segmented-stationary condition: matrices $A_{l,i}(i = 1, \ldots, p)$ satisfy with the stationarity condition of VAR(p) model for each class of states $l \in S(L)$.

M.2. Disturbance assumptions: $\mathbf{E}\eta_{l,r} = 0_N \in \Re^N, \mathbf{E}\{\eta_{l,r}\eta'_{l,s}\} = \delta_{r,s}\Sigma_l$ $(r, s = 1, ..., T,\ l \in S(L))$, where $\delta_{r,s}$ — the Kronecker delta.

M.3. Exogenous variables $z_t = (z_{t1}, ..., z_{tM})' \in \mathfrak{X} \subseteq \Re^M$ are deterministic or stationary time series.

M.4. Structural heterogeneity conditions: $A_l \neq A_k$ and (or) $B_l \neq B_k$ $\forall k \neq l,\ k, l \in S(L)$.

We consider a model with $L\,(2 \leq L < s + 1)$ classes of states, where $s \geq 1$ — number of state switching points $1 < \tau_1 < \ldots < \tau_s < T$. Concerning the sequence of states $d(t) \equiv d_t \in S(L)$ $(t = 1, ..., T)$ there are two types of assumptions:

d.1. d_t $(t = 1, ..., T)$ — independent identically distributed random variables with probability distribution $\mathbf{P}\left\{d_t = l\right\} = \pi_l > 0\,(l \in S(L))\,,\ \sum_{l \in S(L)} \pi_l = 1$; $\mathbf{P}\left\{d_t = l\right\} = \pi_l > 0\,(l \in S(L))\,,\ \sum_{l \in S(L)} \pi_l = 1$;

d.2. $d_t\,(t = 1, ..., T)$ — homogeneous ergodic Markov chain (GCM) with the distribution, which is determined by the vector of probability of the initial state π and a matrix of one-step transition probabilities P:

$$\pi = (\pi_l),\ \pi_l = \mathbf{P}\left\{d_1 = l\right\} > 0\ \ (l \in S(L))\,,\ \sum\nolimits_{l \in S(L)} \pi_l = 1;$$

$$P = (p_{kl})\,, p_{kl} = P\left\{d_{t+1} = l \middle|\, d_t = k\right\} \geq 0,\ \sum\nolimits_{l \in S(L)} p_{kl} = 1,\ k \in S(L).$$

Under the conditions of d.1 and d.2, we get the models IS-VARX and MS-VARX, respectively. Model (1) allows for a number of special cases: a model of multivariate linear regression RS-MLR, if $p = 0,\ M \geq 1$ [4]; a model RS-VAR without exogenous variables, if $p > 0,\ M = 0$ [2].

The true values of a model parameters $\{A_l, B_l, \Sigma_l\}(l \in S(L))$, π, P and the moments of switching state $\{\tau_i\}(i = 1, ..., s)$ are unknown. There are a classified or a non-classified sample of observations $(\bar{X},\ \bar{Z})$ $(\bar{X} = (x_t) \in \Re^{NT},\ \bar{Z} = (z_t) \in \mathfrak{X}^T \subseteq \Re^{MT})$ when a vector of states $\bar{d} = (d_t) \in S^T(L)$ is known and unknown, respectively. We presented two statistical classification algorithms for MS-VARX model in these cases: an EM-algorithm for joint estimation of the parameters and the vector of states for a non-classified sample and a discriminant analysis algorithm in the case of a classified sample for classification of out-of-sample observations. To eliminate short-term fluctuations of states arising from misclassification, we propose a statistical test based on a pointwise classification decision rule. For IS-MLR and IS-VARX models the problems mentioned are solved in [3, 4].

2 Representations for the model parameter estimates

Model (1) under the assumptions M.1–M.4, d.1 or d.2 can be represented in the regression form

$$x_t = \Pi_{d(t)} u_t + \eta_{d(t),t}, \tag{2}$$

where $\Pi_{d(t)} = (A_{d(t),1}, \dots, A_{d(t),p}, B_{d(t)})$ is the block $N \times (pN+M)$-matrix of parameters; $u_t = (x'_{t-1}, \dots, x'_{t-p}, z'_t)' \in \Re^{Np+M}$ — the stacked vector of predetermined variables formed from lagged endogenous and exogenous variables whose values are known at time t.

In this case we use a sample of observations $(\bar{X}, \bar{U})$, where $\bar{X} = (x'_1, \dots, x'_T)' \in \Re^{NT}$ — the values of the endogenous variables, which correspond to the values $\bar{U} = (u'_1, \dots, u'_T)' \in \Re^{NpT} \times \mathfrak{X}^T \subseteq \Re^{(Np+M)^T}$ of predefined exogenous variables. For the model (2) we will also denote:

$\theta_l \in \Re^m$ $(m = N \times (pN+M) + N\,(N+1)/2\ )$ — stacked vector of the parameters for the class $l \in S(L)$ formed of independent elements of matrices $\{\Pi_l,\ \Sigma_l\}$ $(l \in S(L))$;

$\phi \in \Re^q$ $(q = Lm + (L-1)\,(L+1))$ — parameters of a mixture of distributions including $\{\theta_l\}$ and π, P, $\hat{\phi} \in \Re^q$ — statistical estimate of $\phi \in \Re^q$;

$D = (d_1, \dots, d_T)' \in S^T(L)$ — state vector for the period under observation;

$\tilde{\gamma}_{l,t} = \mathbf{P}\{d_t = l | \bar{X},\, \bar{U};\, \tilde{\phi}\}$ — posteriori probability of the class $l \in S(L)$ at the moment t;

$\tilde{\xi}_{kl,t} = \mathbf{P}\{\, d_{t+1} = l |\, d_t = k;\, \bar{X},\, \bar{U};\, \tilde{\phi}\}$ — posteriori probability of a transition from class $k \in S(L)$ to class $l \in S(L)$ at the moment t $(t = 1, \dots, T-1)$.

For a joint estimation of all the parameters and the state vector an EM-algorithm (*Expectation-Maximization algorithm*) is proposed. This algorithm belongs to the family of Baum–Welch algorithms for splitting a mixture of multivariate distributions, controlled by a hidden Markov chain [5]. In accordance with the general approach [4, 5], we obtain an analytical representation for the estimated characteristics.

The representation for an estimate $\hat{\phi} \in \Re^q$ is obtained by maximization of the conditional expectation of the log-likelihood function for some given initial value $\tilde{\phi} \in \Re^q$:

$$\hat{\phi} = \underset{\phi \in \Re^q}{\arg\max}\, \Lambda(\phi, \tilde{\phi}) = \underset{\phi \in \Re^q}{\arg\max}\, E_{\tilde{\phi}}\{l(\phi; \bar{X}, \bar{U}, D) | \bar{X}, \bar{U}; \tilde{\phi}\}, \tag{3}$$

$$l(\phi; \bar{X}, \bar{U}, D) = \ln(\pi_{d_1} p_X(x_1; u_1, \theta_{d_1})) + \sum_{t=2}^{T} \ln(p_{d_{t-1},d_t} p_X(x_t; u_t, \theta_{d_t})). \tag{4}$$

Theorem 1. *If the model (1), (2) satisfy the assumptions of M.1–M.4, d.2, then the estimates* $\{\hat{\Pi}_l, \hat{\Sigma}_l\}$ $(l \in S(L))$, $\hat{\pi}, \hat{P}$ *on a sample* $(\bar{X}, \bar{U})$ *are the solution of the problem (3), (4) for the given vector of parameters* $\tilde{\phi} \in \Re^q$*:*

$$\hat{\pi}_l = \tilde{\gamma}_{l,1}, \hat{p}_{kl} = \sum_{t=2}^{T} \tilde{\xi}_{kl,t} \left(\sum_{t=2}^{T} \tilde{\gamma}_{k,t-1} \right)^{-1}, \hat{\Pi}_l = \sum_{t=1}^{T} \tilde{\gamma}_{l,t} x_t u'_t \left(\sum_{t=1}^{T} \tilde{\gamma}_{l,t} u_t u'_t \right)^{-1}, \tag{5}$$

$$\hat{\Sigma}_l = \sum_{t=1}^{T} \tilde{\gamma}_{l,t}(x_t - \hat{\Pi}_l z_t)(x_t - \hat{\Pi}_l z_t)' \left(\sum_{t=1}^{T} \tilde{\gamma}_{l,t} \right)^{-1}, \tag{6}$$

where analytical representations for the posterior probabilities $\{\tilde{\gamma}_{l,t}\}$, $\{\tilde{\xi}_{kl,t}\}$ are obtained as specified above.

Corollary 1. *Using the known block structure for the matrices $\hat{\Pi}_l$, we can get the estimates $\{\hat{A}_{l,1}, \ldots, \hat{A}_{l,p}, \hat{B}_l\}$ $(l \in S(L))$.*

3 Classification and testing procedure

Bayesian decision rules (BDR) of pointwise and groupwise classification of multivariate observations described by IS-VARX model, have been proposed in [4]. In the case of a Markov-switching model we propose a decision rule of groupwise classification based on the dynamic programming approach described in [6].

Lemma 1. *If the model (1), (2) satisfy the assumptions of M.1–M.3, d.2, and parameters $\phi \in \Re^q$ are known, then a BDR of groupwise classification is determined by the condition*

$$\hat{D} \equiv \hat{D}(\bar{X}_1^T, \bar{U}_1^T) = \arg\max_{D \in S^T(L)} l(\phi; \bar{X}_1^T, \bar{U}_1^T, D), \tag{7}$$

where $(\bar{X}_1^T, \bar{U}_1^T)$ $(\bar{X}_1^T = (x'_1, ..., x'_T)' \in \Re^{NT}, \bar{U}_1^T = (u'_1, ..., u'_T)' \in \Re^{NpT} \times \mathfrak{X}^{MT} \subseteq \Re^{(Np+M)T})$ is a sample of observations to be classified.

To solve the problem (7) with a help of a dynamic programming approach, we use a special representation of the log-likelihood function $l(\phi; \bar{X}, \bar{U}, D)$ through the Bellman function [7].

Theorem 2. *Under the conditions of Lemma 1, a BDR of groupwise classification of sample $(\bar{X}_1^T, \bar{U}_1^T)$ is implemented using dynamic programming method in accordance with the following relationships:*

$$\hat{d}_T = \arg\max_{k \in S(L)} F_T(k),\ \hat{d}_t = \arg\max_{k \in S(L)} \Big(f_t(k, \hat{d}_{t+1}) + F_t(k) \Big),\ t = T-1, T-2, ..., 1, \tag{8}$$

$$F_1(l) \equiv 0,\ F_{t+1}(l) = \max_{k \in S(L)} \left(f_t(k, l) + F_t(k) \right),\ l \in S(L), \quad t = 1, ..., T-1, \tag{9}$$

where $\{F_t(k)\}$ are Bellmans functions and $\{f_t(k, l)\}$ are determined by formula

$$f_t(k,\, l) = \delta_{t,1} \left(\ln \pi_k + \ln p_X \left(x_1;\, u_1,\, \theta_k \right) \right) + \ln p_{kl} + \ln \left(x_{t+1};\, u_{t+1},\, \theta_l \right), \tag{10}$$

δ_{t1} — Kronecker symbol, $t = 1, \ldots, T-1$.

If $\{\hat{\theta}_l\}$ $(l \in S(L)), \hat{\pi}, \hat{P}$ are estimates of a model parameters, then using them in (9) we obtain a consistent "plug-in" decision rule. To find these estimates, it is advisable to apply the EM-algorithm proposed here. The "plug-in" BDR of group classification can be used to classify out-of-sample observations (x_τ, u_τ) $(\tau = T+1, \dots, T+h)$, that is, to forecast future states of a system.

We also suggest a procedure that allows to eliminate short-term (acyclic) fluctuations in system states, which caused by errors of classification of the proposed decision rules. It is based on application of algorithms implementing the Bayesian plug-in decision rule of pointwise classification and subsequent use of a statistical test for expected probability of misclassification [8].

References

[1] Hamilton J.D. (2008). Regime-switching models. *New Palgrave Dictionary of Economics. 2nd Edition. Palgrave Macmillan, Basingstoke*. pp. 1755-1804.

[2] Krolzig H.M. (1997). *Markov-switching vector autoregressions. Modelling statistical inference and application to business cycle analysis*. Springer, Berlin.

[3] Malugin V.I., Kharin Yu.S. (1986). On optimal classification or random observations different in regression equations. *Automation and Remote Control*. Vol. **7**, pp. 61-69 (in Russian).

[4] Malugin V.I. (2014). *Methods of analysis of multivariate econometric models with heterogeneous structure*. BSU, Minsk (in Russian).

[5] Bilmes J.A. (1998). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models: Technical Report. *Int. Computer Science Institute*. Berkeley, USA.

[6] Kharin Yu.S. (1996). *Robustness in statistical pattern recognition*. Kluwer Academic Publishers, Dordrecht.

[7] Malugin V.I., Novopoltsev A.Yu. (2015). Analysis of multivariate statistical models with heterogeneous structure in the case of hidden Markov dependence of states. *Proc. National Academy of Science of Belarus: physics and mathematics series*. Vol. **2**, pp. 26-36 (in Russian).

[8] Malugin V.I. (2015). Algorithms of testing the cyclic structural changes in the vector autoregression models with switching states. *Informatica*. Vol. **4**, pp. 5-16 (in Russian).