# Classification and Prediction of the Gaming Activity States in Online-Games Based on the Regime Switching Models

**V. Malugin, Y.Babakhin**
Faculty of Applied Mathematics and Computer Science, Belarusian State University
Postal address: 4, Nezavisimosti avenue, 220030, Minsk, Republic of Belarus
e-mail: malugin@bsu.by

***Abstract:*** *The goal of the study is to construct an algorithm of determining the overall gaming activity state in online-games. Apart from the classification of states, the problem of predicting future gaming activity states has arisen. Both goals are accomplished by using multivariate econometric models with heterogeneous structure and the assumption of the hidden Markov dependency of the classes of states. In particular, Markov-Switching Vector Autoregression Models (MS-VAR) have been used. The constructed theoretical methods have been applied to the measuring gaming activity in the famous online-game: "World of Tanks".*

***Keywords*:** Online-games, multivariate econometric models, MS-VAR, EM-algorithm, HMM.

## 1. INTRODUCTION

There is a steady growth in video games industry with the development of high-speed Internet, enlargement of various platforms and active marketing strategy of the developers. As a result, the revenue of the industry amounted in \$32.8B in 2015. Moreover, the major percentage of the revenue, about 60%, was accounted for MMOG (Massively-Multiplayer Online Games) [1].

In MMOG thousands of players are interacting with each other online in the virtual world over the long-time period. Our focus during the study has been devoted exactly to such type of online-games.

Multivariate econometric models are widely used for complex systems description in various applications [2]. For a number of different complex systems there are a few regimes of functioning (classes of states). This results in parametric heterogeneity of considered econometric models [3, 4]. Markov-Switching Vector Autoregression Models (MS-VAR) are applied for describing the sequence of independent states starting from the study [5].

In the current study defined models are applied for classifying overall gaming activity states in MMOG. Various phases of activity are used as classes of states in the complex system, e.g. a lack of activity or a stable level of activity.

## 2. MS-VAR MODEL

Let the complex system at the moment $t$ is described by a random vector of observations $y_t \in \Re^N$, which is defined on the probability space $(\Omega, \mathcal{F}, P)$, where $\Omega$ – sample space ($\omega \in \Omega$ – single outcome); $\mathcal{F}$ – $\sigma$ – algebra of subsets from $\Omega$; $P$ – probability measure: $P(A) = P\{\omega \in A\}, A \in \mathcal{F}$.

Let $\{\Omega_1, \ldots, \Omega_M\}$ be a partition of $\Omega$ in the finite number of not empty, disjoint subsets such, that

$$\Omega_m \in \mathcal{F}, P\{\Omega_m\} = P\{\omega \in \Omega_m\} > 0,$$
$$\bigcup_{m \in S(m)} \Omega_m = \Omega, \; S(M) = \{1, \ldots, M\}$$

Let's call subsets $\{\Omega_m\}(m \in S(M))$ the *classes of states* in the complex system. Its number is equal to $M$ .

In the general case it is assumed that the time series $y_t \in \Re^N$ is described by the model MS($M$)–VAR($p$) $(p \geq 1)$ of the following form:

$$y_t = v(s_t) + \sum_{i=1}^{p} A_i(s_t) y_{t-i} + \eta_t(s_t), \; t = 1, \ldots, T, \quad (1)$$

where $y_{1-p}, \ldots, y_0 \in \Re^N$ are predefined initial values; $\eta_t(s_t)$ random vectors of errors defined on $(\Omega, \mathcal{F}, P)$; $(s_t) \in S(M) = \{1, \ldots, M\}$, $M \geq 2$ is the number of classes of states; $A_i(s_t)$ are coefficient matrices of the size $N \times N$ for the corresponding state of the system.

**Model assumptions.**

The model (1) should satisfy the following assumptions [3]:

M1. Coefficient matrices $A_i(m), i = 1, \ldots, p$ satisfy the stability condition of the VAR($p$) model [2] for each class of the states $m \in S(M)$. It should be noted that stable VAR($p$)-process is also a stationary one.

M2. Random vectors of errors $\{\eta_t(m)\}, t = 1, \ldots, T$ are independent, Gaussian random vectors with zero expectation vector and covariance matrix $\Sigma(m) \in \Re^{N \times N}$ for each class of the states $m \in S(M)$.

M3. Model is satisfied the condition of parametric heterogeneity [3]:

$$\left(v(k), A_1(k), \ldots, A_p(k)\right) \neq \left(v(l), A_1(l), \ldots, A_p(l)\right),$$
$$\forall k \neq l; k, l \in S(M)$$

and (or)

$$\Sigma(k) \neq \Sigma(l), \forall k \neq l; k, l \in S(M)$$

The number of the class of states $s_t \in S(M), t = 1, \ldots, T$ should satisfy the following condition:

s1. $s_t, t = 1, \ldots, T$ is a time-homogeneous, irreducible, ergodic Markov chain with the distribution defined by the initial state vector $\pi$ and transition matrix $P$, respectively:

$$\pi = (\pi_m), \pi_m = \mathbf{P}\{s_m = m\} > 0 (m \in S(M)),$$
$$\sum_{m \in S(M)} \pi_m = 1$$
$$P = (p_{kl}),\ p_{kl} = \mathbf{P}\{s_{t+1} = l \mid s_t = k\} > 0$$
$$k, l \in S(M); t = 1, \ldots, T-1$$

## 3. CLASSIFICATION OF GAMING ACTIVITY STATES

For the joint estimation of all model parameters and a vector of classes of states the EM-algorithm has been used [6-8].

To construct estimates of the activity state we have to conduct two following steps:

1) Find the posterior probability distribution of the classes of states $s_t, t = 1, \ldots, T$.

2) Classification of gaming activity states.

Define $\mathrm{Y}_t = (y_t', y_{t-1}', \ldots, y_1')'$ as vector of available observations by the time $t$.

To find the posterior probability distribution of the classes of states the following recursive formula is used:

$$\mathrm{P}(s_t \mid \mathrm{Y}_T) = \sum_{s_{t+1}}^{M} \frac{\mathrm{P}(s_{t+1} \mid s_t)\,\mathrm{P}(s_t \mid \mathrm{Y}_t)}{\mathrm{P}(s_{t+1} \mid \mathrm{Y}_t)} \mathrm{P}(s_{t+1} \mid \mathrm{Y}_T) \quad (2)$$

The estimate of the state at the moment $t$ is found using the following rule:

$$\hat{s}_t = \arg\max_{m \in \{1, \ldots, M\}} \mathrm{P}(s_t = m \mid \mathrm{Y}_T) \quad (3)$$

## 4. "WORLD OF TANKS" EXAMPLE

As an example, consider the classification problem of gaming activity states for the famous MMOG "World of Tanks".

The initial sample consists of CIS region players, who have registered in the game on 29-31 August 2013 (about 10 thousand players). The time period under consideration is about 2.5 years: 2 September 2013 – 1 May 2016 (139 full weeks). The whole time period has been divided into equal parts with one week duration.

Three major measures of players' activity have been chosen for each player:

- *Number of battles.* The total number of battles during the specific week
- *Number of active days.* The total number of days with at least one battle during the specific week
- *Win rate.* The percentage of won battles. It is calculated by a 4-week window.

Thus, for each player there is a multivariate time series with the dimension 3 and 139 observations.

Let's assume that there are three classes of state: the lack of activity ($s_t = 1$); a stable level of activity ($s_t = 3$); an intermediate state of activity ($s_t = 2$).

During the study MS(3)-VAR(2) model has been constructed for each player separately.

However, each player individually is not a good indicator of the overall gaming activity state in the game. So, for each moment of time the aggregated state of the activity in the game is calculated as the average of the number of class for each player. The results are presented in Fig. 1. Moreover, the forecast for the following 8 weeks (May-June 2016) has been constructed (dashed line in Fig. 1).

The next conclusions are followed from the Fig. 1:

- *The overall descending trend.* It is explained by the fact that the certain number of players leave the game during the time
- *Considerable activity growth during the winter holidays.* There are a lot of sales and various events for players over December-January periods.

Thus, the proposed method gives an opportunity not only to explain the activity state of the game but also to measure the success of a certain in-game event.

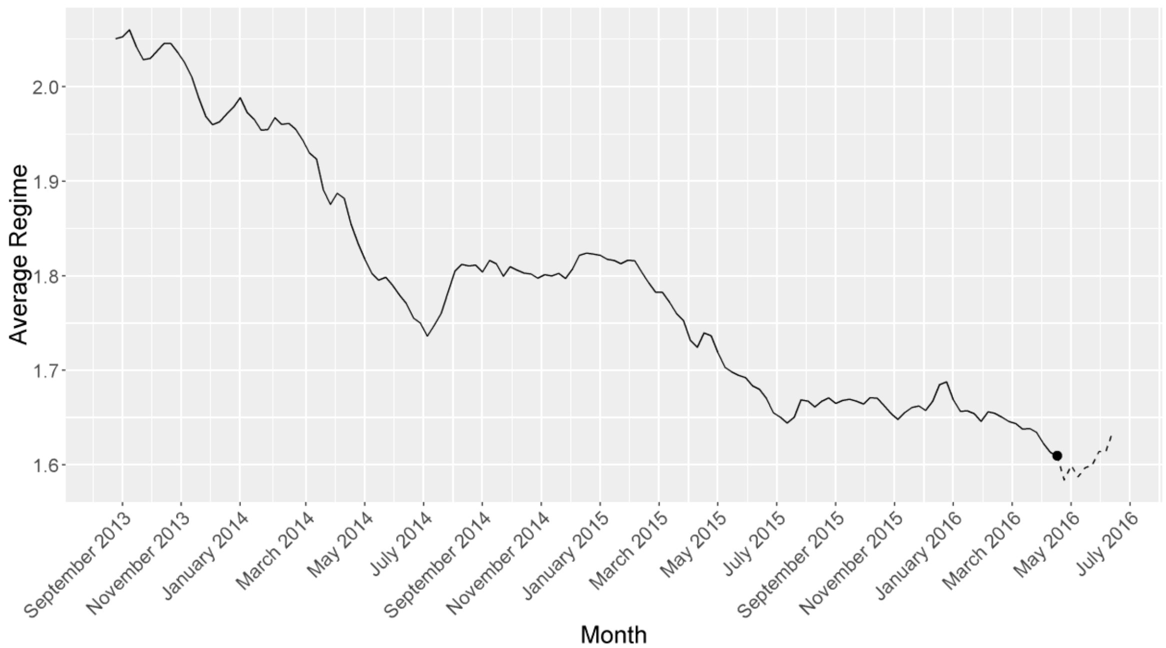


**Fig. 1 – The states of gaming activity in the MMOG "World of Tanks" and its predictions (dashed line).**

The mentioned above "overall descending trend" in activity is connected with a term called "Retention Rate" – the share of players, who has stayed in the game over a specific time period. This measure for our sample equals about 30% on 1 May 2016.

To exclude the effect of another 70% of players consider gaming activity states only for the active players. Similarly, predicted states for the following 8 weeks have been constructed. The results are presented in Fig. 2.

The next conclusions are followed from the Fig. 2:

- *The overall ascending trend*. It is explained by the fact that the most involved group of players is being considered.
- *Considerable activity growth during the winter holidays.* As in the previous case, it follows from the various holiday sales.
- *The rapid growth of activity in March 2016*. The seasonal effect in January 2016 has been followed by a rapid increase in gaming activity in March. Is is explained by the new game update. Update 9.14 has released on 10 March 2016 and introduced new audio mounting and also absolutely new physics of tank movement. These innovations have been positively met by the audience.
- *Prediction of the decrease in activity.* There is a gradual decrease in activity after a reached peak in March 2016. Predicted values also show the downward trend in the gaming activity during May-June 2016.

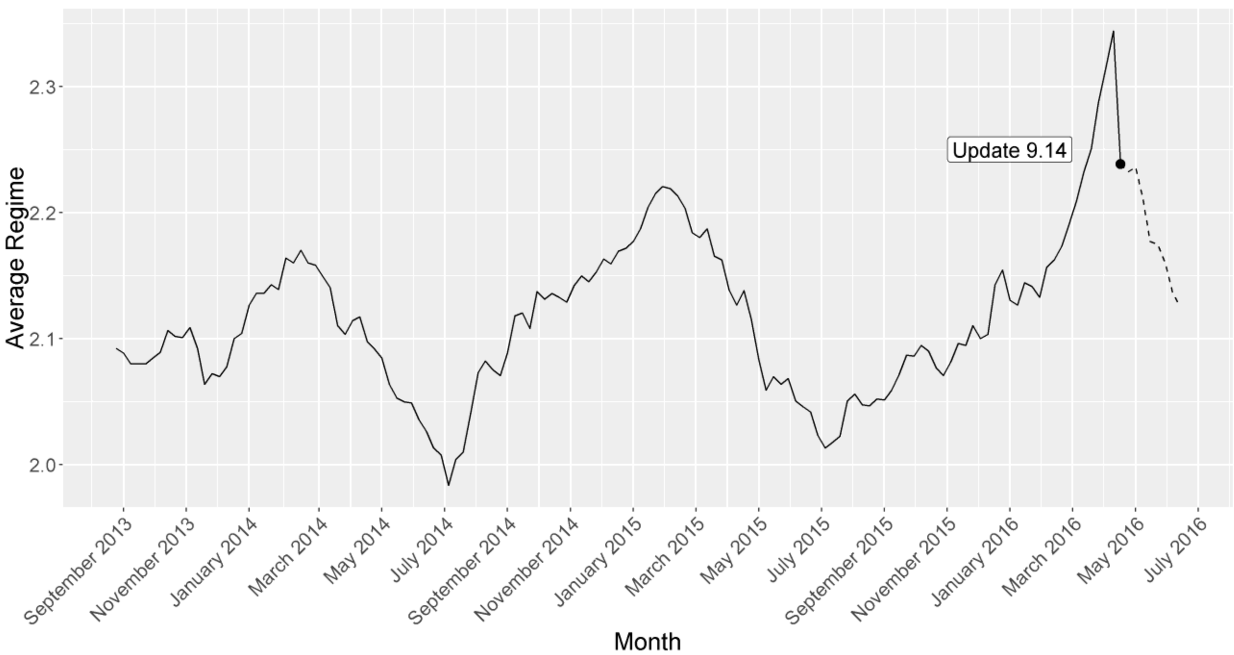


**Fig. 2 – The states of gaming activity in the MMOG "World of Tanks" in terms of the active audience. Also its predictions (dashed line).**

## 5. CONCLUSIONS

During the study, classification rules for estimating the gaming activity states have been proposed. The classification is based on the MS-VAR models. All the implemented results are developed using the R programming language.

Proposed methods and models could be used in various MMOG in order to evaluate the level of success of the specific game events and also to increase the efficiency of the CRM-campaigns.

Moreover, the obtained results have been already implemented in the real projects at Wargming.net – the developer of the game "World of Tanks"

## 6. REFERENCES

[1] J.V. Dreunen. State of the Industry, *SuperData Digital Goods Measurement*. 2016.

[2] H. Lutkepohl. *New Introduction to Multiple Time Series Analysis*. Berlin, 2005. p. 764.

[3] V.I. Malugin, Yu.S. Kharin. *On optimal classification or random observations different in regression equations.* Automation and Remote Control, No 7, 1986, pp. 61-69 (in Russian).

[4] V. Malugin. Analysis of Multivariate Econometric Models with Heterogeneous Structure. Minsk, 2014. p. 351 (in Russian).

[5] J.D. Hamilton. *A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*. Econometrica, Vol. 57, No. 2, 1989. pp. 357–384.

[6] H.-M. Krolzig. *Constructing Turning Points Chronologies with Markov–Switching Vector Autoregressive Models: the Euro–Zone Business Cycle.* Oxford, 2003. 38 p.

[7] V. Malugin, A. Novopoltsev. *Analysis of Multivariate Statistical Models with Heterogeneous Structure and the Assumption of the Hidden Markov Dependency of the Classes of States.* Proceedings of the National Academy of Science of Belarus: physics and mathematics series. Vol. 2, 2015, pp. 26-36 (in Russian).

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