ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 21, Number 1, June 2017
Available online at http://acutm.math.ut.ee

Markov-modulated multivariate linear regression

Alexander Andronov

Abstract. The article concerns parameter estimation for the Markov-modulated multivariate linear regression model. It is supposed that the parameters of the linear regression are dependent from states of a random environment. The last is described as a continuous-time homogeneous irreducible Markov chain with known parameters. The procedure of estimating the regression parameters is established.

1. Introduction

We consider the case where a process, described by multivariate linear regression (Srivastava [6], Turkington [7], Kollo and von Rosen [4]), operates in a random environment. The last is presented (Pacheco et al. [5]) as a continuous-time homogeneous irreducible Markov chain $J(t)$, $t \geq 0$, with finite state set $N = \{1, 2, \ldots, k\}$. Let $\lambda_{i,j}$ be the known transition rate from state i to state j $(\lambda_{j,j} = 0)$.

The following notation will be used for the η-th observation ($\eta = 1, \ldots, n$): $x^*_{(\eta)} = (x^*_{\eta,1}, \ldots, x^*_{\eta,q})$ is the q-row vector of known independent variables, $Y^*_{(\eta)} = (Y^*_{\eta,1}, \ldots, Y^*_{\eta,p})$ is the p-row vector of observed dependent variables, $Z_{(\eta)} = (Z_{\eta,1}, \ldots, Z_{\eta,p})$ is the p-row vector of random variables, $Z_{(\eta)} \in N_p(0, \Sigma)$, t_η is the observation time, and $T_{\eta,j}$ is an unobserved sojourn time in the state $j \in N$, $(T_{\eta,1} + \cdots + T_{\eta,k} = t_\eta)$. Further, $B(j) = (b_{v,\mu}(j))$ is the $q \times p$-matrix of regression parameters for the j-th state of the random environment $(j = 1, \ldots, k)$ and $B = (B(1)^T, \ldots, B(k)^T)^T$ is the $kq \times p$-matrix. Let the positive definite matrix $\Sigma = (\sigma_{v,\mu})_{p \times p}$ be unknown and identical for all observations.

Received September 30, 2016.
2010 *Mathematics Subject Classification.* G2J05; G2J12.
Key words and phrases. Continuous-time Markov chain; multivariate linear regression; point estimation.
http://dx.doi.org/10.12697/ACUTM.2017.21.03

Thus, if $T_{(\eta)} = (T_{\eta,1}, \dots, T_{\eta,k})$, then we have the model for the η-th observation:

$$Y^*_{(\eta)} = \left(Y^*_{\eta,1}, \dots, Y^*_{\eta,p}\right) = (T_{(\eta)} \otimes x^*_{(\eta)})B + \sqrt{t_\eta} Z_{(\eta)}, \quad \eta = 1, \dots, n.$$

Using notation

$$x_{(\eta)} = x^*_{(\eta)}/\sqrt{t_\eta}, Y_{(\eta)} = Y^*_{(\eta)}/\eta,$$

we rewrite this formula as

$$Y_{(\eta)} = (Y_{\eta,1}, \dots, Y_{\eta,p}) = (T_{(\eta)} \otimes x_{(\eta)})B + Z_{(\eta)}, \quad \eta = 1, \dots, n. \tag{1.1}$$

Further, we use the notation

$$Z = \begin{pmatrix} Z_{(1)} \\ \dots \\ Z_{(n)} \end{pmatrix} = \begin{pmatrix} Z_{1,1} & \dots & Z_{1,p} \\ \dots & \dots & \dots \\ Z_{n,1} & \dots & Z_{n,p} \end{pmatrix}.$$

Now the general model is of the form

$$Y = \begin{pmatrix} Y_{1,1} & \dots & Y_{1,p} \\ \dots & \dots & \dots \\ Y_{n,1} & \dots & Y_{n,p} \end{pmatrix} = \begin{pmatrix} T_{(1)} \otimes x_{(1)} \\ \dots \\ T_{(n)} \otimes x_{(n)} \end{pmatrix} B + Z. \tag{1.2}$$

Let us use the notation

$$V_{(\eta)} = (V_{\eta,1}, \dots, V_{\eta,p}) = T_{(\eta)} - E(T_{(\eta)}) = (T_{\eta,1} - E(T_{(\eta,1)}), \dots, T_{\eta,p} - E(T_{(\eta,p)})).$$

Then

$$Y = \begin{pmatrix} Y_{(1)} \\ \dots \\ Y_{(n)} \end{pmatrix} = \begin{pmatrix} E(T_{(1)}) \otimes x_{(1)} \\ \dots \\ E(T_{(n)}) \otimes x_{(n)} \end{pmatrix} B + Z + \begin{pmatrix} V_{(1)} \otimes x_{(1)} \\ \dots \\ V_{(n)} \otimes x_{(n)} \end{pmatrix} B. \tag{1.3}$$

This expression is the main one for the statistical analysis. We note that the rows $Z_{(\eta)} = (Z_{\eta,1}, \dots, Z_{\eta,p})$ and $Y_{(\eta)} = (Y_{\eta,1}, \dots, Y_{\eta,p})$ for different observations of matrices Z and Y are independent. The same property holds for the rows $V_{(\eta)} = (V_{\eta,1}, \dots, V_{\eta,p})$ of matrix V. Further, we get formulas for the expectation and the covariance matrix of the response (1.1):

$$E(Y_{(\eta)}) = (E(T_{(\eta)}) \otimes x_{(\eta)})B,$$

$$\begin{aligned}
&\mathrm{Cov}(T_{(\eta)} \otimes x_{(\eta)}) \\
&\quad = E((T_{(\eta)} \otimes x_{(\eta)} - E(T_{(\eta)}) \otimes x_{(\eta)})^T (T_{(\eta)} \otimes x_{(\eta)} - E(T_{(\eta)}) \otimes x_{(\eta)})) \\
&\quad = E(((T_{(\eta)} - E(T_{(\eta)})) \otimes x_{(\eta)})^T ((T_{(\eta)} - E(T_{(\eta)})) \otimes x_{(\eta)})) \\
&\quad = E((T_{(\eta)} - E(T_{(\eta)}))^T (T_{(\eta)} - E(T_{(\eta)}))) \otimes (x^T_{(\eta)} x_{(\eta)}) \\
&\quad = \mathrm{Cov}(T_{(\eta)}) \otimes (x^T_{(\eta)} x_{(\eta)}),
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(Y_{(\eta)}) &= \mathrm{Cov}((T_{(\eta)} \otimes x_{(\eta)})B) + \Sigma \\
&= B^T (\mathrm{Cov}(T_{(\eta)}) \otimes (x^T_{(\eta)} x_{(\eta)}))B + \Sigma.
\end{aligned} \tag{1.4}$$

In particular,

$$E(Y_{\eta,\mu}) = (E(t_{(\eta)}) \otimes x_{(\eta)})B^{<\mu>},$$
$$D(Y_{\eta,\mu}) = (B^{<\mu>})^T(\mathrm{Cov}(T_{(\eta)}) \otimes (x^T_{(\eta)}x_{(\eta)}))B^{<\mu>} + \sigma^2_\mu, \tag{1.5}$$

where $B^{<\mu>}$ denotes the μ-th column of matrix B. Let

$$Y^{<\mu>} = \begin{pmatrix} Y_{1,\mu} \\ \dots \\ Y_{n,\mu} \end{pmatrix} = \begin{pmatrix} T_{(1)} \otimes x_{(1)} \\ \dots\dots\dots \\ T_{(n)} \otimes x_n \end{pmatrix} B^{<\mu>} + Z^{<\mu>}. \tag{1.6}$$

As $Y_{(\eta)} = (Y_{\eta,1}, \dots, Y_{\eta,p})$, $\eta = 1, \dots, n$, are independent, we have

$$\mathrm{Cov}(Y^{<\mu>}) = \mathrm{diag}(D(Y_{1,\mu}), \dots, D(Y_{n,\mu}))$$

and we can use formula (1.5).

Now let $v \neq \mu$. Then from (1.4) it follows that

$$\begin{aligned} \mathrm{Cov}(Y_{\eta,\mu}, Y_{\eta,v}) &= \mathrm{Cov}((T_{(\eta)} \otimes x_{(\eta)})B^{<\mu>}, (T_{(\eta)} \otimes x_{(\eta)})B^{<v>}) \\ &\quad + \mathrm{Cov}(Z_{\eta,\mu}, Z_{\eta,v}) \\ &= (B^{<\mu>})^T \mathrm{Cov}(T_{(\eta)} \otimes x_{(\eta)})B^{<v>} + \sigma_{\mu,v}. \end{aligned} \tag{1.7}$$

The last formulas determine the covariance matrix $\mathrm{Cov}(Y^{<\mu>}, Y^{<v>})$ for arbitrary μ and v.

We consider estimators $\widetilde{B}$ and $\widetilde{\Sigma}$ of B and $\Sigma = (\sigma^2_{v,\mu})$ for the following data on n observations, $\eta = 1, ..., n$: the vectors $Y_{(\eta)}$ and $x_{(\eta)}$ denote dependent and independent variables, respectively, observation time is t_η, the initial state of Markov chain is $J(0)$ and finite state is $J(t_n)$. It is supposed that all n observations are independent.

Note that a case of multiple regression was considered by Andronov [1].

We need to know expressions for $E(T_{(\eta)})$ and $\mathrm{Cov}(T_{(\eta)})$ for statistical inference. Therefore a random environment must be considered.

2. Random environment as Markov chain

Let $\lambda = (\lambda_{i,j})$ be a $k \times k$ matrix, $\Lambda = \mathrm{diag}\left(\sum_j \lambda_{i,j}\right)$ be a diagonal matrix, $P_{i,j}(t) = P\{X(t) = j|X(0) = i\}$ be the transition probability of Markov chain $X(t)$, and let $P(t) = (P_{i,j}(t))_{k\times k}$ denote the corresponding matrix. If all eigenvalues of the matrix $A = \lambda - \Lambda$ are different, then probabilities $P(t) = (P_{i,j}(t))_{k\times k}$ can be represented simply. Let γ_η and χ_η, $\eta = 1, ..., k$, be the eigenvalue and the corresponding eigenvector of A, $\chi = (\chi_1, ..., \chi_k)$ be the matrix of the unit length eigenvectors, and $\bar{\chi} = \chi^{-1} = (\bar{\chi}^T_1, ..., \bar{\chi}^T_k)$ be the corresponding inverse matrix (here $\bar{\chi}_\eta$ is the η-th row of $\bar{Z}$). Then (Bellman [3], Pacheco et al. [5])

$$P(t) = \exp(tA) = \chi \operatorname{diag}(\exp(\gamma_1 t), \ldots, \exp(\gamma_k t))\chi^{-1}$$
$$= \sum_{\eta=1}^{k} \chi_\eta \exp(\gamma_\eta t)\bar{\chi}_\eta.$$

It is known that for the considered Markov chain one eigenvalue equals 0 (let it be the first), and other eigenvalues (with numbers 2,…,k) are negative.

Let us fix the initial state i and the final state j of the Markov chain $X(t)$ and consider the sojourn time $T_v(t)$ in the state $v \in N$ on the interval $(0, t)$. Then for the conditional expectation

$$\tau_v(t, i, j) = E(T(t)_v | X(0) = i, \ \ X(t) = j)$$

we have

$$\tau_v(t, i, j) = \frac{1}{P_{i,j}(t)} \int_0^t P_{i,v}(u) P_{v,j}(t-u) du, \quad v = 1, \ldots, k. \tag{2.1}$$

Further,

$$\int_0^t P_{i,v}(u) P_{v,j}(t-u) du$$
$$= \int_0^t \sum_{\eta=1}^{m} \chi_{i,\eta} \exp(\gamma_\eta u) \bar{\chi}_{\eta,v} \sum_{\theta=1}^{m} \chi_{v,\theta} \exp(\gamma_\theta (t-u)) \bar{\chi}_{\theta,j} du$$
$$= \sum_{\eta=1}^{m} \chi_{i,\eta} \bar{\chi}_{\eta,v} \sum_{\theta=1,\theta\neq\eta}^{m} \chi_{v,\theta} \bar{\chi}_{\theta,j} \exp(\gamma_\theta t) \frac{1}{\gamma_\theta - \gamma_\eta}$$
$$\times (1 - \exp(-t(\gamma_\theta - \gamma_\eta))) + t \sum_{\eta=1}^{m} \chi_{i,\eta} \bar{\chi}_{\eta,v} \chi_{v,\eta} \bar{\chi}_{\eta,j} \exp(\gamma_\eta t)$$
$$= t \sum_{\eta=1}^{m} \chi_{i,\eta} \bar{\chi}_{\eta,v} \chi_{v,\eta} \bar{\chi}_{\eta,j} \exp(\gamma_\eta t)$$
$$+ \sum_{\eta=1}^{m} \chi_{i,\eta} \bar{\chi}_{\eta,v} \sum_{\theta=1,\theta\neq\eta}^{m} \chi_{v,\theta} \bar{\chi}_{\theta,j} \frac{1}{\gamma_\theta - \gamma_\eta} (\exp(\gamma_\theta t) - \exp(\gamma_\eta t)).$$

Now we can apply formula (2.1). The conditional mixed second order moments

$$\tau_{v,v^*}(t, i, j) = E(T_v(t) T_{v^*}(t) | X(0) = i, \ \ X(t) = j)$$

and conditional covariance

$$C_{v,v^*}(t, i, j) = \operatorname{Cov}(T_v(t), T_{v^*}(t) | X(0) = i, \ X(t) = j)$$

of the sojourn time in the states $v, v^* \in N$ on the interval $(0,t)$ are calculated as

$$\tau_{v,v^*}(t,i,j) = \frac{1}{P_{i,j}(t)}\left(\int_0^t P_{i,v}(u)P_{v,j}(t-u)\tau_{v^*}(t-u,v,j)du\right.$$
$$\left.+\int_0^t P_{i,v^*}(u)P_{v^*,j}(t-u)\tau_v(t-u,v^*,j)du\right),$$
$$C_{v,v^*}(t,i,j) = \tau_{v,v^*}(t,i,j) - \tau_v(t,i,j,)\tau_{v^*}(t,i,j).$$

3. Method of least squares

We begin with the estimation of the regression coefficients $b_{\varsigma,\mu}(j)$, $j = 1,\ldots,k$, $\varsigma = 1,\ldots,q$, $\mu = 1,\ldots,p$. Let

$$H = \begin{pmatrix} E(T_{(1)}) \otimes x_{(1)} \\ \ldots\ldots\ldots\ldots \\ E(T_{(n)}) \otimes x_{(n)} \end{pmatrix}_{n \times kq}. \tag{3.1}$$

The ordinary-least-squares (OLS) estimator for the multivariate linear regression (1.3) is the following (Srivastava [6], p. 279):

$$\widetilde{B} = (H^T H)^{-1} H^T Y.$$

This estimator is unbiased, because it follows from (1.2) and (3.1) that

$$E(Y) = HB.$$

The iterative joint-generalized-least-squares estimator takes into account unequal weights of the observations (Turkington [7]). The iterative procedure estimates alternately the regression parameters B and the covariance matrices $\mathrm{Cov}(Y_{(\eta)})$, $\eta = 1,...,n$, $\mathrm{Cov}(Y^{<\mu>})$, $\mu = 1,...,p$, of the responses Y and Σ of the random term Z.

The procedure begins with the OLS $\widetilde{B}$. On the first step, estimation of the covariance matrices is based on $\widetilde{B}$. For that, let us consider residuals from (1.3):

$$U(B)_{(\eta)} = Y_{(\eta)} - (E(T_{(\eta)}) \otimes x_{(\eta)})B$$
$$= (V_{(\eta)} \otimes x_{(\eta)})B + Z_{(\eta)}, \quad \eta = 1,\ldots,n,$$
$$\mathrm{Cov}(U(B)_{(\eta)}) = \mathrm{Cov}((V_{(\eta)} \otimes x_{(\eta)})B) + \mathrm{Cov}(Z_\eta).$$

The estimator $\widetilde{\mathrm{Cov}}(Y_{(\eta)})$ of $\mathrm{Cov}(Y_{(\eta)})$ is calculated as follows. If $y_{(\eta)}$ and

$$u(B)_{(\eta)} = y_{(\eta)} - (E(T_{(\eta)}) \otimes x_{(\eta)})B$$

are the observed values of $Y_{(\eta)}$ and

$$U(B)_{(\eta)} = Y_{(\eta)} - (E(T_{(\eta)}) \otimes x_{(\eta)})B,$$

then

$$\widetilde{\mathrm{Cov}}(Y_{(\eta)}) = \mathrm{Cov}(U(\widetilde{B})_{(\eta)}) = (u(\widetilde{B})_{(\eta)})^T u(\widetilde{B})_{(\eta)},$$
$$\widetilde{\Sigma}(\widetilde{B}, \eta) = \mathrm{Cov}(U(\widetilde{B})_{(\eta)}) - \mathrm{Cov}((V_{(\eta)} \otimes x_{(\eta)})\widetilde{B}),$$

where $\widetilde{\Sigma}(\widetilde{B}, \eta)$ is the estimator of Σ calculated for fixed $\widetilde{B}$ and the η-th observation.

As

$$\mathrm{Cov}(V_{(\eta)} \otimes x_{(\eta)}) = \mathrm{Cov}(T_{(\eta)}) \otimes (x^T_{(\eta)} x_{(\eta)}),$$

we get that

$$\widetilde{\Sigma}(\widetilde{B}, \eta) = (u(\widetilde{B})_{(\eta)})^T (u(\widetilde{B})_{(\eta)} - \widetilde{B}^T (\mathrm{Cov}(T_{(\eta)}) \otimes (x^T_{(\eta)} x_{(\eta)}))\widetilde{B}$$

and

$$\begin{aligned}\widetilde{\Sigma}(\widetilde{B}) &= \frac{1}{n}\sum_{\eta=1}^{n} \widetilde{\Sigma}(\widetilde{B}, \eta) \\ &= \frac{1}{n}\sum_{\eta=1}^{n} [(u(\widetilde{B})_{(\eta)})^T u(\widetilde{B})_{(\eta)} - \widetilde{B}^T \mathrm{Cov}(T_{(\eta)}) \otimes (x^T_{(\eta)} x_{(\eta)})\widetilde{B}].\end{aligned}$$

Finally we correct the previous estimator of $\mathrm{Cov}(Y_{(\eta)})$ with respect to (1.4):

$$\widetilde{\mathrm{Cov}}(Y_{(\eta)}) = \widetilde{B}^T (\mathrm{Cov}(T_{(\eta)}) \otimes (x^T_{(\eta)} x_{(\eta)}))\widetilde{B} + \widetilde{\Sigma}(\widetilde{B}).$$

The second step consists in estimation of B. Let us describe the corresponding procedure suggested by Turkington [7] , p. 114. We remind that if A is an $m \times n$-matrix and a_i is its i-th column, then vec$A$ denotes the mn-column vector

$$\mathrm{vec}A = \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix}.$$

Let $y = \mathrm{vec}Y$, $z = \mathrm{vec}Z$ be the np-column vectors, and let $\beta = \mathrm{vec}B$ be the qkp-column vector. Let $\vec{H}_{np\times kqp}$ and $\vec{V}_{np\times kqp}$ be the block-diagonal matrices (Kollo and von Rosen [4], p. 73) with H and V, respectively, in the all p diagonal positions. Here H is given in (3.1) and

$$V = \begin{pmatrix} (T_{(1)} - E(T_{(1)})) \otimes x_{(1)} \\ \dots\dots\dots\dots\dots\dots \\ (T_{(n)} - E(T_{(n)}) \otimes x_{(n)} \end{pmatrix}_{n\times kq}.$$

Then the model (1.3) is presented as

$$y = \vec{H}\beta + \vec{V}\beta + z.$$

Further, let $C(\widetilde{B})_{\mu,v} = \widetilde{\mathrm{Cov}}(Y^{<\mu>}, Y^{<v>})$. It is a diagonal $n \times n$ matrix with the estimators of $\mathrm{Cov}(Y_{\eta,\mu}, Y_{\eta,v})$, $\eta = 1, ..., n$, on the main diagonal (see formula (1.7)). The covariance matrix of the vector y is represented as a partitioned $np \times np$ matrix with $C(\widetilde{\beta})_{\mu,v}$ on the (μ, ν)-th position:

$$\widetilde{\mathrm{Cov}}(y) = \begin{pmatrix} C(\widetilde{\beta})_{1,1} & C(\widetilde{\beta})_{1,2} & \dots & C(\widetilde{\beta})_{1,p-1} & C(\widetilde{\beta})_{1,p} \\ C(\widetilde{\beta})_{2,1} & C(\widetilde{\beta})_{2,2} & \dots & C(\widetilde{\beta})_{2,p-1} & C(\widetilde{\beta})_{2,p} \\ \dots & \dots & \dots & \dots & \dots \\ C(\widetilde{\beta})_{p-1,1} & C(\widetilde{\beta})_{p-1,2} & \dots & C(\widetilde{\beta})_{p-1,p-1} & C(\widetilde{\beta})_{p-1,p} \\ C(\widetilde{\beta})_{p,1} & C(\widetilde{\beta})_{p,2} & \dots & C(\widetilde{\beta})_{p,p-1} & C(\widetilde{\beta})_{p,p} \end{pmatrix}.$$

The joint-generalized-least-squares estimator is the following (Turkington [7], p. 114):

$$\widetilde{\beta} = [\vec{H}^T \widetilde{\mathrm{Cov}}(y) \vec{H}]^{-1} \vec{H}^T \widetilde{\mathrm{Cov}}(y)\ y.$$

The next iteration begins from the first step using the last estimate $\widetilde{\beta}$ and so on. The iterative procedure ends when the estimate changes become small.

4. Discussion

Our previous experience (Andronov [2]) has shown that the obtained estimates converge to the true values of the parameters very slowly. We cannot use maximum likelihood method because the density of the distribution of these sojourn times is unknown. We know only the Laplace transformation. In future we plan to use this transformation for parameter estimation directly as we have done in the paper Andronov [2].

Acknowledgment

The author thanks the anonymous referee for careful reading and valuable comments.

References

[1] A. Andronov, *Parameter statistical estimates of Markov-modulated linear regression*, in: Statistical Methods of Parameter Estimation and Hypothesis Testing **24**, Perm State University, Perm, Russia, 2012, pp. 163–180. (Russian)

[2] A. Andronov, *On a reward estimation for the finite irreducible continue time Markov chain*, J. Statist. Theory Practice, 2017, in press.

[3] R. Bellman, *Introduction to Matrix Analysis*, McGraw Hill Book Co., Inc., New York – Toronto – London, 1960.

[4] T. Kollo and D. von Rosen, *Advanced Multivariate Statistics with Matrices*, Springer, Dordrecht, 2005.

[5] A. Pacheco, L. C. Tang, and N. U. Prabhu, *Markov-Modulated Processes & Semiregenerative Phenomena.*, World Scientific Publishing Co. Pte. Ltd., Hackensack, New York, 2009.

[6] M. S. Srivastava, *Methods of Multivariate Statistics*, Wiley-Interscience [John Wiley & Sons], New York, 2002.

[7] D. A. Turkington, *Matrix Calculus* & *Zero-One Matrices. Statistical and Econometric Applications*, Cambridge University Press, New York, 2002.

Transport and Telecommunication Institute, Lomonosova Str.1, Riga, Latvia
E-mail address: `lora@mailbox.riga.lv`