МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Международная лаборатория статистики случайных
процессов и количественного финансового анализа

Международная научная конференция «Робастная статистика и финансовая математика – 2017»

(03–05 июля 2017 г.)

Сборник статей

Под редакцией
д-ра физ.-мат. наук, профессора С.М. Пергаменщикова,
канд. физ.-мат. наук, доцента Е.А. Пчелинцева

Томск
Издательский Дом Томского государственного университета
2017

Model selection for a semi - Markov continuous time regression observed in the discrete time moments *

Barbu V.S., Beltaief S., Pergamenshchikov S.

Université de Rouen Normandie, Rouen,
Tomsk State University, Tomsk
e-mail: Serge.Pergamenshchikov@univ-rouen.fr

Abstract

In this article we consider the nonparametric robust estimation problem for regression models in continuous time with semi-Markov noises observed in discrete time moments. An adaptive model selection procedure is proposed. A sharp non-asymptotic oracle inequality for the robust risks is obtained. We obtain sufficient conditions on the frequency observations under which the robust efficiency is shown. It turns out that for the semi-Markov models the robust minimax convergence rate may be faster or slower than the classical one.

Keywords: Non-asymptotic estimation; Robust risk; Model selection; Sharp oracle inequality; Asymptotic efficiency

In this paper we consider the semi-Markov regression model in continuous time introduced in [1], i.e.

$$d\,y_t = S(t)d\,t + d\,\xi_t\,,\quad 0\le t\le n\,,\tag{1}$$

where $S(\cdot)$ is an unknown 1-periodic function defined on $\mathbf{R}$ with values on $\mathbf{R}$, $(\xi_t)_{t\ge 0}$ is the unobserved noise process defined through a certain semi-Markov process in [1].

Our problem in the present paper is to estimate the unknown function S in the model (1) on the basis of observations

$$(y_{t_j})_{0\le j\le np},\quad t_j=j\Delta,\quad \Delta=\frac{1}{p},\tag{2}$$

where the integer $p\ge 1$ is the observation frequency. Firstly, this problem was considered in the framework "signal+white noise" (see, for example, [3] or [11]). Later, to introduce a dependence in the continuous time regression model in [10], [6], [4], [5] [7], the Ornstein-Uhlenbeck processes has been used to model the "color noise". Moreover, in order to introduce the dependence and the jumps in the regression model

*The work was supported by the RSF grant 17-11-01049.

(1), the papers [8] and [9] use the non Gaussian Ornstein-Uhlenbeck processes defined in [2]. The problem in all these papers is that the introduced Ornstein-Uhlenbeck type of dependence decreases with a geometric rate. So, asymptotically when the duration of observations goes to infinity, we obtain the same "signal+white noise" model very quick. To keep the dependence for sufficiently large duration of observations, in [1] it was proposed the model (1) with a semi-Markov component in the jumps of the noise process $(\xi_t)_{t\ge 0}$.

The main goal of this paper is to develop adaptive robust method from [1], that was based on continuous observations, to the estimation problem based on discrete observations given in (2). In this paper we use quadratic risk defined as

$$\mathcal{R}_Q(\tilde{S}_n, S) = \mathbf{E}_{Q,S}\, \|\tilde{S}_n - S\|^2\,, \tag{3}$$

where $\tilde{S}_n(\cdot)$ is some estimate (i.e. any periodical function measurable with respect to the observations $\sigma\{y_{t_0},\ldots y_{t_{pn}}\}$), $\|f\|^2 = \int_0^1 f^2(s)ds$ and $\mathbf{E}_{Q,S}$ is the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (1) corresponding to the unknown noise distribution Q in the Skorokhod space $\mathcal{D}[0,n]$. We assume that this distribution belongs to some distribution family $\mathcal{Q}_n$ specified in [1].

To study the properties of the estimators uniformly over the noise distribution (what is really needed in practice), we use the robust risk defined as

$$\mathcal{R}^*_n(\tilde{S}_n, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\tilde{S}_n, S)\,. \tag{4}$$

Thus the goal of this paper is to develop a robust efficient model selection method based on the observations (2) for the model (1) with the semi-Markov components in the jumps of the noise $(\xi_t)_{t\ge 0}$. We use the approach proposed by Konev and Pergamenshchikov in [9] for continuous-time regression models observed in the discrete time moments. Unfortunately, we cannot use directly this method for semi-Markov regression models, since their tool essentially uses the fact that the Ornstein-Uhlenbeck dependence decreases with geometrical rate and obtain sufficiently quickly the "white noise" case. In the present paper, in order to obtain the sharp non-asymptotic oracle inequalities, we use the renewal methods from [1] developed for the model (1). As a consequence, we can obtain the constructive sufficient conditions that provide the robust efficiency for proposed model selection procedures.

In this paper we construct some special family of the weight least square estimators $(\hat{S}_\lambda)_{\lambda\in\Lambda}$, where Λ is some finite set from $[0,1]^n$. In the sequel we assume that the number of the vector of the $\nu = \mathrm{card}\Lambda$

is a function of n, i.e. $\nu = \nu_n$, such that for any $\gamma > 0$

$$\lim_{n\to\infty} \frac{\nu}{n^\gamma} = 0\,. \tag{5}$$

The model select procedure for this family is defined as

$$\hat{S}_* = \hat{S}_{\hat{\lambda}} \quad \text{and} \quad \hat{\lambda} = \text{argmin}_{\lambda\in\Lambda} J_{n,\delta}(\lambda)\,, \tag{6}$$

where $J_{n,\delta}(\lambda)$ is the cost function defined for any in [1] and $0 < \delta < 1$ is the penalty parameter.

We assume that the renewal distribution in the semi-markov component of the noise process in (1) has a density g that satisfies the following conditions.

$\mathbf{H}_1$) *Assume that, for any $x \in \mathbf{R}$, there exist the finite limits*

$$g(x-) = \lim_{z\to x-} g(z) \quad \textit{and} \quad g(x+) = \lim_{z\to x+} g(z)$$

and, for any $K > 0$, there exists $\delta = \delta(K) > 0$ for which

$$\sup_{|x|\le K} \int_0^\delta \frac{|g(x+t) + g(x-t) - g(x+) - g(x-)|}{t}\, dt \,<\, \infty. \tag{7}$$

$\mathbf{H}_2$) *For any $\gamma > 0$,*

$$\sup_{z\ge 0} z^\gamma |2g(z) - g(z-) - g(z+)| \,<\, \infty\,.$$

$\mathbf{H}_3$) *There exists $\beta > 0$ such that $\int_{\mathbf{R}} e^{\beta x}\, g(x)\, dx < \infty$.*

$\mathbf{H}_4$) *There exists $t^* > 0$ such that the function $\hat{g}(\theta - it)$ belongs to $\mathbf{L}_1(\mathbf{R})$ for any $0 \le t \le t^*$, where*

$$\hat{g}(\theta) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{i\theta x}\, g(x)\, dx\,.$$

Theorem 1. *Assume that the function S is continuously differentiable and that Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) hold true. Then there exists some constant $l^* > 0$ such that for any noise distribution Q, the weight vectors set Λ, for any periodic function S for any $n \ge 1$, $p \ge 3$ and $0 < \delta \le 1/6$, the procedure (6) satisfies the following oracle inequality*

$$\mathcal{R}_Q(\hat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\hat{S}_\lambda, S) + l^* \frac{\nu}{\delta n} \tag{8}$$

and

$$\mathcal{R}^*(\hat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(\hat{S}_\lambda, S) + \frac{U_n^*}{\delta n}\,, \tag{9}$$

where the the rest term U_n^* is such that for any $\gamma > 0$

$$\lim_{n\to\infty} \frac{U_n^*}{n^\gamma} = 0\,.$$

Remark 1. *The oracle inequalities* (8) *and* (9) *are called sharp oracle inequalities, since the coefficients in the main terms in the right side closed to one. As we will see below this property allows to show asymptotic efficiency in the adaptive setting.*

In order to obtain the efficiency property, we specify the weight coefficients $(\lambda(j))_{1\le j\le n}$ in the procedure (6). Consider, for some fixed $0<\varepsilon<1$, a numerical grid of the form

$$\mathcal{A}=\{1,\ldots,k^*\}\times\{\varepsilon,\ldots,m\varepsilon\}, \tag{10}$$

where $m=[1/\varepsilon^2]$. We assume that both parameters $k^*\ge 1$ and ε are functions of n, i.e. $k^*=k^*(n)$ and $\varepsilon=\varepsilon(n)$, such that

$$\begin{cases} \lim_{n\to\infty} k^*(n)=+\infty\,, \quad \lim_{n\to\infty}\dfrac{k^*(n)}{\ln n}=0\,, \\ \\ \lim_{n\to\infty}\varepsilon(n)=0 \quad \text{and} \quad \lim_{n\to\infty} n^{\gamma}\varepsilon(n)\,=+\infty \end{cases} \tag{11}$$

for any $\gamma>0$. One can take, for example, for $n\ge 2$

$$\varepsilon(n)=\frac{1}{\ln n} \quad \text{and} \quad k^*(n)=k^*_0+\sqrt{\ln n}\,, \tag{12}$$

where $k^*_0\ge 0$ is some fixed constant. For each $\alpha=(\beta,\tau)\in\mathcal{A}$, we introduce the weight sequence

$$\lambda_\alpha=(\lambda_\alpha(j))_{1\le j\le p}$$

with the elements

$$\lambda_\alpha(j)=\mathbf{1}_{\{1\le j<j_*\}}+\left(1-(j/\omega_\alpha)^{\beta}\right)\,\mathbf{1}_{\{j_*\le j\le\omega_\alpha\}}, \tag{13}$$

where $j_*=1+[\ln \upsilon_n]$, $\omega_\alpha=(\mathbf{d}_\beta\,\tau\upsilon_n)^{1/(2\beta+1)}$,

$$\mathbf{d}_\beta=\frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} \quad \text{and} \quad \upsilon_n=n/\varsigma^*\,.$$

Threshold ς^* is a function of n, i.e. $\varsigma^*=\varsigma^*(n)$ such that for any $\gamma>0$

$$\liminf_{n\to\infty} n^{\gamma}\,\varsigma^*(n)=+\infty \quad \text{and} \quad \limsup_{n\to\infty} n^{-\gamma}\,\varsigma^*(n)=0\,. \tag{14}$$

Now we define the set Λ as

$$\Lambda\,=\,\{\lambda_\alpha\,,\,\alpha\in\mathcal{A}\}\,. \tag{15}$$

These weight coefficients are used in [8, 9] for continuous time regression models to show the asymptotic efficiency. Note also that in this case the cardinal of the set Λ is $\nu=k^*m$. Therefore, the conditions in (11) yield the property (5). Moreover, to obtain the efficiency for the model selection procedure we assume the following condition on the frequency of the observations.

$\mathbf{H}_5$) *Assume that for any $n\ge 3$ the observation frequence $p\ge n^{5/6}$.*

To this end, we assume that the unknown function S in the model

(1) belongs to the Sobolev ball

$$W_r^k = \{f \in \mathcal{C}_{per}^k[0,1]\,,\, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le r\}\,, \tag{16}$$

where $r > 0\,,\; k \ge 1$ are some parameters, $\mathcal{C}_{per}^k[0,1]$ is the set of k times continuously differentiable functions $f\,:\,[0,1] \to \mathbf{R}$ such that $f^{(i)}(0) = f^{(i)}(1)$ for all $0 \le i \le k$. The function class W_r^k can be written as an ellipsoid in l_2, i.e.

$$W_r^k = \{f \in \mathcal{C}_{per}^k[0,1]\,:\, \sum_{j=1}^{\infty} a_j\, \theta_j^2\, \le r\} \tag{17}$$

where $a_j = \sum_{i=0}^{k} \left(2\pi[j/2]\right)^{2i}$.

Similarly to [8,9] we will show here that the asymptotic sharp lower bound for the robust risk (4) is given by

$$r_k^* = l\, \left((2k+1)r\right)^{1/(2k+1)} \left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)}\,. \tag{18}$$

Note that this is the well-known Pinsker constant obtained for the nonadaptive filtration problem in “signal + small white noise” model (see, for example, [11]).

Let Π_n be the set of all estimators $\hat{S}_n$ measurable with respect to the sigma-algebra $\sigma\{y_t\,,\, 0 \le t \le n\}$ generated by the process (1).

Theorem 2. *Under Conditions* (14)

$$\liminf_{n\to\infty}\, \upsilon_n^{2k/(2k+1)}\, \inf_{\hat{S}_n\in\Pi_n}\, \sup_{S\in W_r^k}\, \mathcal{R}_n^*(\hat{S}_n, S) \ge r_k^*\,, \tag{19}$$

where $\upsilon_n = n/\varsigma^*$.

Note that, if the parameters r and k are known, i.e. for the nonadaptive estimation case, in order to obtain the efficient estimation for the “signal+white noise” model, Pinsker proposed in [11] to use the estimate $\hat{S}_{\lambda_0}$ with the weights (13) in which

$$\lambda_0 = \lambda_{\alpha_0} \quad \text{and} \quad \alpha_0 = (k, \tau_0)\,, \tag{20}$$

where $\tau_0 = [r/\varepsilon]\varepsilon$. For the semi-markov regression model (1) we show the same result.

Proposition 3. *The estimator* $\hat{S}_{\lambda_0}$ *satisfies the following asymptotic upper bound*

$$\lim_{n\to\infty}\, \upsilon_n^{2k/(2k+1)}\, \sup_{S\in W_r^k}\, \mathcal{R}_n^*(\hat{S}_{\lambda_0}, S) \le r_k^*\,.$$

For the adaptive estimation we user the model selection procedure (6) with the parameter δ defined as a function of n satisfying

$$\lim_{n\longrightarrow\infty}\, \delta_n = 0 \quad \text{and} \quad \lim_{n\longrightarrow\infty}\, n^{\gamma}\, \delta_n = 0 \tag{21}$$

for any $\gamma > 0$. For example, we can take $\delta_n = (6 + \ln n)^{-1}$. Now the oracle inequality (9) for the family (15) and Proposition 3 imply the following upper bound.

Theorem 4. *Assume that Conditions* $\mathbf{H}_1)$–$\mathbf{H}_5)$ *hold true. Then the robust risk defined in* (4) *for the procedure* (6) *with the coefficients* (13) *and the parameter* $\delta = \delta_n$ *satisfying* (21) *has the following asymptotic upper bound*

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_r^k} \mathcal{R}_n^*(\hat{S}_*, S) \le r_k^* \,. \tag{22}$$

Theorem 2 and Theorem 4 imply the following result.

Corollary 5. *Under the conditions of Theorem 4, the model selection procedure* $\hat{S}_*$ *is efficient, i.e.*

$$\lim_{n\to\infty} \frac{\inf_{\hat{S}_n\in\Pi_n} \sup_{S\in W_r^k} \mathcal{R}_n^*(\hat{S}_n, S)}{\sup_{S\in W_r^k} \mathcal{R}_n^*(\hat{S}_*, S)} = 1 \,. \tag{23}$$

Remark 2. *It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball* W_r^k *is* $n^{2k/(2k+1)}$ *(see, for example, [11]). We see here that the efficient robust rate is* $\upsilon_n^{2k/(2k+1)}$*, i.e. if the distribution upper bound* $\varsigma^* \to 0$ *as* $n \to \infty$ *we obtain a faster rate with respect to* $n^{2k/(2k+1)}$*, and if* $\varsigma^* \to \infty$ *as* $n \to \infty$ *we obtain a slower rate. In the case when* ς^* *is constant the robuste rate is the same as the classical non robuste convergence rate.*

Conclusion. In the conclusion we would like to emphasize that in this paper :

- we construct a selection model procedure based on the weight least square estimators;
- we find conditions for which we obtained an sharp non asymptotic oracle inequalities for the simple quadratic risks and for the robust risks as well;
- using the Pinsker method we obtain a lower bound for the robust quadratic risks, then, through the obtained sharp oracle inequalities we show that the risk upper bound for the constructed procedure matters this lower bound, i.e. the procedure is efficient in the adaptative setting.

References

1. Barbu V.S. Robust adaptive efficient estimation for semi-Markov nonparametric regression models / Barbu V.S., Beltaif S., Pergamenchtchikov S.M. // Statistical inference for stochastic processes. 25 March 2017.— (Preprint / Statistical inference for stochastic processes) – URL: https://arxiv.org/pdf/1604.04516.pdf

2. Barndorff-Nielsen O. E. and Shephard N. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial mathematics // J. Royal Stat. Soc. 2001. V. B 63. P. 167—241.

3. Ibragimov I. A. Statistical Estimation: Asymptotic Theory / Ibragimov I. A., Khasminskii R. Z.—Springer, Berlin, New York. 1981.

4. Höpfner R. and Kutoyants Yu. A. On LAN for parametrized continuous periodic signals in a time inhomogeneous diffusion // Statist. Decisions. 2009. V. 27. No. 4. P. 309—326.

5. Höpfner R., Kutoyants Yu. A. Estimating discontinuous periodic signals in a time inhomogeneous diffusion // Statistical Inference for Stochastic Processes. 2010. V.13. No. 3. P. 193—230.

6. Konev V. V., Pergamenshchikov S. M. Sequential estimation of the parameters in a trigonometric regression model with the gaussian coloured noise // Statistical Inference for Stochastic Processes. 2003. V. 6. P. 215—235.

7. Konev V. V., Pergamenshchikov S. M. General model selection estimation of a periodic regression with a Gaussian noise // Annals of the Institute of Statistical Mathematics. 2010. V. 62. P. 1083—1111.

8. Konev V. V., Pergamenshchikov S.M. Efficient robust nonparametric estimation in a semimartingale regression model // Ann. Inst. Henri Poincaré Probab. Stat. 2012. V. 48. No. 4. P. 1217—1244.

9. Konev V.V. and Pergamenshchikov S.M. Robust model selection for a semimartingale continuous time regression from discrete data //Stochastic processes and their applications. 2015. V. 294—326.

10. Kutoyants Yu. A. Estimation of signal parameter in Gaussian noise // Problems of Information Transmission. 1977. V. 13. P. 266-271.

11. Pinsker M. S. Optimal filtration of square integrable signals in gaussian white noise // Problems of Transimission information. 1981. V. 17. P. 120–133.