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

Международная научная конференция «Робастная статистика и финансовая математика – 2019»

(04–06 июля 2019 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
2019

Efficient Nonparametric Estimation of Square-integrable Functions in Continuous Time Regression Models [1]

Povzun M. A., Pchelintsev E. A., Pergamenshchikov S. M.

Tomsk State University, Tomsk,
Rouen University, Rouen, France
e-mail: povzunyasha@gmail.com

Abstract

In this paper we study an asymptotic efficiency property of the weighted least squares estimates for unknown square integrable functions in Gaussian regression models. We use the Pinsker approach. It is established that it is impossible to directly use the Pinsker method. Some conditions and an a priori distribution for the Fourier coefficients are obtained under which the asymptotic lower bound for the mean square risk was proved.

Keywords: regression model, weighted least squares estimates, asymptotic efficiency, Pinsker method.

We will observe the following process

$$\mathrm{d}\,y_t = \theta(t)\mathrm{d}\,t + \varepsilon\mathrm{d}\,W_t\,, \quad 0 \le t \le 1\,. \tag{1}$$

This model is a non-parametric regression in continuous time with noises of small intensity, where $\theta(t) \in \mathcal{L}_2[0,1]$ – unknown function, $(W_t)_{0\le t\le 1}$ is Brownian motions and $\varepsilon > 0$ is the noise intensity. Another words we have got a "signal+white noise" model. Such models are widely used in statistical radio-physics and financial mathematics.

Our goal is to study an asymptotic efficiency property of estimate the unknown function $\theta(t)$ as $\varepsilon \to 0$, i.e. in the case when the ratio "signal/noise" tends to infinity.

The asymptotically efficient estimates for Gaussian models have been constructed in [3, 6, 9]. In [1, 2] was proposed a new approach to study an asymptotic efficiency for estimating problem in the non-parametric models via sharp oracle inequalities method. This method has been developed for semimartingale regression models in continuous time by [4, 5]. For improved estimates of unknown regression functions was expanded such approach in [7, 8].

Estimation of the function will be considered in the sense of mean-square accuracy:

$$\mathcal{R}(\hat\theta,\theta) := \mathbf{E}\|\hat\theta-\theta\|^2 \qquad \text{and} \qquad \|\theta\|^2 = \int_0^1 \theta^2(t)\mathrm{d}\,t\,. \tag{2}$$

[1] The work was supported by the RSF, the project No 17-11-01049.

Let (φ_j) be an orthonormal basis, $(\varphi_j) \in \mathcal{L}_2[0,1]$ and we represent $\theta(t)$ as

$$\theta(t) = \sum_{j=1}^{\infty} \theta_j \varphi_j(t), \tag{3}$$

where the Fourier coefficients can be represented as

$$\theta_j = \int_0^1 \theta(t)\varphi_j(t)\mathrm{d}\, t.$$

Further we need the following set

$$\Theta = \left\{ \theta(t) \in \mathcal{L}_2[0,1] : \sum_{j=1}^{\infty} a_j \theta_j^2 \leqslant \mathbf{r} \right\}, \tag{4}$$

where $\mathbf{r} > 0$.
In order to construct the estimate, one needs to know how many parameters should be choosing. In the sequel we will use the following threshold

$$n_* = \max_{l \geqslant 1} \left\{ \sqrt{a_l} \sum_{j=1}^{l} \sqrt{a_j} - \sum_{j=1}^{l} a_j \leqslant \frac{\mathbf{r}}{\varepsilon^2} \right\}. \tag{5}$$

Note that n_* is finite, since the sums, in view of (4), are bounded by the constant $\mathbf{r}$.

So, we turn on from the non-parametric estimation problem to the parametric one. The Fourier coefficients in (3) can be estimated as

$$\hat{\theta}_j = \int_0^1 \varphi_j(t)\mathrm{d}\, y_t = \theta_j + \varepsilon \xi_j, \tag{6}$$

where

$$\xi_j = \int_0^1 \varphi_j(t)\, \mathrm{d}\, W_t.$$

Now we consider projection estimates in the following form:

$$\tilde{\theta}_j = \alpha_j \hat{\theta}_j,$$

where $(\alpha_j)_{j\geqslant 1}$ is a non-random sequence.

We find such a case that the difference between the estimate and the true value of the function will be minimal.
We set

$$\nu^2 = \inf_{v} \sup_{\theta \in \Theta} \sum_{j=1}^{\infty} \mathbf{E}_\theta \left(\alpha_j \hat{\theta}_j - \theta_j \right)^2, \tag{7}$$

where inf is taken take out all coefficients $v = (v_i)_{i \geqslant 1}$. Using Lemma of Pinscer [?] we can find that

$$\nu^2 = \varepsilon^2 \sum_{j=1}^{n_*} \frac{\theta_j^{*2}}{\theta_j^{*2} + \varepsilon^2} = n_* \varepsilon^2 - \frac{\varepsilon^2 (\sum_{j=1}^{n_*} \sqrt{a_j})^2}{\mathbf{r}/\varepsilon^2 + \sum_{j=1}^{n_*} a_j}. \tag{8}$$

where optimal points $(\theta^*_j)_{1\leqslant j\leqslant n_*}$ are defined as:

$$\theta^{*2}_j = \varepsilon^2\left(\frac{\mu}{\sqrt{a_j}} - 1\right) \quad \text{and} \quad \mu = \frac{\mathbf{r}/\varepsilon^2 + \sum_{j=1}^{n_*} a_j}{\sum_{j=1}^{n_*}\sqrt{a_j}}. \tag{9}$$

The parameter μ is taken from the next condition

$$\sum_{j=1}^{n_*} a_j\theta^*_j = \mathbf{r}. \tag{10}$$

In the future, we will need important conditions for the proof of the theorems:

A$_1$) A sequence $(a_j)_{j\geqslant 1}$ increasing numbers such that $a_j > a_{j-1}$ and

$$\sum_{j=1}^{\infty}\sqrt{a_j}(\sqrt{a_{j+1}} - \sqrt{a_j}) = +\infty;$$

For a fixed $h > 0$

$$N_h = \max_{j\geqslant 1}\{a_j \leqslant h\}. \tag{11}$$

A$_2$) For any $0 < \gamma < 1$,

$$N_h - N_{\gamma h} - \frac{1}{\sqrt{a_{N_h}}}\sum_{j=N_{\gamma h}+1}^{N_h}\sqrt{a_j} \to \infty.$$

Without these conditions, Pinsker theorem cannot be proved.

First we obtain the lower bound.

Theorem 1. *For the model* (1) *the following lower bound holds*

$$\varliminf_{\varepsilon\to 0}\sup_{\theta\in\Theta}\inf_{\hat\theta}\frac{\mathcal{R}(\hat\theta,\theta)}{\nu^2} \geqslant 1. \tag{12}$$

where inf *is taken over all possible estimates* $\hat\theta$*.*

Moreover, consider $\nu\prime 2$ in the following form

$$\nu'^2 = \sum_{j=n_0+1}^{n_1}\frac{\theta^{*2}_j\varepsilon^2}{\theta^{*2}_j + \varepsilon^2}. \tag{13}$$

This is a key property, which is necessary for the proof of the lower bound.

Proposition 1. *The condition* A_2 *imply*

$$\lim_{\varepsilon\to 0}\frac{\nu'^2}{\varepsilon^2} = +\infty, \tag{14}$$

where

$$n_0 = \max_{j\geqslant 1}\left\{a_j \leqslant \frac{\mu^2}{(1+q)^2}\right\} \quad \textit{and} \quad n_1 = \max_{j\geqslant 1}\left\{a_j \leqslant \frac{\mu^2 q^2}{(1+q)^2}\right\}. \tag{15}$$

If $a_j = e^{2j}$, then conditions $A_1) - -A_2)$ are not satisfied, so

$$\lim_{\varepsilon\to 0}\frac{\nu'^2}{\varepsilon^2} < +\infty. \tag{16}$$

It means that the theorem of Pinsker without conditions is not true. To confirm the obtained results, we consider the following example. Let $a_j = j^{2k}$ and k is fixed. Also fix some $q > 0$. Check the condition A_1):

$$\sum_{j=1}^{\infty} \sqrt{j^{2k}}(\sqrt{(j+1)^{2k}} - \sqrt{j^{2k}}) = +\infty\,.$$

As you can see, this condition is satisfied. Next, we check the condition A_2):

$$N_h - N_{\gamma h} - \sum_{j=1}^{N_h} \left(\frac{j}{N_h}\right)^k + \sum_{j=1}^{N_{\gamma h}} \left(\frac{j}{N_h}\right)^k = +\infty\,,$$

we obtain the condition A_2 satisfy. Thus, this coefficients can be used for estimation procedure. Next, we calculate some more characteristics. Note that, in equation (5) for n_*

$$n_* \approx \left(\frac{\mathbf{r}}{\varepsilon^2}\right)^{\frac{1}{2k+1}} \qquad \text{and} \qquad \mu \approx \left(\frac{1}{\varepsilon^2}\right)^{\frac{k}{2k+1}}\,.$$

Note that if we take $a_j = e^{2j}$, we get a counter–example. So, conditions A_1 and A_2 are not satisfied for this.

We estimate the function $\theta(t)$ as

$$\tilde{\theta}(t) = \sum_{j=1}^{n_*} \lambda_j \hat{\theta}_j \varphi_j(t), \qquad \lambda_j = \left(1 - \frac{\sqrt{a_j}}{\mu}\right). \tag{17}$$

Let us proceed to the formulation of the theorem on the upper boundary.

Theorem 2. *The estimator* (17) *admits the following upper bound*

$$\overline{\lim_{\varepsilon\to 0}} \sup_{\theta\in\Theta} \frac{\mathcal{R}(\tilde{\theta},\theta)}{\nu^2} \leqslant 1. \tag{18}$$

Using these theorems, we can formulate the efficiency property of estimates.

Corollary 1. *The theorems 2 and 1 imply*

$$\underline{\lim_{\varepsilon\to 0}} \sup_{\theta\in\Theta} \frac{\mathcal{R}(\tilde{\theta},\theta)}{\nu^2} = 1. \tag{19}$$

In this paper we constructed an estimator for which we showed efficiency property. To this end we use the approach proposed by Pinsker (1980) for this problem. Unfortunately, we cannot use directly the Pinsker method for since his main theorem about lower bound is not true without conditions on the coefficient of the Sobolev ball. We found this conditions for which we provide the efficiency property. The main difficulty is to find the prior distribution for Fourier coefficient. Moreover, we gave the constrictive sufficient conditions for which this property holds. And as example we checked the obtained conditions for Sobolev coefficients which goes to infinity power function.

References

[1] Galtchouk L.I., Pergamenshchikov S.M. Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes // Stat. Inference Stoch. Processes. 2006. Vol.9, No1. P. 1–16.

[2] Galtchouk L., Pergamenshchikov S. Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression // Journal of Korean Statistical Society. 2009. Vol. 38, No 4. P. 305–322.

[3] Ibragimov I. A., Khasminskii R. Z. Statistical Estimation: Asymptotic Theory. – New York: Springer, 1981.

[4] Konev V. V., Pergamenshchikov S. M. Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency // Journal of Mathematics and Mechanics of Tomsk State University. 2009. No 4. P. 31–45.

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

[6] Kutoyants Yu. A. Identification of dynamical systems with small noise. – Dordrecht: Kluwer Academic Publishers Group, 1994.

[7] Pchelintsev E., Pergamenshchikov S. Oracle inequalities for the stochastic differential equations // Stat. Inference Stoch. Process. 2018. Vol. 21, No 2. P. 469–483.

[8] Pchelintsev E.A., Pchelintsev V.A., Pergamenshchikov S.M. Improved robust model selection methods for a Lévy nonparametric regression in continuous time // Journal of Nonparametric Statistics. 2019. Vol. 31, No 3. P. 612–628.

[9] Pinsker M.S. Optimal filtration of square integrable signals in Gaussian white noise // Problems of Transimission information. 1981. Vol. 17. P. 120–133.

Повзун М. А., Пчелинцев Е. А., Пергаменщиков С. М. (Томский государственный университет, Томск, Руанский Университет, Руан, 2019) **Эффективное непараметрическое оценивание квадратично интегрируемых функций в моделях непрерывной регрессии.**

Аннотация. В данной работе изучено свойство асимптотической эффективности взвешенных оценок наименьших квадратов для неизвестных квадратично интегрируемых функций в регрессионных моделях с гауссовскими шумами. Применяется подход Пинскера. Установлено, что напрямую использовать данный метод невозможно. Получены некоторые условия и априорное распределение для коэффициентов Фурье, при которых доказана асимптотическая нижняя оценка для среднеквадратичного риска.

Ключевые слова: регрессионная модель, взвешенные оценки наименьших квадратов, асимптотическая эффективность, метод Пинскера.