сти. / А.Н. Важдаев // Известия Южного федерального университета. Технические науки, № 4, 2014. – 236 с. – с. 197-204.

13. Белов А.А., Баллод Б.А., Елизарова Н.Н. Теория вероятностей и математическая статистика / А.А. Белов, Б.А. Баллод, Н.Н. Елизарова. – Ростов н/Д: Феникс, 2008. – 318с.

АЛГОРИТМЫ ИДЕНТИФИКАЦИИ И ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ ВЗВЕШЕННЫХ ОЦЕНОК РЕГРЕССИИ И АПРИОРНОЙ ДОГАДКИ

Дмитриев Ю.Г., Кошкин Г.М., Луков В.Ю.
(г. Томск, Томский государственный университет)
e-mail: dmit@mail.tsu.ru, kgm@mail.tsu.ru, lukov_vadim@rambler.ru

COMBINED IDENTIFICATION AND PREDICTION ALGORITHMS

Dmitriev Yu.G., Koshkin G.M., Lukov V.Yu.
(Tomsk, Tomsk State University)

Аннотация. Рассматривается задача построения математической модели зависимости выходных переменных от входных переменных стохастического объекта с учетом априорных знаний о зависимости. Для решения этой проблемы используются как параметрические, так и непараметрические подходы. В работе предлагаются комбинированные алгоритмы идентификации и прогнозирования стохастических объектов с использованием линейной комбинации непараметрических и параметрических оценок регрессии.

Ключевые слова – непараметрическая оценка Надарая-Ватсона, параметрическая оценка, априорная догадка, регрессия, комбинированный алгоритм, идентификация, прогнозирование, бутстрэп.

Введение. Имеется обширная литература по оцениванию вероятностных характеристик с использованием дополнительной информации (априорной догадки). Комбинированные статистические адаптивные оценки с априорной догадкой и их свойства рассматривались в ряде работ, например, в [1-5]. В работе рассматривается случай, когда имеется предположение о виде оцениваемой функции, интерпретируемое как априорная догадка.

Рассмотрим стохастический объект, который описывается функцией регрессии

$$r(\vec{x}) = E\left(Y \mid \vec{X} = \vec{x}\right) = \int y \cdot p(y \mid \vec{x}) dy = \frac{\int y \cdot p(x, y) dy}{p(x)}, \tag{1}$$

где $\left(\vec{X}, Y\right) = \left(X^{\langle 1\rangle}, ..., X^{\langle p\rangle}, Y\right)$ - $(p+1)$-мерный вектор p входов и выхода, $p(\vec{x}, y)$ - их общая плотность распределения, $p(\vec{x})$ - плотность распределения входов, $p(y \mid \vec{x})$ является условной плотностью распределения.

Пусть $\left(\vec{X}_i, Y_i\right) = \left(X_i^{\langle 1\rangle}, ..., X_i^{\langle p\rangle}, Y_i\right)$, $i = 1, ..., n,$ - независимые наблюдения случайного вектора $\left(\vec{X}, Y\right)$. В качестве непараметрической оценку функции регрессии (1) возьмем оценку Надарая-Ватсона [6]

$$\hat{r}(\vec{x}) = \hat{r}(\vec{x}; X_1, ..., X_n) = \frac{\sum_{i=1}^{n} Y_i \cdot K\left(\frac{\vec{x} - \vec{X}_i}{\vec{h}_n}\right)}{\sum_{i=1}^{n} K\left(\frac{\vec{x} - \vec{X}_i}{\vec{h}_n}\right)}, \tag{2}$$

где $K\left(\frac{\vec{x}-\vec{X}_i}{\vec{h}_n}\right)=K\left(\frac{x_1-X_i^{\langle 1\rangle}}{h_n^{\langle 1\rangle}}\right)...\,K\left(\frac{x_p-X_i^{\langle p\rangle}}{h_n^{\langle p\rangle}}\right)$ - p-мерное ядро (произведение p одномерных ядер), $\vec{h}_n=\left(h_n^{\langle 1\rangle},...,h_n^{\langle p\rangle}\right)$- p-мерный вектор параметров размытости.

Обычно исследователь имеет некоторую информацию о характере зависимости выхода объекта от входов. Предположим, что он может выразить это знание в виде заданной функции $\varphi\left(\vec{x},\vec{\theta}\right)$, где $\vec{\theta}=\left(\theta^{(1)},...,\theta^{(s)}\right)$- вектор известных параметров. Этот тип информации мы назовем априорной догадкой. Рассмотрим задачу совместного использования непараметрической оценки регрессии и априорной догадки. Подход с использованием комбинаций различных оценок изучался, например, в [7-10].

Комбинированные оценки для статической модели. В случае статических моделей в качестве комбинированной оценки регрессии возьмем [7-9]

$$\hat{R}_{\lambda}\left(\vec{x}\right)=\left(1-\lambda\right)\cdot\hat{r}\left(\vec{x}\right)+\lambda\cdot\varphi\left(\vec{x},\vec{\theta}\right), \tag{3}$$

где λ есть весовой коэффициент, определяемый из минимума критерия

$$M\left\{\hat{R}_{\lambda}(\vec{x})-r(\vec{x})\right\}^2. \tag{4}$$

Таким образом, из (4) получаем оптимальное λ

$$\lambda(\vec{x})=\frac{M\left\{\left(\hat{r}\left(\vec{x}\right)-r\left(\vec{x}\right)\right)\left(\hat{r}\left(\vec{x}\right)-\varphi\left(\vec{x},\vec{\theta}\right)\right)\right\}}{M\left\{\hat{r}\left(\vec{x}\right)-\varphi\left(\vec{x},\vec{\theta}\right)\right\}^2}. \tag{5}$$

Подставляя (5) в (4) и делая преобразования, получаем:

$$M\left\{\hat{R}_{\lambda}(\vec{x})-r(\vec{x})\right\}^2=M\left\{\hat{r}\left(\vec{x}\right)-r\left(\vec{x}\right)\right\}^2-\frac{\left[M\left\{\left(\hat{r}\left(\vec{x}\right)-r\left(\vec{x}\right)\right)\cdot\left(\hat{r}\left(\vec{x}\right)-\varphi\left(\vec{x},\vec{\theta}\right)\right)\right\}\right]^2}{M\left\{\hat{r}\left(\vec{x}\right)-\varphi\left(\vec{x},\vec{\theta}\right)\right\}^2}. \tag{6}$$

Второй член в (6) показывает, насколько СКО комбинированной оценки $\hat{R}_{\lambda}\left(\vec{x}\right)$, с учетом априорной догадки, уменьшается по сравнению с $r\left(\vec{x}\right)$. Так как оптимальное λ (5) обычно неизвестно, рассмотрим её оценку, построенную методом бутстрэп

$$\hat{\lambda}_B(\vec{x})=\frac{\sum_{j=1}^{B}\left(\hat{r}(\vec{x};\vec{X}_j^*)-\hat{r}(\vec{x})\right)\left(\hat{r}(\vec{x};\vec{X}_j^*)-\varphi\left(\vec{x},\vec{\theta}\right)\right)}{\sum_{j=1}^{B}\left[\hat{r}(\vec{x};\vec{X}_j^*)-\varphi\left(\vec{x},\vec{\theta}\right)\right]^2}. \tag{7}$$

где $\left(\vec{X}_j^*,Y_j^*\right)=\left(\vec{X}_{1,j}^*,...,\vec{X}_{n,j}^*,Y_j^*\right)$, $j=1,...,B$ - сгенерированная бутстрэп-выборка.

Использование (6) в (3) приводит к адаптивной комбинированной оценке

$$\hat{R}_{\hat{\lambda}_B}\left(\vec{x}\right)=\left(1-\hat{\lambda}_B(\vec{x})\right)\hat{r}\left(\vec{x}\right)+\hat{\lambda}_B(\vec{x})\varphi\left(\vec{x},\vec{\theta}\right). \tag{8}$$

Если θ оценивается по выборке, то оценку (8) будем обозначать как $\tilde{R}_{\hat{\lambda}_B}\left(\vec{x}\right)$.

Комбинированные оценки для динамической модели. Рассмотрим динамическую модель (см. [10-15])

$$Y_t=f\left(Y_{t-1},...,Y_{t-s},X_{t-1},...,X_{t-q}\right)+\xi_t, \tag{9}$$

где Y_t - выход объекта в момент времени t, X_t является значением экзогенного фактора в момент t, f является неизвестной функцией, ξ_t является последовательностью независимых и одинаково распределенных случайных величин с неотрицательной плотностью распределения.

Предположим, что f ограничена и ее вид не меняется в изучаемом интервале времени. В качестве априорной догадки о виде f возьмем функцию $\varphi\left(\vec{x},\vec{\theta}\right)$ и рассмотрим следующую комбинированную адаптивную оценку:

$$\hat{R}\left(\vec{y}_t,\vec{x}_t\right)=\left(1-\hat{\lambda}_B(\vec{y}_t,\vec{x}_t)\right)\hat{r}\left(\vec{y}_t,\vec{x}_t\right)+\hat{\lambda}_B(\vec{y}_t,\vec{x}_t)\,\varphi\left(\vec{y}_t,\vec{x}_t;\vec{\theta}\right) \qquad (10)$$

где

$$\hat{\lambda}_B(\vec{y}_t,\vec{x}_t)=\frac{\sum_{j=1}^{B}\left(\hat{r}(\vec{y}_t,\vec{x}_t;\vec{Y}_j^*,\vec{X}_j^*)-\hat{r}(\vec{y}_t,\vec{x}_t)\right)\left(\hat{r}(\vec{y}_t,\vec{x}_t;\vec{Y}_j^*,\vec{X}_j^*)-\varphi\left(\vec{y}_t,\vec{x}_t;\vec{\theta}\right)\right)}{\sum_{j=1}^{B}\left[r(\vec{y}_t,\vec{x}_t;\vec{Y}_j^*,\vec{X}_j^*)-\varphi\left(\vec{y}_t,\vec{x}_t;\vec{\theta}\right)\right]^2}.$$

Анализ реальных данных. Анализ цен акций Газпрома за 2016 год проводится на основе (10) с s=1 и q=1:

$$Y_t=f\left(Y_{t-1},X_{t-1}\right)+\xi_t, \qquad (11)$$

где $t=1,...,n,$ Y_t цена акции в момент времени t. Возьмем параметрическую функцию в виде $\varphi(X_{t-1};\theta^{(1)},\theta^{(2)})=\theta^{(1)}+\theta^{(2)}X_{t-1},$ где для простоты положим $\theta^{(1)}=0,\ \theta^{(2)}=2,$ т.е.

$$\varphi(X_{t-1};0,2)=2X_{t-1}.$$

Модификацию оценки (2) интерполяционного прогноза для Y_t зададим в виде

$$\hat{Y}_t=\hat{r}\left(Y_{t-1},X_{t-1}\right)=\frac{\sum_{j\ge 2,\,j\ne t}Y_tK\left(\frac{Y_{t-1}-Y_{j-1}}{h_n^{\langle y\rangle}}\right)K\left(\frac{X_{t-1}-X_{j-1}}{h_n^{\langle x\rangle}}\right)}{\sum_{j\ge 2,\,j\ne t}K\left(\frac{Y_{t-1}-Y_{j-1}}{h_n^{\langle y\rangle}}\right)K\left(\frac{X_{t-1}-X_{j-1}}{h_n^{\langle x\rangle}}\right)}. \qquad (12)$$

Тогда комбинированная оценка, для которой $\varphi(X_{t-1};0,2)=2X_{t-1},$ задается формулой

$$\bar{Y}_t=\hat{R}_{\hat{\lambda}_B}\left(Y_{t-1},X_{t-1}\right)=\left(1-\hat{\lambda}_B(Y_{t-1},X_{t-1})\right)\hat{r}\left(Y_{t-1},X_{t-1}\right)+2\hat{\lambda}_B(Y_{t-1},X_{t-1})\,X_{t-1}, \qquad (13)$$

Аналогичным образом строятся прогнозы $\hat{Y}_{n+3},...,\hat{Y}_{n+L}$ и $\bar{Y}_{n+3},...,\bar{Y}_{n+L}$. Качество идентификации и прогнозирования будем характеризовать средними относительными ошибками $\delta_{real}\left(\hat{r}\right)$ и $\eta_{real}\left(\hat{r}\right)$:

$$\delta_{real}\left(\hat{r}\right)=\frac{1}{n-1}\sum_{i=2}^{n}\frac{\left|Y_i-\hat{r}(Y_{i-1},X_{i-1})\right|}{Y_i}\cdot 100\%,\quad \eta_{real}\left(\hat{r}\right)=\frac{1}{L}\sum_{i=n+1}^{n+L}\frac{\left|Y_i-\hat{r}(Y_{i-1},X_{i-1})\right|}{Y_i}\cdot 100\%.$$

В Таблице I приведены средние относительные ошибки идентификации и прогнозирования цен акций для разных объемов наблюдений.

Таблица I. Средние относительные ошибки идентификации и прогнозирования

Средние ошибки	Объемы наблюдений		
	10	50	100
$\delta_{real}\left(\hat{r}\right)$	1,84	1,40	1,36
$\delta_{real}\left(\hat{R}\right)$	1,31	1,23	1,23
$\delta_{real}\left(\tilde{R}\right)$	1,40	1,25	1,24
$\eta_{real}\left(\hat{r}\right)$	1,94	1,27	1,69

Средние ошибки	Объемы наблюдений		
	10	50	100
$\eta_{real}\left(\hat{R}\right)$	1,70	1,17	1,65
$\eta_{real}\left(\tilde{R}\right)$	1,75	1,24	1,66

Из полученных результатов следует, что на практике предпочтительнее использовать комбинированную оценку по сравнению с непараметрической оценкой, особенно в случае небольших объемов выборок.

Заключение. Результаты численного моделирования при конечных объемах наблюдений показывают преимущество (в точности оценивания) адаптивных комбинированных оценок по сравнению с оценками непараметрической регрессии. Особенно это преимущество проявляется для малых размеров выборок и большого уровня шума. Целесообразность применения предлагаемого подхода на практике иллюстрируется анализом цен акций «Газпрома» за 2016 год.

ЛИТЕРАТУРА

1. Dmitriev Yu.G., Tarasenko P.F. The Use of a Priori Information in the Statistical Processing of Experimental Data // Russian Physics Journal. – 1992. – V. 35. – P. 888-893.
2. Dmitriev Yu.G., Tarima S.S. Statistical estimation with possibly incorrect model assumptions // The Bulletin of Tomsk State University : control, computing, informatics. – 2009. – V. 3. – N. 8. – P. 87-99.
3. Dmitriev Yu., Tarassenko P., Ustinov Yu. On Estimation of Linear Functional by Utilizing a Prior Guess // A. Dudin et al. (Eds.): ITMM 2014. – 2014. – CCIS 487. – P. 82–90.
4. Dmitriev Yu., Tarassenko P. On Adaptive Estimation Using a Prior Guess // Proceedings The International Workshop, Applied Methods of Statistical Analysis. Nonparametric Approach. Novosibirsk, Russia. Novosibirsk, 14-15 September, 2015: NSTU publisher. – 2015. – P. 49-55.
5. Tarima S. Statistical estimation in the presence of possibly incorrect model assumptions // Journal of Statistical Theory and Practic. – 2017. – V. 11(3). – P. 449-467.
6. Koshkin G.M. Asymptotic properties of functions of statistics and their application to nonparametric estimation // Automat. and Remote Control. – 1990. – Vol. 51. – N. 3. – P. 345-357.
7. Dmitriev Yu.G., Koshkin G.M. On the Use of a Priori Information in Nonparametric Regression Estimation // IFAC Proceedings Series. – 1987. – V. 2. – P. 223– 228.
8. Dmitriev Yu.G., Koshkin G.M. Using Additional Information in Nonparametric Estimation of Density Functionals // Automat. and Remote Control. – 1987. – V. 48. – N. 10. – P. 1307–1316.
9. Skripin S.V. Properties of a Combined Regression Estimate for Finite Sample Sizes. // Bulletin of the Tomsk Polytechnic University. – 2008. – V. 313. – N. 5. – P. 10–14. (in Russian).
10. Dmitriev Yu.G., Koshkin G.M., Lukov V.Yu Combined Identification Algorithms // Applied Methods of Statistical Analysis. Nonparametric Methods in Cybernetics and System Analysis - AMSA’2017, Krasnoyarsk, Russia, 18-22 September, 2017: Proceedings of the International Workshop. – Novosibirsk: NSTU publisher. – 2017. – P. 19-27.
11. Kitayeva A.V., Koshkin G.M. Nonparametric Identification of the Production Functions // Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2011, WCE 2011, (6-8 July, 2011, Imperial College London, London, U.K.). – Hong Kong: Newswood Limited. – 2011. – V. 1. – P. 276-280.
12. Kitayeva A.V., Koshkin G.M. Nonparametric semirecursive identification in a wide sense of strong mixing processes // Problems Inform. Trans. – 2010. – V. 46. – N. 1. – P. 22-37.

13. Kitayeva A.V., Koshkin G.M. Semi-recursive nonparametric identification in the general sense of a nonlinear heteroscedastic autoregression. // Automat. and Remote Control. – 2010. – V. 71. – N. 2. – P. 257-274.
14. Kitayeva A.V., Koshkin G.M. Nonparametric Identification of Static and Dynamic Production Functions // IAENG International Journal of Applied Mathematics. – 2011. – V. 41. – Issue 3. – P. 228-234.
15. Vasiliev V.A., Koshkin G.M. Nonparametric Identification of Autoregressions // Theory Probab. – 1998. – V. 43. – N. 3. – P. 507–517.

АНАЛИЗ НАЛОГОВОЙ НАГРУЗКИ ПО СТРАНАМ МИРА

К.А. Баннова, Н.Е. Актаев, Т.А. Пенкина
(г. Томск, Томский политехнический университет,
г. Тюмень, Тюменский государственный университет)
e-mail: bannovaka@yandex.ru

ANALYSIS OF TAX RATES BY COUNTRY

K.A. Bannova, N.E. Aktaev, T.A. Penkina
(Tomsk, Tomsk Polytechnik University, Tyumen, Tyumen State University)

Annotation: Taxes and fees are the main component of budget revenues at various levels. The set of rules for the calculation and payment of taxes and fees form the tax system. An effective tax system determines the development of the economy of the state, as well as citizens residing on its territory. In this article we present an analysis of the tax burden on the OECD, BRICS countries and the Eurasian Economic Union.

Key words: tax rates, taxes, income tax, budget

Введение. Налоги и сборы являются основной составляющей доходов бюджетов различных уровней. Совокупность правил по начислению и уплате налогов и сборов образуют налоговую систему. Эффективная налоговая система определяет развитие экономики государства, а также проживающих на ее территории граждан. В данной статье приведем анализ налоговой нагрузки на страны ОЭСР, БРИКС и Евразийского экономического союза.

Основная часть. Согласно анализу приведенных данных в Таблице 1, средний уровень налоговой нагрузки на экономику в странах-членах ОЭСР в 2015 году составил 34,3% к ВВП, что на 0,4 процентных пункта выше значения в России - 33,9% к ВВП. При этом уровень налоговой нагрузки в России без учета нефтегазовых доходов в 2015 году составил - 23,1% к ВВП, что на 11,2 п. п. ниже среднего значения по ОЭСР.

Таблица 1 -Налоговая нагрузка на экономику в странах ОЭСР (% ВВП).

Страна	2008	2009	2010	2011	2012	2013	2014	2015
Австралия	27,06	25,82	25,58	26,29	27,29	27,4	27,6	27,8
Австрия	42,70	42,45	40,90	41,03	41,67	42,52	42,8	43,5
Бельгия	44,16	43,10	42,37	42,90	43,95	44,64	45,0	44.8
Великобритания	33,0	31,5	32,5	33,4	32,7	32,5	32,1	32,5
Венгрия	39,7	39,2	37,56	36,5	38,6	38,2	38,2	39.4
Германия	36,97	37,37	35,03	35,70	36,45	36,4	36,6	36,9