

ИССЛЕДОВАНИЕ ПОСЛЕДОВАТЕЛЬНОГО ПОДХОДА К ОЦЕНИВАНИЮ ПАРАМЕТРОВ АВТОРЕГРЕССИОННОЙ МОДЕЛИ

Емельянова Татьяна Вениаминовна

кандидат физико-математических наук, Томский государственный университет – ТГУ, РФ, г. Томск

Иващенко Анна Олеговна

магистрант, Томский государственный университет – ТГУ, РФ, г. Томск

Researching of sequential estimation of the autoregression model's parameters

Emelyanova Tatyana

Candidate of Physical and Mathematical Sciences **Tomsk State University – TSU, Russia, Tomsk**

Ivashchenko Anna

Master, Tomsk State University – TSU, Russia, Tomsk

Аннотация. В задачах обработки временных рядов, прогнозировании, идентификации, экономических исследованиях широко применяются авторегрессионные модели, параметры которых, как правило, неизвестны. Рассматривается задача оценивания параметров модели устойчивой авторегрессии первого порядка, а также исследование качества полученных оценок.

Abstract. The problems of processing time series, prediction, identification, economical researches are widely used autoregression models. The parameters of that models are usually unknown. The article discusses the problem of estimating and researching of quality the estimators of autoregression model's parameters.

Ключевые слова: модель авторегрессии; последовательный подход к оцениванию параметров; момент остановки.

Keywords: autoregression model; sequential estimation of the parameter; stopping rule.

Рассматривается авторегрессионный процесс первого порядка,

описанный разностным уравнением

$$X_i = \beta X_{i-1} + \varepsilon_i, i = 1, 2, \dots, \quad (1)$$

где $X_0 = 0$, ε_i – независимые одинаково распределенные случайные величины с нулевым математическим ожиданием и конечной дисперсией, $|\beta| < 1$. Чаще всего для оценивания параметра модели авторегрессии первого порядка AR(1) применяют метод наименьших квадратов (МНК) [1]. Можно рассмотреть последовательную процедуру оценивания по методу наименьших квадратов, в которой число наблюдений не фиксируется заранее. Процесс наблюдений останавливается в некоторый случайный момент времени [3]. Рассмотрим последовательную оценку, представляющую собой оценку по методу наименьших квадратов, вычисленную в момент остановки

$$\tilde{\beta}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2} \quad (2)$$

При использовании байесовского подхода к оцениванию параметров вводится некоторая функция потерь $L_n(\tilde{\beta}_n, \beta)$, которая является мерой расхождения между истинным значением параметра и его оценкой [4]. В свою очередь риск $\tilde{R}_n$ определяется как математическое ожидание функции потерь.

Проведя некоторые математические расчеты, можно получить, $\tilde{R}_n$ минимизируется объемом выборки $n_0 \approx A \frac{1}{\sigma}$.

Ситуация осложняется в том случае, когда неизвестны значения дисперсии и параметра [1]. Для такой ситуации предлагается последовательный подход к оцениванию [5]. При

этом пользуются правилом остановки, построенным по наблюдаемому процессу

$$t_A = \inf \left\{ n \geq m_A : n \geq A^2 \hat{\sigma}_n \right\} \quad (3)$$

Здесь $\hat{\sigma}_n$ – выборочная оценка дисперсии и m_A заранее заданный объем выборки. Таким образом, риск, с учетом заданного правила остановки

$$R_{t_A} = EL_{t_A}(\tilde{\beta}_{t_A} \cdot \beta)$$

Оценка $\bar{X}_{t_A}$ является асимптотически эффективной, следовательно, $\frac{R_{t_A}}{R_n} \rightarrow 1$ при $t_A \rightarrow 0$. Более того, правило остановки также является асимптотически эффективным, то есть $\frac{E(t_A)}{n_0} \rightarrow 1$ [3].

При исследовании полученных оценок параметров авторегрессионной модели было сделано предположение о том, что разность рисков $R_{t_A} - \tilde{R}_{n_0}$ является бесконечно малой величиной при $A \rightarrow \infty$. Для подтверждения данного предположения, проведено имитационное моделирование, результатами которого продемонстрировано, что разность рисков при использовании последовательной процедуры оценивания может принимать отрицательные значения даже для распределений с нормальными ошибками [2].

В качестве иллюстрации, рассмотрим результаты

имитационного моделирования, проведенного для смешанного нормального распределения:

$$\xi_1 \in N(0,1), \xi_2 \in N(0,9) \varepsilon^2 \sim 0.95\xi_1 + 0.05\xi_2, \sigma^2 = 1.4$$

Рассмотрим разность рисков при последовательном и байесовском подходе к оцениванию параметра в случае смешанного нормального распределения. В таблице указаны

результаты вычислений при различных значениях $A = \frac{1}{c}$. Здесь n_0

- количество наблюдений в момент остановки t_A ; $R_{\tilde{n}_0}$ - оценка

риска при количестве наблюдений n_0 ; β^{t_A} - оценка параметра

модели авторегрессии в момент остановки t_A ; $\sigma^{(t_A^2)}$ - оценка

дисперсии в момент остановки t_A .

Под разностью рисков понимается разность между риском R_{t_A} , вычисленным в момент

остановки t_A , и оценкой риска $R_{\tilde{n}_0}$ при количестве наблюдений

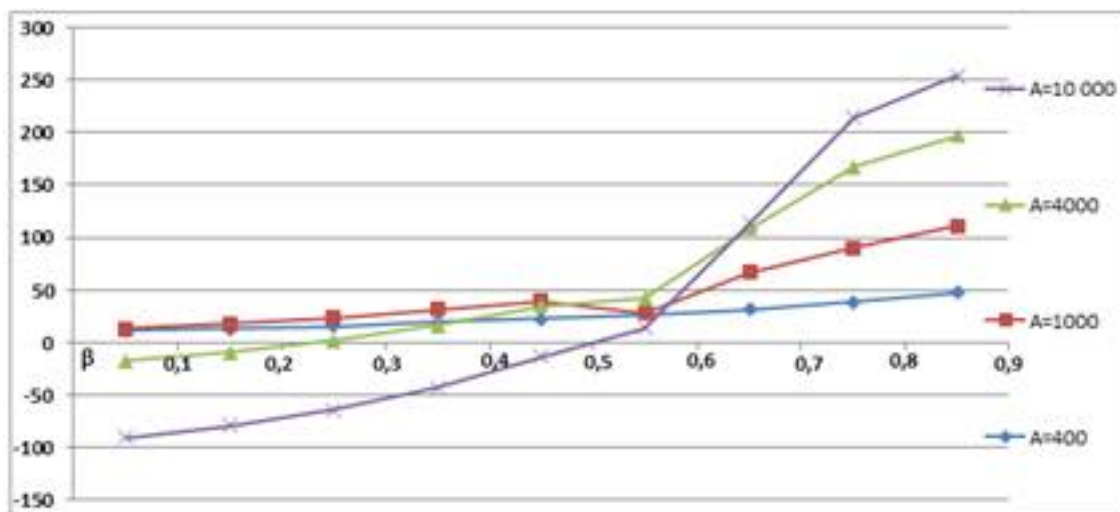
n_0 .

Таблица

Смешанное нормальное распределение:
 $\xi_1 \in N(0,1), \xi_2 \in N(0,9)\varepsilon^2 \sim 0.95\xi_1 + 0.05\xi_2, \sigma^2 = 1.4$

β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
A=400, n0 $\cong$ 23.66									
$R_{n_0}^{\sim}$	47.33	47.33	47.33	47.33	47.33	47.33	47.33	47.33	47.33
Разность рисков	11.28	13.49	15.77	19.81	22.75	26.54	31.74	38.78	47.29
$\beta_{t_A}^{\wedge}$	0.08	0.18	0.28	0.39	0.49	0.6	0.71	0.8	0.9
$\sigma_{t_A}^{\wedge 2}$	2.63	2.63	2.63	2.72	2.71	2.69	2.66	2.62	2.56
A=1000, n0 $\cong$ 37.42									
$R_{n_0}^{\sim}$	74.83	74.83	74.83	74.83	74.83	74.83	74.83	74.83	74.83
Разность рисков	1.6	4.07	7.81	11.49	16.59	24.14	35.27	51.35	63.62
$\beta_{t_A}^{\wedge}$	0.11	0.2	0.3	0.4	0.5	0.57	0.68	0.78	0.94
$\sigma_{t_A}^{\wedge 2}$	2.22	2.22	2.27	2.27	2.26	2.25	2.27	2.27	2.28
A=4000, n0 $\cong$ 74.83									
$R_{n_0}^{\sim}$	149.67	149.67	149.67	149.67	149.67	149.67	149.67	149.67	149.67
Разность рисков	-30.14	-26.59	-22.01	-15.34	-4.55	13.49	41.82	76.77	85.51
$\beta_{t_A}^{\wedge}$	0.1	0.2	0.3	0.4	0.5	0.62	0.7	0.8	0.9
$\sigma_{t_A}^{\wedge 2}$	1.77	1.77	1.77	1.77	1.77	1.77	1.77	1.76	1.76
A=10 000, n0 $\cong$ 118.32									
$R_{n_0}^{\sim}$	236.64	236.64	236.64	236.64	236.64	236.64	236.64	236.64	236.64
Разность рисков	-72.85	-69.81	-65.59	-58.99	-47.81	-28.3	4.25	47.69	56.9
$\beta_{t_A}^{\wedge}$	0.1	0.2	0.3	0.39	0.49	0.6	0.71	0.8	0.9
$\sigma_{t_A}^{\wedge 2}$	1.66	1.66	1.66	1.66	1.66	1.66	1.66	1.66	1.66

На рисунке продемонстрирован график, описывающий разность рисков в случае смешанного нормального распределения.



??????? – ????????? ?????? ? ?????? ????????????? ????????????? ?????????????

В случае смешанного нормального распределения при небольших значениях параметра и больших значениях цены наблюдений разности рисков отрицательны. При $\beta \geq 0,6$ разности рисков становятся положительными.

Таким образом, последовательный метод оценивания параметров позволяет получить оценки с гарантированным качеством в среднеквадратическом смысле за конечное время. Время оценивания определяется правилом остановки, построенным по наблюдаемому процессу.

Имитационное моделирование позволяет продемонстрировать, что использование последовательного оценивания параметров эффективно минимизирует риск функции потерь.

Список литературы:

1. Иващенко А. Идентификация параметров модели устойчивой авторегрессии/А. О. Иващенко//Молодёжь и современные информационные технологии. – 2016 г. – Т. 1 – С. 85-86.
2. Иващенко А. Последовательный подход к оцениванию параметра авторегрессии первого порядка/ А. О. Иващенко//Современные проблемы математических и естественных наук в мире. – 2015 г. – С. 24-27.
3. Sriram T. Sequential Estimation for Time Series Models / T.N.Sriram, R.Iaci // Sequential Analysis: Design Methods and Applications. – 2014. – V. 33. - P. 136-157.

4. Sriram T. Sequential Estimation of the autoregressive parameter in a first order autoregressive process / T.N.Sriram// Sequential Analysis: Design Methods and Applications. – 1988. – V. 7(1). – P. 53-74.
5. Sriram T. Sequential Estimation of the Mean of a First-Order Stationary Autoregressive Process / T.N.Sriram // The Annals of Statistics. – 1987. – V. 15. – P. 1079-1090.