

АНАЛИЗ МНК-ОЦЕНОК ДЛЯ ИДЕНТИФИКАЦИИ РАЗНОСТНЫХ УРАВНЕНИЙ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

М. Г. Матвеев, Е. А. Сирота

Воронежский государственный университет

Поступила в редакцию 10.09.2018 г.

Аннотация. Целью предлагаемой работы является исследование условий, при которых можно ожидать получение приемлемых МНК-оценок параметров разностного уравнения по результатам наблюдений нестационарных временных рядов его переменных. В частности, решается задача идентификации термодинамических процессов. В работе проводится сравнительный анализ смещения МНК-оценок параметров исследуемых уравнений авторегрессии и стандартных ошибок отклонения. Делаются выводы о возможности применения прямых МНК-оценок без существенной потери их адекватности.

Ключевые слова: параметрическая идентификация, МНК-оценки параметров, авторегрессия, смещение, разностное уравнение, стандартные ошибки отклонения.

Annotation. The aim of the present paper is to investigate the conditions under which it is possible to expect the derivation of acceptable LSM estimates of the parameters of the difference equation from the results of observations of non-stationary time series of its variables. In particular, the problem of identifying thermodynamic processes is being solved. In the paper, a comparative analysis of the displacement of LSM estimators of the parameters of the investigated autoregression equations and standard deviation errors is carried out. Conclusions are made about the possibility of applying direct LSM estimates without a significant loss of their adequacy.

Keywords: Parametric identification, LSM estimations of parameters, autoregression, displacement, difference equation, standard deviation errors.

ВВЕДЕНИЕ

Смещенность МНК-оценок может привести к ряду проблем при параметрической идентификации авторегрессионной модели нестационарных временных рядов [1, 2]. Известно, что процессы «остационарирования» могут обеспечить получение несмещенных оценок. Если временной ряд относится к классу TS-рядов [2], то возможности параметрической идентификации ограничиваются методами «остационарирования» ряда: либо последовательным применением оператора конечных разностей, либо методами коинтегрирования [2]. Однако процессы «остационарирования» достаточно сложны, громоздки и требуют подтверждения ряда статистических гипотез и, вообще говоря, не обеспечивают эффективности оценок. Поэтому часто

в практических задачах пытаются использовать простые и эффективные МНК-оценки. Если смещенность таких оценок мала или удастся уменьшить смещение до приемлемых значений [3], то смещением можно пренебречь, например, при условии, когда его величина соизмерима с величиной стандартной ошибки статистически значимой оценки.

Целью предлагаемой работы является исследование условий, при которых можно ожидать получения приемлемых МНК-оценок параметров разностного уравнения по результатам наблюдений нестационарных временных рядов его переменных.

АНАЛИЗ РАЗНОСТНОГО УРАВНЕНИЯ И АВТОРЕГРЕССИИ

Пусть модель динамического распределенного процесса, описывается системой приведенных разностных схем [4]

$$y_i^{k+1} = (y^k)^T a; \quad \forall i \quad (1)$$

с заданными начальными и краевыми условиями

$$y_i^0 = c, \quad y_{i-1}^k = b_1, \quad y_{i+1}^k = b_2, \quad \forall k; \quad (2)$$

где i – дискретные значения пространственной координаты, k – дискретное время; $y^k = (y_{i-1}^k; y_i^k; y_{i+1}^k)^T$; $y_i = (y_i^0; y_i^1; \dots; y_i^k)^T$ – векторы переменных процесса, c – начальное значение в нулевой момент времени; b_1, b_2 – наблюдаемые векторы значений краевых условий; $a = (a_1; a_2; a_3)^T$ – вектор параметров разностной схемы; T – знак транспонирования. В соответствии с условием консервативности $\sum_{i=1}^3 a_i = 1$.

К схеме (1) приводится, например, разностный вид уравнения конвективной диффузии.

Переменные процесса наблюдаются в каждом i -м узле с некоторой погрешностью e

$$x_i^k = y_i^k + e_i^k. \quad (3)$$

Будем считать, что случайный процесс e_i^k , $k = 0; 1; \dots$ имеет характер нормального «белого шума» во всех пространственных узлах i . Таким образом, в каждом i -м узле формируется временной ряд наблюдений x_i^k . Запишем выражение (1) с учетом наблюдаемых значений (3)

$$x_i^{k+1} = (x^k)^T a - (e^k)^T a + e_i^{k+1}, \quad (4)$$

где $(e^k)^T = (e_{i-1}^k; e_i^k; e_{i+1}^k)$.

Вычислим скалярное произведение $(e^k)^T a$:

$$(e^k)^T a = \tilde{e}_i^{k+1}. \quad (5)$$

Перепишем выражение (4) в виде

$$x_i^{k+1} = (x^k)^T a + \delta_i^{k+1}, \quad (6)$$

где $\delta_i^{k+1} = e_i^{k+1} - \tilde{e}_i^{k+1}$.

Поскольку $\sum_{i=1}^3 a_i = 1$, выражение (5) можно рассматривать как аффинную комбинацию случайных величин – компонент вектора e^k . При этом случайная величина \tilde{e}_i^{k+1} сохраняет свойства нормального «белого шума» [5].

Для перехода к уравнению авторегрессии найдем условное математическое ожидание (6):

$$\hat{x}_i^{k+1} = M(x_i^{k+1} | X) = (x^k)^T a + M(\delta_i^{k+1} | X) \quad (7)$$

или с учетом $M(e_i^{k+1} | X) = M(\tilde{e}_i^{k+1}) = 0$

$$\tilde{x}_i^{k+1} = (x^k)^T a - M(\tilde{e}_i^{k+1} | X), \quad (8)$$

где $X = X_{k \times 3}$ – матрица результатов наблюдений в узлах $i-1; i; i+1$, формируемая в моменты времени от 0 до k .

Выражение (8) отличается от классической авторегрессии на величину $M(\tilde{e}_i^{k+1} | X)$, что является следствием зависимости помехи от объясняющих регрессоров, обусловленной выражением (3). Это порождает смещение МНК-оценки параметров a .

ИССЛЕДОВАНИЕ СМЕЩЕНИЯ МНК-ОЦЕНКИ ПАРАМЕТРОВ АВТОРЕГРЕССИИ

Необходимые условия минимума критерия наименьших квадратов позволяют получить МНК-оценку параметров авторегрессии в виде [6]

$$\tilde{a} = (X^T X)^{-1} X^T x^{k+1}. \quad (9)$$

Вектор наблюдаемых значений $(x^{k+1})^T = (x_i^1; \dots; x_i^{k+1})$ перепишем с учетом (6) следующим образом

$$x^{k+1} = Xa + \delta^{k+1}, \quad (10)$$

где $\delta^{k+1} = (\delta_i^1; \dots; \delta_i^{k+1})^T$.

Тогда выражение для вычисления оценки принимает вид

$$\begin{aligned} \tilde{a} &= (X^T X)^{-1} X^T (Xa + \delta^{k+1}) = \\ &= a - (X^T X)^{-1} X^T \delta^{k+1}. \end{aligned} \quad (11)$$

Найдем математическое ожидание левой и правой части (11). Так как вектор e^{k+1} не зависит от X , получим

$$M(\tilde{a}) = a - M[(X^T X)^{-1} X^T (\tilde{e}^{k+1} | X)].$$

Тогда вектор абсолютных величин смещения оценок параметров будет определяться выражением

$$\Delta = a - M(\tilde{a}) = M[(X^T X)^{-1} X^T (\tilde{e}^{k+1} | X)] \neq 0. \quad (12)$$

Математическое ожидание $M[(X^T X)^{-1} X^T (\tilde{e}^{k+1} | X)]$ следует рассматривать как условное, т. е. как вектор со случайными компонентами. Очевидно, что в этом случае и смещение Δ является случайным вектором.

Наличием смещения можно пренебречь, в случаях, когда величина стандартной ошибки оценки параметров существенно больше ве-

личины смещения. Величина стандартной ошибки оценки параметров, при фиксированном объеме выборки, зависит от погрешности наблюдения e . Величина смещения также зависит от этой погрешности. Стандартная ошибка рассчитывается при МНК-оценки параметров, а величину смещения можно рассчитать с помощью выражения (11). Построим эти зависимости на основе модельного эксперимента.

**МОДЕЛЬНАЯ ОЦЕНКА
ЗАВИСИМОСТИ СТАНДАРТНОЙ
ОШИБКИ И ВЕЛИЧИНЫ СМЕЩЕНИЯ
ОЦЕНКИ ПАРАМЕТРОВ
ОТ ПОГРЕШНОСТИ НАБЛЮДЕНИЯ**

В качестве процесса (1, 2) было использовано аналитическое решение дифференциального уравнения [7], далее была выполнена регулярная дискретизация полученного решения и добавлена погрешность ε в виде «белого шума» с различной интенсивностью. В качестве параметров приведенной разностной схемы (1) были заданы следующие значения $a_1 = 0.2583$, $a_2 = 0.5$, $a_3 = 0.2417$.

Полученные статистические данные были использованы для получения оценок известных параметров. Для проведения экспериментов были выбраны разные уровни погрешности.

Таким образом, к полученным значениям переменных $y^k = (y_{i-1}^k; y_i^k; y_{i+1}^k)$ были добавлены случайные гауссовские переменные с нулевым математическим ожиданием и переменной дисперсией $-k\sigma^2$, где $\sigma^2 = 1$; $k \in \{0.0005, 0.0007, 0.0008, 0.001\}$. Модельный эксперимент повторялся при каждом уровне интенсивности помехи многократно (1000 раз).

Окончательно сформированные выборки использовались для получения МНК-оценок параметров моделируемого процесса.

При различной интенсивности помех были получены МНК-оценки параметров $\hat{a}_1; \hat{a}_2; \hat{a}_3$, приведенные в табл. 1.

По полученным выборкам для различных интенсивностей помех были рассчитаны оценки величины смещения по формуле:

$$\hat{\Delta} = (X^T X)^{-1} X^T \delta^{k+1}. \quad (13)$$

Результаты расчетов помещены в табл. 2.

Сравнительные зависимости изменения стандартных ошибок оценок параметров и

Таблица 1

МНК-оценки параметров при различной интенсивности помех

	a_1	Стандартная ошибка оценки	a_2	Стандартная ошибка оценки	a_3	Стандартная ошибка оценки
Исходные значения параметров	0,2583		0,5		0,2417	
$k = 0.0005$	0.2584	0.000267	0.4995	0.000802	0.2421	0.000670
$k = 0.0007$	0.26	0.000316	0.4974	0.001103	0.2426	0.00213
$k = 0.0008$	0.2594	0.00041	0.4981	0.001268	0.2424	0.00267
$k = 0.001$	0.2605	0.000554	0.4957	0.00175	0.2439	0.00381

Таблица 2

Оценки величины смещения при различных интенсивностях помех

	$\hat{\Delta}a_1$	$\hat{\Delta}a_2$	$\hat{\Delta}a_3$
$k = 0.0005$	-0.00015	-0.00016	-0.000345
$k = 0.0007$	-0.00026	-0.00088	-0.000231
$k = 0.0008$	-0.0004	-0.00141	-0.00271
$k = 0.001$	-0.00066	-0.00271	-0.00420



Рис. 1. Сравнительные зависимости изменения стандартных ошибок оценок параметров и абсолютной величины их смещения от погрешности наблюдений

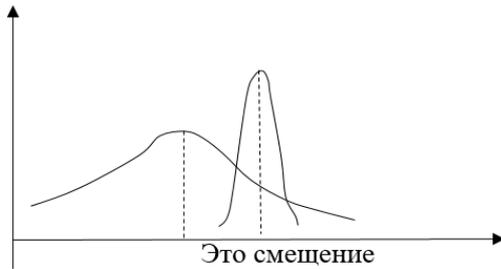


Рис. 2. Смещение МНК-оценок

абсолютной величины их смещения от погрешности наблюдений в логарифмических координатах схематично показаны на рис. 1.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные результаты показывают, что при небольших помехах величина смещения МНК-оценок параметров приведенной разностной схемы (1) существенно меньше стандартной ошибки и таким смещением можно пренебречь. В этом случае, вместо громоздких алгоритмов стационарирования временного ряда наблюдений или использования приближенного метода инструментальной переменной [8] можно применять прямые МНК-оценки без существенной потери их адекватности. С увеличением помех величина смещения остается того же порядка, что и стандартная ошибка. Это означает, что смещение не выходит, например, за границы доверительного интервала с 5 % уровнем значимости.

МНК-оценки обладают свойством эффективности при выполнении условий [8]: $\sigma_e^2 = const$; $cov(e_i; e_j) = 0$, которые выполняются в рассматриваемом случае. Тогда применение методов обеспечивающих несмещенные, но не обязательно эффективные оценки может привести к ситуации, показанной на рис. 2. В такой ситуации, смещенные эффек-

тивные МНК-оценки будут более предпочтительны, чем несмещенные оценки со значительной дисперсией.

СПИСОК ЛИТЕРАТУРЫ

1. Patterson, K. An Introduction to Applied Econometrics: A Time Series Approach / K. Patterson. – New York : Palgrave, 2000. – 832 p.
2. Гребенюк, Е. А. Методы анализа нестационарных временных рядов с неявными изменениями свойств / Е. А. Гребенюк // Автоматика и телемеханика, 2005. – № 12. – С. 3–29.
3. Заусаева, М. А. Определение параметров двумерных динамических процессов на основе разностных схем / М. А. Заусаева, В. Е. Зотеев // Вестн. Сам. гос. техн. ун-та. Сер. Физ.-мат. науки. – 2010. – Т. 20, № 1. – С. 154–161.
4. Матвеев, М. Г. Моделирование динамики атмосферных процессов на основе анализа многомерных временных рядов / М. Г. Матвеев, Е. А. Сирота, И. В. Приходько, С. А. Кузнецов // Воздушно-космические силы. Теория и практика. Электронное периодическое издание. – 2017. – №1 (1). – С. 98–110.
5. Бахтин, В. И. Введение в прикладную статистику / В. И. Бахтин. – Минск : БГУ, 2012. – 99 с.
6. Демиденко, Е. З. Линейная и нелинейная регрессия / Е. З. Демиденко. – М. : Финансы и статистика, 1981. – 302 с.
7. Матвеев, М. Г. Верификация процесса конвективной диффузии на основе анализа многомерных временных рядов / М. Г. Матвеев, Е. А. Сирота, М. Е. Семенов, А. В. Копытин // Аналитика и управление данными в областях с интенсивным использованием данных: сб. науч. тр. XIX Междунар. конф. DAMDID RCDL'2017 (Москва, 10–13 октября 2017 г). – Москва, 2017. – С. 433–438.
8. Носко, В. Н. Эконометрика. Кн. 2, части 3, 4 / В. Н. Носко. – М. : Дело, 2011. – 259 с.

Матвеев М. Г. – д-р техн. наук, профессор, заведующий кафедрой информационных технологий управления, факультет компьютерных наук, Воронежский государственный университет
E-mail: mgmatveev@yandex.ru

Matveev M. G. – doctor of technical sciences, professor department of programming and information technologies, Voronezh State University.
E-mail: mgmatveev@yandex.ru

Сирота Е. А. – канд. физ.-мат. наук, доцент кафедры цифровых технологий, факультет компьютерных наук, Воронежский государственный университет
E-mail: sirota_ea@sc.vsu.ru

Sirota E. A. – candidate of physics-math. sciences, assistant professor, the dept. of digital technologies Faculty of Computer Science, Voronezh State University.
E-mail: sirota_ea@sc.vsu.ru