

## АНАЛИЗ СВОЙСТВ МНК-ОЦЕНОК В СЛУЧАЕ УСТРАНЕНИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ В ЗАДАЧЕ ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ РАСПРЕДЕЛЕННЫХ ДИНАМИЧЕСКИХ ПРОЦЕССОВ

© 2020 М. Г. Матвеев, Е. А. Сирота 

*Воронежский государственный университет  
Университетская пл., 1, 394018 Воронеж, Российская Федерация*

**Аннотация.** Одной из наиболее сложных задач параметрической идентификации является задача оценки параметров моделей распределенных динамических процессов на основе статистических методов. Известно, что одним из факторов, влияющих на качество полученной модели, например, регрессии, является мультиколлинеарность. Если оцененную модель регрессии предполагается использовать для изучения связей, например, экономических или иного характера, то устранение мультиколлинеарных факторов является обязательным, потому что их наличие в модели может привести к смещенным коэффициентам регрессии. В рамках данной статьи был предложен метод понижения размерности с целью устранения мультиколлинеарности в задаче параметрической идентификации распределенных динамических процессов, а также проведено исследование возможности применения метода наименьших квадратов (МНК) для параметрической идентификации моделей распределенных динамических процессов в случае смещенных оценок. МНК-оценки, полученные в результате метода понижения размерности, названы альтернативными. Необходимо проверить, эффективность применения альтернативных оценок по сравнению с обычными МНК-оценками.

Проведенное исследование показало, что при низком уровне погрешностей наблюдения (1 % и ниже) применение прямых МНК-оценок для идентификации параметров распределенных динамических процессов дает удовлетворительные результаты. При этом величина смещения всегда несколько больше величины стандартного отклонения оценки параметра, что не позволяет пренебрегать смещением, особенно при высоком и среднем уровне погрешностей наблюдения. Применение альтернативных МНК-оценок позволяет уменьшить мультиколлинеарность, а, следовательно, понизить размерность задачи. В выборочной статистике и при любом уровне погрешностей наблюдения предложенный метод существенно снижает стандартную ошибку оценки параметров.

**Ключевые слова:** авторегрессионная модель, конечно-разностные уравнения, идентификация, МНК-оценки, смещенные оценки параметров модели, понижение размерности.

### ВВЕДЕНИЕ

Одной из наиболее сложных задач параметрической идентификации является задача оценки параметров моделей распределенных динамических процессов на основе статистических методов [1, 2].

Известно, что одним из факторов, влияющих на качество полученной модели, например, регрессии, является мультиколлинеарность. Если оцененную модель регрессии предполагается использовать для изучения связей, например, экономических или иного характера, то устранение мультиколлинеарных факторов является обязательным, потому что их наличие в модели может привести к смещенным коэффициентам регрессии [3].

 Сирота Екатерина Александровна  
e-mail: [sirota\\_ea@sc.vsu.ru](mailto:sirota_ea@sc.vsu.ru)



Контент доступен под лицензией Creative Commons Attribution 4.0 License.  
The content is available under Creative Commons Attribution 4.0 License.

При построении прогноза на основе модели регрессии с мультиколлинеарными факторами необходимо оценивать ситуацию по величине ошибки прогноза. Если её величина является удовлетворительной, то модель можно использовать, несмотря на мультиколлинеарность. Если же величина ошибки прогноза большая, то устранение мультиколлинеарных факторов из модели регрессии является одним из методов повышения качества прогноза. На сегодняшний день существует ряд основных способов устранения мультиколлинеарности в модели множественной регрессии. Один из наиболее простых способов устранения мультиколлинеарности состоит в получении дополнительных данных. Однако на практике в некоторых случаях реализация данного метода может быть весьма затруднительна [3, 4], также известен способ преобразования переменных, например, вместо значений всех переменных, участвующих в модели (и результативной в том числе) можно взять их логарифмы. Однако данный способ также не способен гарантировать полного устранения мультиколлинеарности факторов [5].

Если рассмотренные способы не помогли устранить мультиколлинеарность факторов, то переходят к использованию смещённых методов оценки неизвестных параметров модели регрессии, или методов исключения переменных из модели множественной регрессии [5], другими словами методов понижения размерности задачи, например, методу главных компонент или ридж регрессии. В работах [6] предлагается свой метод, основанный на исключении коррелирующих признаков, которые заранее неизвестны. Отметим, что в нашем случае, коррелирующие признаки, напротив, заранее известны.

В рамках данной статьи был предложен свой метод понижения размерности с целью устранения мультиколлинеарности, а также проведено исследование возможности применения МНК для параметрической идентификации моделей распределённых динамических процессов в случае смещённых оценок.

## 1. ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ

Если процессы адекватно описываются линейными дифференциальными уравнениями, то для получения оценок удобно перейти к разностным уравнениям. Например, для однородного уравнения конвективной диффузии с одной пространственной переменной приведенное разностное уравнение для  $i$ -го узла в  $k+1$  момент времени будет иметь вид [7]:

$$y_i^{k+1} = a_1 y_{i-1}^k + a_2 y_i^k + a_3 y_{i+1}^k, \quad (1)$$

с заданными начальными и краевыми условиями:

$$y_i^0 = c_i, \quad y_{i-1}^k = b_{i-1}^k, \quad y_{i+1}^k = b_{i+1}^k, \quad \forall k, \quad (2)$$

где  $i$  — дискретные значения пространственной координаты,  $k$  — дискретное время; вообще говоря, необходимый индекс  $i$  у параметров  $a = (a_1; a_2; a_3)$  здесь и далее опущен для упрощения записи;  $a_1 + a_2 + a_3 = 1$ , что обеспечивает консервативность схемы. Напомним, что схема называется консервативной, если она отражает на сетке те же законы сохранения, которые присутствовали в исходной дифференциальной задаче [8].

Значения переменной  $y_i^k$  измеряется в каждом узле  $i$  с погрешностью  $\xi_i^k$ , формируемой случайным процессом типа «белый шум». Измеренное значение будем обозначать  $x_i^k = y_i^k + \xi_i^k$ . Тогда выражения (1) и (2) можно записать в форме авторегрессионной зависимости, описывающей, вообще говоря, нестационарный временной ряд:

$$x_i^{k+1} = a_1 (x_{i-1}^k - \xi_{i-1}^k) + a_2 (x_i^k - \xi_i^k) + a_3 (x_{i+1}^k - \xi_{i+1}^k) + \xi_i^{k+1} = a^T \cdot x^k + \omega_i, \quad (3)$$

где  $\omega_i = \xi_i^{k+1} - a^T \cdot \xi^k$ ;  $T$  — здесь и далее знак транспонирования. Значения начальных и краевых условий также определяются результатами измерений. Как правило, временные ряды в смежных узлах аппроксимирующей сетки сильно коррелированы между собой, что вместе с нестационарностью обуславливает проблемы получения МНК-оценок  $\hat{a}$  с удовлетворительными статистическими свойствами. В частности, в этом случае оценки параметров будут смещёнными [3]. Несмещённые

ные оценки можно получить с применением методов «остационарирования» [9] и связанных с ними процедур и критериев определения типа процесса, существенно усложняющие алгоритмы нахождения оценок по сравнению с прямым применением МНК. Поэтому представляется актуальным исследование границ применимости прямого использования МНК, тем более, что в ряде наших предыдущих исследований нестационарных процессов изменения температурных полей атмосферы [10, 11, 12] МНК демонстрировал вполне приемлемые результаты.

Наличие смещения, вообще говоря, не всегда определяет отказ от применения МНК, т. к. качество оценки задается комбинацией стандартной ошибки оценки и ее смещения. В этой связи актуальным представляется исследование указанных свойств при оценке параметров авторегрессии (3) с условиями (2) при различных уровнях помех наблюдения. Нас также будет интересовать, как изменяются свойства оценок при изменении степени корреляции между регрессорами.

## 2. МЕТОДИКА ИССЛЕДОВАНИЯ

Мы будем исследовать зависимость стандартной ошибки оценки параметров,  $\delta$  и смещения,  $\Delta$  от уровня помехи наблюдения,  $\xi$  и коэффициента парной корреляции  $r_{ij}$  между  $i$ -м и  $j$ -м временными рядами в соответствующих узлах. Статистические данные для исследования были получены в виде аналитического решения однородного уравнения конвективной диффузии [7]

$$\frac{\partial y}{\partial t} + v \frac{\partial y}{\partial l} = D \frac{\partial^2 y}{\partial l^2},$$

$$y(0, l) = \phi(l),$$

$$y(t, l^{\min}) = f_1(t),$$

$$y(t, l^{\max}) = f_2(t)$$
(4)

$$y(t, l) = e^{\frac{v}{2D}\left(l - \frac{vt}{2}\right)} [e^{-Dt} \sin(l) + e^{-4Dt} \sin(2l) + e^{-9Dt} \sin(3l)].$$
(5)

Решение  $y(t, l)$  было дискретизировано в узлах регулярной сетки  $t_k, l_i$ . В каждом узле  $k$

полученному значению  $y_i^k$  добавлялась аддитивная помеха  $\xi_i^k$ , интенсивность которой менялась в ходе исследования.

Для оценки влияния мультиколлинеарности рядов на стандартную ошибку и смещение оценки использовалась следующая процедура. Поскольку временные ряды в смежных узлах сильно коррелированы, можно предположить наличие линейной зависимости уровней соответствующих временных рядов записать линейную комбинацию

$$y_i^k = \beta_1 y_{i-1}^k + \beta_2 y_{i+1}^k,$$

где

$$y_i^k = x_i^k - \xi_i^k$$

тогда

$$x_i^k = \beta_1 x_{i-1}^k + \beta_2 x_{i+1}^k + \xi^k, \tag{6}$$

где

$$\xi^k = \xi_i^k - \beta_1 \xi_{i-1}^k - \beta_2 \xi_{i+1}^k$$

Подставим выражение (6) в (3), получим:

$$x_i^k = (a_1 + a_2 \beta_1) x_{i-1}^k + (a_2 \beta_2 + a_3) x_{i+1}^k + \xi_{\Sigma}^{k+1} - \xi_{\Sigma}^k = \theta_1 x_{i-1}^k + \theta_2 x_{i+1}^k + \Delta \xi, \tag{7}$$

где  $\xi_{\Sigma}^k$  — выпуклая линейная комбинация помех  $\xi_{i-1}^k, \xi_i^k, \xi_{i+1}^k$ .

Выражение (7) позволяет построить систему линейных уравнений относительно параметров  $a_1, a_2, a_3$ :

$$\begin{aligned} a_1 + a_2 \beta_1 &= \theta_1, \\ a_2 \beta_2 + a_3 &= \theta_2, \\ a_1 + a_2 + a_3 &= 1. \end{aligned} \tag{8}$$

Посчитаем определитель системы (8)

$$\begin{vmatrix} 1 & \beta_1 & 0 \\ 0 & \beta_2 & 1 \\ 1 & 1 & 1 \end{vmatrix} : \beta_1 + \beta_2 - 1 \neq 0,$$

т. е. система является невырожденной в случае, когда  $\beta_1 + \beta_2 \neq 1$ .

В дальнейшем, оценки параметров  $a$ , полученные решением системы (8) будем называть альтернативными МНК-оценками, в отличие от прямых МНК-оценок.

МНК-оценки параметров  $\beta_1$  и  $\beta_2$  можно получить в виде

$$\hat{\beta} = (X^T X)^{-1} X^T x_k, \tag{9}$$

где  $X$  — матрица наблюдений в узлах  $i-1$  и  $i+1$ ;  $x_k$  — вектор наблюдений в узле  $i$  в моменты времени  $k$ .

Оценки параметров  $\theta_1$  и  $\theta_2$  получаются в виде

$$\hat{\theta} = (X^T X)^{-1} X^T x_{k+1}, \quad (10)$$

где  $x_{k+1}$  — вектор наблюдений в узле  $i$  в моменты времени  $k+1$ .

Оценки (9) и (10) будут смещенными. Величина смещения определяется соответствующими выражениями:

$$\begin{aligned} \Delta\beta &= (X^T X)^{-1} X^T \varepsilon \\ \Delta\theta &= (X^T X)^{-1} X^T \Delta\xi. \end{aligned} \quad (11)$$

Альтернативные МНК-оценки вектора параметров  $a$  будут отличаться от соответствующих оценок, полученных с использованием прямого МНК. Отличие, в основном, будет обусловлено разной степенью коррелированности регрессоров, поскольку коррелированность не смежных регрессоров в (6) и (7) будет ниже, чем коррелированность смежных регрессоров в (3). Прямой МНК использует три регрессора с высокими значениями парной корреляции, а альтернативный подход два регрессора, корреляция которых объективно ниже. Альтернативный подход соответствует переходу от исходной, плохо обусловленной системы уравнений к лучше обусловленной эквивалентной системе, в случае понижения размерности задачи. В дальнейшем приведенный подход для альтернативных МНК-оценок будет называть первым способом.

Можно рассмотреть и другой (второй) способ получения альтернативных МНК-оценок, который также в отличии от основного уравнения будет использовать только два регрессора, корреляция которых ниже, а сама система обусловлена лучше исходной. Используя условие консервативности, переписав уравнение (3) следующим образом

$$\begin{aligned} x_i^{k+1} &= a_1 x_{i-1}^k + (1 - a_1 - a_3) x_i^k + a_3 x_{i+1}^k = \\ &= x_i^k + a_1 (x_{i-1}^k - x_i^k) + a_3 (x_{i+1}^k - x_i^k); \\ x_i^{k+1} - x_i^k &= a_1 (x_{i-1}^k - x_i^k) + a_3 (x_{i+1}^k - x_i^k). \end{aligned}$$

Таким образом, относительно разностей

$$\begin{aligned} \Delta x_i^k &= x_i^{k+1} - x_i^k; \Delta x_{i-1}^k = x_{i-1}^k - x_i^k; \\ \Delta x_{i+1}^k &= x_{i+1}^k - x_i^k. \end{aligned}$$

получаем новое уравнение, которое содержит лишь два регрессора

$$\Delta x_i^k = a_1 \Delta x_{i-1}^k + a_3 \Delta x_{i+1}^k. \quad (12)$$

Таким образом, находим МНК — оценки параметров  $a_1$  и  $a_3$ , а параметр  $a_2$  находим из условия консервативности

$$a_2 = 1 - a_1 - a_3.$$

### 3. ОПИСАНИЕ УСЛОВИЙ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ И ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Выборочная статистика, необходимая для модельного исследования была получена с использованием аналитического решения (4) дифференциального уравнения (5). Аналитическое решение было дискретизировано по времени и пространственной координате. Было получено разностное уравнение (1) со следующими значениями параметров:  $a_1 = 0.2583$ ,  $a_2 = 0.5$ ,  $a_3 = 0.2417$ .

В соответствии с методикой исследования к значениям переменной  $y_i$  в соответствующих узлах сетки была добавлена аддитивная помеха наблюдения  $\xi_i^k$ , полученная с помощью генератора независимых случайных чисел гауссовского типа с нулевым математическим ожиданием и единичной дисперсией  $\sigma_\xi^2$ . Интенсивность помехи  $c \cdot \sigma_\xi^2$  задавалась на уровнях:  $c = 0.1$ ,  $c = 0.01$ ,  $c = 0.0025$ ,  $c = 0.001$ . При этом доля помехи составила соответственно 40 %; 4 %; 1 % и 0,4 % от среднеквадратической величины амплитуды полезного сигнала, что позволило разделить полученную статистику на данные: с высоким (40 %), средним (4 %) и низким (1 % и 0,4 %) уровнями помех.

Изменение мультиколлинеарности достигалось выбором метода получения решения задачи идентификации параметров: либо прямыми МНК-оценками для уравнения (3), либо альтернативными МНК-оценками в виде решения системы уравнений (8), а также уравнения (12). Корреляционная матрица регрессоров с усредненными по узлам сетки элементами, имеет следующий вид:

	$x_{i-1}$	$x_i$	$x_{i+1}$
$x_{i-1}$	1	0,98	0,76
$x_i$	0,98	1	0,97
$x_{i+1}$	0,76	0,97	1

Нетрудно заметить, что уровень корреляции в смежных узлах отличается от соответствующего значения в несмежных узлах примерно на 20 %.

Для получения надежных результатов оценок смещения и стандартной ошибки, допускающих сравнения числовых величин, эксперименты во всех режимах повторялись 1000 раз и их результаты усреднялись.

Прямые МНК-оценки параметров  $s$ , а также альтернативные МНК-оценки приведены в табл. 1 для различных уровней помехи наблюдения. Для анализа эффективности рассматриваемых методов в табл. 2 и табл. 3 приведены значения смещений для каждого метода, а также стандартных ошибок оцениваемых параметров.

Анализ табл. 1 и 2 показывает, что удовлетворительные оценки параметров во всех трех подходах достигаются только при низком уровне помех. Но фундаментальное свойство консервативности сохраняется с высокой точностью на всех уровнях помех даже в прямых МНК-оценках, где соотношение  $a_1 + a_2 + a_3 = 1$  явно не заложено в решение задачи идентификации.

Возможность более подробного анализа полученных результатов представляют табл. 3, 4.

Анализ таблиц демонстрирует явное преимущество альтернативных подходов, как в первом, так и во втором случае, особенно для низкого уровня помехи, из самих данных таблиц не совсем ясно какой из способов первый или второй является более предпочтительным. Однако, второй подход представляет явные преимущества в самой реализации вычислительного алгоритма, так как более прост в исполнении.

Анализ полученных результатов удобно проводить, основываясь на двух источниках смещения и стандартной ошибки: наличия ошибки наблюдения и мультиколлинеарности.

Смещения прямых МНК-оценок при наличии коррелированности регрессоров с погрешностями наблюдения определяется следующим образом

$$\begin{aligned} \hat{a} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x} = \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}a + \omega) = \\ &= a + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \omega \end{aligned} \quad (13)$$

$$E(\hat{a}) = a + E\{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \omega\}, \quad (14)$$

где  $E\{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \omega\}$  — величина смещения прямой МНК-оценки.

Приблизительное выражение смещения альтернативных МНК-оценок можно получить с учетом (10) следующим образом:

$$\begin{aligned} \hat{a} &= \hat{B}^{-1} \hat{\theta} = \hat{B}^{-1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}\theta + \Delta\xi) \approx \\ &\approx a + \hat{B}^{-1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Delta\xi \end{aligned} \quad (15)$$

$$E(\hat{a}) = a + E\{\hat{B}^{-1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Delta\xi\}, \quad (16)$$

где  $E\{\hat{B}^{-1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Delta\xi\}$  — приближительная величина смещения альтернативной МНК-оценки;  $\hat{B}$  — матрица системы (8), невырожденность которой обуславливается смещенными оценками  $\beta$ .

Выражения (14) и (16) показывают, что величина смещения зависит от двух факторов: помеха наблюдения,  $\xi_i^k$  и погрешности обращения матрицы  $(\mathbf{X}^T \mathbf{X})$ , связанные с ее плохой обусловленностью. Числа обусловленности этой матрицы для различных помех наблюдения и приведены в табл. 4.

Данные табл. 5 показывают, что существенное изменение числа обусловленности матрицы  $\mathbf{X}^T \mathbf{X}$  за счет применения альтернативного подхода (расчеты производились для первого способа, в случае второго результаты аналогичны) достигается только при низком уровне помех. Как следствие, при высоком и среднем уровне помех изменение числа обусловленности практически не влияет на величину смещения (табл. 2, 3). Отсюда можно сделать вывод о том, что при высоком и среднем уровне помех наблюдения именно они оказывают доминирующее влияние на величину смещения оценок  $a$  при обоих подходах. При малых помехах доминирующую роль в формировании смещения начинает играть погрешность обращения матрицы  $\mathbf{X}^T \mathbf{X}$ . В этом случае применение альтернативных МНК-оценок существенно уменьшает смещение (табл. 2, 3).

Таблица 1. Сравнение МНК-оценок при различной интенсивности помех (первый способ)  
 [Table 1. Comparison of the OLS estimates at various noise intensity (first method)]

c	МНК-оценки параметров при различной интенсивности помех							
	Прямые МНК-оценки Коэфф. парн. корр. регрессоров $\geq 0,97$				Альтернативные МНК-оценки (I способ) Коэфф. парн. корр. регрессоров = 0,76			
	$a_1$	$a_2$	$a_3$	$\Sigma a$	$a_1$	$a_2$	$a_3$	$\Sigma a$
0,1	0,3958	0,3293	0,2661	0,9912	0,3904	0,3310	0,2786	1
0,01	0,3520	0,3375	0,3100	0,9995	0,3483	0,3399	0,3118	1
0,0025	0,2732	0,4730	0,2540	1,0002	0,2657	0,4889	0,2454	1
0,001	0,2613	0,4951	0,2434	0,9998	0,2594	0,4992	0,2415	1
Истинные значения	0,2583	0,5000	0,2417	1	0,2583	0,5000	0,2417	1

Таблица 2. Сравнение МНК-оценок при различной интенсивности помех (второй способ)  
 [Table 2. Comparison of the OLS estimates at various noise intensity (second method)]

c	МНК-оценки параметров при различной интенсивности помех							
	Прямые МНК-оценки Коэфф. парн. корр. регрессоров $\geq 0,97$				Альтернативные МНК-оценки (II способ) Коэфф. парн. корр. регрессоров = 0,76			
	$a_1$	$a_2$	$a_3$	$\Sigma a$	$a_1$	$a_2$	$a_3$	$\Sigma a$
0,1	0,3958	0,3293	0,2661	0,9912	0,390376	0,33097	0,278627	1
0,01	0,3520	0,3375	0,3100	0,9995	0,348316	0,339912	0,311772	1
0,0025	0,2732	0,4730	0,2540	1,0002	0,2656	0,48919	0,24521	1
0,001	0,2613	0,4951	0,2434	0,9998	0,260553	0,497056	0,242391	1
Истинные значения	0,2583	0,5000	0,2417	1	0,2583	0,5000	0,2417	1

Таблица 3. Оценки средней величины смещения рассматриваемых методов при различной интенсивности помех (прямой МНК и первый способ)  
 [Table 3. The estimates of the mean displacement of the considered methods with various levels of noise intensity (direct OLS method and the first method)]

c	Оценки средней величины смещения					
	Смещение прямых МНК-оценок, $r_{xx} \geq 0,97$			Смещение альтернативных МНК-оценок, $r_{xx} = 0,76$		
	$\Delta a_1$	$\Delta a_2$	$\Delta a_3$	$\Delta a_1$	$\Delta a_2$	$\Delta a_3$
0,1	0,1375	0,1707	0,0244	0,1321	0,1690	0,0369
0,01	0,0937	0,1625	0,0683	0,0900	0,1601	0,0701
0,0025	-0,0140	0,0270	0,0125	0,0074	0,0111	-0,0037
0,001	-0,0030	0,0049	-0,0017	-0,0010	0,0008	0,0002

Таблица 4. Оценки средней величины смещения рассматриваемых методов при различной интенсивности помех (прямой МНК и второй способ)

[Table 4. The estimates of the mean displacement of the considered methods with various levels of noise intensity (direct OLS method and the second method)]

c	Оценки средней величины смещения					
	Смещение прямых МНК-оценок, $r_{xx} \geq 0,97$			Смещение альтернативных МНК-оценок, $r_{xx} = 0,76$		
	$\Delta a_1$	$\Delta a_2$	$\Delta a_3$	$\Delta a_1$	$\Delta a_2$	$\Delta a_3$
0,1	0,1375	0,1707	0,0244	0,13207	0,16903	0,03692
0,01	0,0937	0,1625	0,0683	0,09002	0,160088	0,07007
0,0025	-0,0140	0,0270	0,0125	0,00737	0,011	-0,00364
0,001	-0,0030	0,0049	-0,0017	-0,00225	0,0029	0,00069

Таблица 5. Числа обусловленности матрицы  $X^T X$

[Table 5. Condition numbers of the matrix  $X^T X$ ]

c	Для прямых МНК-оценок	Для альтернативных МНК-оценок (первый способ)
0,1	40	16
0,01	100	60
0,0025	$4,1 \cdot 10^3$	$2,1 \cdot 10^2$
0,001	$8,6 \cdot 10^5$	$3,2 \cdot 10^3$

Стандартная ошибка оценки параметра  $a_i$  определяется следующим образом

$$\delta_{ai} = \sigma_{\xi} \cdot \sqrt{x_{ii}}, \quad (17)$$

где  $x_{ii}$  — соответствующий элемент главной диагонали матрицы  $(X^T X)^{-1}$ . То есть стандартная ошибка зависит от тех же факторов, что и смещение — среднего квадратического отклонения помехи наблюдения,  $\sigma_{\xi}$  и погрешности обращения матрицы  $(X^T X)$ , анализ табл. 3 показывает явное преимущество альтернативных МНК-оценок на всех уровнях погрешностей наблюдения.

## ЗАКЛЮЧЕНИЕ

Проведенное исследование показало, что при низком уровне погрешностей наблюдения (1 % и ниже) применение прямых МНК-оценок для идентификации параметров распределенных динамических процессов дает удовлетворительные результаты. При этом величина

смещения всегда несколько больше величины стандартного отклонения оценки параметра, что не позволяет пренебрегать смещением, особенно при высоком и среднем уровне погрешностей наблюдения. Применение альтернативных МНК-оценок позволяет уменьшить мультиколлинеарность, а, следовательно, понизить размерность задачи. В выборочной статистике и при любом уровне погрешностей наблюдения существенно снижает стандартную ошибку оценки параметров. На величину смещения альтернативные МНК-оценки оказывают положительное влияние лишь при низком уровне погрешностей наблюдения.

Прямые МНК-оценки параметров можно улучшить с помощью альтернативных МНК-оценок, но только при низком уровне погрешностей наблюдения.

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

## СПИСОК ЛИТЕРАТУРЫ

1. Guo, L. Z. Identification of partial differential equation models for a class of multiscale spatio-temporal dynamical systems / L. Z. Guo, S. A. Billings, D. Coca // International Journal of Control. – 2010. – No 83:1. – С. 40–48. DOI: 10.1080/00207170903085597.

2. Xiaolei, X. Parameter Estimation of Partial Differential Equation Models / Xiaolei Xun, Jiguo Cao, Bani Mallick, Raymond J. Carroll, Arnab Maity // Journal of the American Statistical Association. – 2013. – No 108:503. – P. 1009–1020. DOI: 10.1080/01621459.2013.794730
3. Носко, В. П. Эконометрика. Введение в регрессионный анализ временных рядов / В. П. Носко. – М. : НФПК, 2002. – 273 с.
4. Яковлева, А. В. Эконометрика. Конспект лекций / А. В. Яковлева. – М. : Эксмо, 2008. – 224 с.
5. Yan, X. Linear Regression Analysis: Theory and Computing / X. Yan, X. G. Su. // Singapore: World Scientific Publishing Co. Pte. Ltd. – 2009. – 348 с. DOI: 10.1142/6986
6. Стрижов, В. В. Методы индуктивно-го порождения регрессионных моделей / В. В. Стрижов. – М. : ВЦ РАН, 2008. – 55 с.
7. Matveev, M. G. Parameters identification of a distributed dynamical model using combined approach / M. G. Matveev, A. V. Kopytin, E. A. Sirota // Journal of Physics: Conference Series. – 2018. – No 1096. – 012068. DOI:10.1088/1742-6596/1096/1/012068
8. Самарский, А. А. Теория разностных схем / А. А. Самарский – М. : Наука, 1978. – 376 с.
9. Гребенюк, Е. А. Методы анализа нестационарных временных рядов с неявными изменениями свойств / Е. А. Гребенюк // Автоматика и телемеханика. – 2005. – № 66. – С. 1–28. DOI: 10.1007/s10513-005-0221-z
10. Матвеев, М. Г. Комбинированные модели нестационарных временных рядов с изменяющимися состояниями / М. Г. Матвеев, Е. А. Сирота // Вестник Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии. – 2016. – № 3. – С. 50–61.
11. Matveev, M. G. Modeling of Nonstationary Distributed Processes on the Basis of Multidimensional Time Series / M. G. Matveev, A. V. Kopytin, E. A. Sirota, E. A. Kopytina // Procedia Engineering. – 2017. – No 201. – P. 1–862. DOI: 10.1016/j.proeng.2017.09.643
12. Копытин, А. В. Применение расширенного фильтра Калмана для идентификации параметров распределенной динамической системы / А. В. Копытин, Е. А. Копытина, М. Г. Матвеев // Вестник Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии. – 2018. – № 3. – С. 44–50.

**Матвеев Михаил Григорьевич** – д-р тех. наук, проф., заведующий кафедрой информационных технологий управления факультета компьютерных наук Воронежского государственного университета.

E-mail: mgmatveev@yandex.ru

ORCID iD: <https://orcid.org/0000-0002-6528-6420>

**Сирота Екатерина Александровна** – канд. физ.-мат. наук, доцент кафедры цифровых технологий факультета компьютерных наук Воронежского государственного университета.

E-mail: sirota\_ea@sc.vsu.ru

ORCID iD: <https://orcid.org/0000-0002-0338-7461>

## AN ANALYSIS OF THE OLS ESTIMATES FOR THE PROBLEM OF PARAMETRIC IDENTIFICATION OF DISTRIBUTED DYNAMIC PROCESSES IN THE SITUATION OF ELIMINATION OF MULTICOLLINEARITY

© 2020 M. G. Matveev, E. A. Sirota 

Voronezh State University  
1, Universitetskaya square, 394018 Voronezh, Russian Federation

**Annotation.** One of the most difficult problems of parametric identification is the estimation of the parameters of the models of distributed dynamic processes using statistical methods. It is known that one of the factors affecting the quality of the resulting model (for example, of regression), is multicollinearity. If the estimated regression model is intended to be used to study relationships (for example, economic relationships, etc.), then the elimination of multicollinear factors is mandatory, because their presence in the model can result in biased regression coefficients. The article suggests a method for reducing the dimensionality in order to eliminate multicollinearity in the problem of parametric identification of distributed dynamic processes. It also considers the possibility of using the Ordinary Least Squares (OLS) method for parametric identification of models of distributed dynamic processes in the case of biased estimates. We gave the name «alternative estimates» to the OLS estimates obtained as a result of the dimensionality reduction. It was then necessary to assess the effectiveness of the alternative estimates as compared to the normal OLS estimates.

The study demonstrated that with a low level of observational errors (1 % or less), the use of direct OLS estimates for the identification of the parameters of distributed dynamic processes yields satisfactory results. At the same time, the displacement is always slightly greater than the standard deviation of the parameter estimate, which does not allow us to neglect this displacement, especially with high and average levels of observational errors. The use of alternative OLS estimates reduces the multicollinearity and, consequently, reduces the dimensionality of the problem. In the sample statistics and at any level of observational errors, the proposed method significantly reduces the standard error of parameter estimation.

**Keywords:** autoregressive model, finite difference equations, identification, OLS estimates, biased estimates of model parameters, dimensionality reduction.

### CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

### REFERENCES

1. Guo L. Z., Billings S. A., Coca D. Identification of partial differential equation models for a class of multiscale spatio-temporal dynamical systems // *International Journal of Control*. 2010. V. 83:1. P. 40–48. DOI: 10.1080/00207170903085597.

2. Xiaolei Xun, Jiguo Cao, Bani Mallick, Raymond J. Carroll, Arnab Maity. Parameter Estimation of Partial Differential Equation Models // *Journal of the American Statistical Association*. 2013. V. 108:503. P. 1009–1020. DOI: 10.1080/01621459.2013.794730

3. Nosko V. P. *Econometrics. Introduction to Regression Analysis of the Time Series*. Moscow: NFPK, 2002. – 273 p.

4. Yakovleva A. V. *Econometrics. Lecture notes*. Moscow: Exmo, 2008. – 224 p.

5. Yan X, Su X. G. *Linear Regression Analysis: Theory and Computing* // Singapore: World Scientific Publishing Co. Pte. Ltd. 2009. – 348 p. DOI: 10.1142/6986.

 Sirota Ekaterina. A.  
e-mail: [sirota\\_ea@sc.vsu.ru](mailto:sirota_ea@sc.vsu.ru)

6. *Strizhov V. V.* Methods of inductive generation of regression models. Moscow : VZ RAN, 2008. – 55 p.
7. *Matveev M. G., Kopytin A. V., Sirota E. A.* Parameters identification of a distributed dynamical model using combined approach // *Journal of Physics: Conference Series*. 2018. V. 1096. 012068. DOI:10.1088/1742-6596/1096/1/012068
8. *Samarskii A. A.* The Theory of Difference Schemes. Moscow : Nauka, 1978. 376 p.
9. *Grebenyuk E. A.* Methods of analyzing the nonstationary time series with implicit changes in their properties // *Automation and Remote Control*. 2005. V. 66. P. 1–28. DOI: 10.1007/s10513-005-0221-z
10. *Matveev M. G., Sirota E. A.* Combined models of non-stationary time series with varying condition // *Proceedings of Voronezh State University, Series: Systems Analysis and Information Technologies*. 2016. V. 3. P. 50–61.
11. *Matveev M. G., Kopytin A. V., Sirota E. A., Kopytina E. A.* Modeling of Nonstationary Distributed Processes on the Basis of Multidimensional Time Series // *Procedia Engineering*. 2017. V. 201. P. 1–862. DOI: 10.1016/j.proeng.2017.09.643
12. *Kopytin A. V., Kopytina E. A., Matveev M. G.* Application of the Extended Kalman Filter to Parameters Identification of a Distributed Dynamic System // *Proceedings of Voronezh State University, Series: Systems Analysis and Information Technologies*. 2018. V. 3. P. 44–50.

**Matveev Mikhail G.** – DSc in Technical Sciences, Professor, Head of the Department of Information Technologies in Management, Faculty of Computer Sciences, Voronezh State University.

E-mail: [mgmatveev@yandex.ru](mailto:mgmatveev@yandex.ru)

ORCID iD: <https://orcid.org/0000-0002-6528-6420>

**Sirota Ekaterina A.** – PhD in Physics and Mathematics, Associate Professor, Department of Digital Technologies, Faculty of Computer Sciences, Voronezh State University.

E-mail: [sirota\\_ea@sc.vsu.ru](mailto:sirota_ea@sc.vsu.ru)

ORCID iD: <https://orcid.org/0000-0002-0338-7461>