

УДК 519.23

Сравнительное исследование точности локального и глобального оценивания параметров регрессионных моделей при использовании различных типов М-оценок*

А.А. ПОПОВ¹, А.А. ХОЛДОНОВ²

¹ 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, доктор технических наук, профессор. E-mail: a.porov@corp.nstu.ru

² 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, аспирант. E-mail: firuz_530_11_29@mail.ru

В работе рассматриваются локальный и глобальный методы оценивания параметров регрессионных моделей при построении робастных решений. Робастные решения строятся на основе М-оценок по итерационному методу наименьших квадратов. Рассматриваются несколько типов М-оценок на основе функций потерь Хьюбера, Эндрюса и Тьюки. Цель данного исследования состояла в сравнении точности получаемых решений при использовании локального и глобального оценивания параметров регрессионных моделей в условиях симметричного и несимметричного засорения. При использовании локального метода наименьших квадратов параметры отдельных линейных моделей, входящих в систему правил, оцениваются независимо. В качестве систем правил использовалась модель Такаги–Сугено. При разбиении области определения входных факторов использовались трапециевидные функции принадлежности. Для оценивания точности получаемых решений в работе использовалась среднеквадратичная ошибка (MSE). Для проведения вычислительного эксперимента было разработано соответствующее программное обеспечение. Вычислительный эксперимент проводился на модельных данных. В качестве модели, порождающей данные, использовалась линейная зависимость от входного фактора. Рассматривается два варианта засорения выборки данных: с использованием симметричного и несимметричного засорения. Дисперсия помехи (уровень шума) определялась в процентах от мощности сигнала. Результаты проведенных исследований приведены в отдельных таблицах. Сделан вывод, что в практическом плане для построения устойчивых моделей по загрязненным выборкам можно использовать любые из рассмотренных вариантов М-оценок. Рассмотренные локальный и глобальный методы оценивания параметров показали близкие результаты. Преимущество локального метода оценивания может проявиться в случае, когда области пересечения функций принадлежности узкие.

Ключевые слова: нечеткие системы (Fuzzy System), регрессионная модель, метод наименьших квадратов (МНК), модель Такаги–Сугено, М-оценивание, функции потерь, робастные оценки, среднеквадратичная ошибка, симметричное засорение, несимметричное засорение

DOI: 10.17212/1814-1196-2017-4-67-84

* Статья получена 29 сентября 2017 г.

ВВЕДЕНИЕ

Технология построения регрессионных зависимостей в рамках концепции нечетких систем (FS) является достаточно удобным инструментом моделирования при отсутствии априорных предположений о структуре модели (составе регрессоров) [1–10]. В этом случае выбор модели происходит, как говорят, на основе самих данных. Универсальность данной методологии базируется на возможности управлять сложностью модели через выбор числа и формы нечетких партиций для входных факторов. Однако эти гибкость и универсальность создают и определенные трудности. Перечислим некоторые из них: необходимо, например, контролировать полноту покрытия области определения факторов нечеткими партициями, высока вероятность получения переусложненной модели, есть сложности моделирования при наличии выбросов. В число факторов, действующих на объект, могут входить факторы, измеренные не только в абсолютной шкале или шкале отношений, но и в номинальной шкале. В этом случае необходимо учитывать условия идентифицируемости таких моделей [6, 11].

Сложность моделирования при наличии выбросов связана с тем, что рассматриваемая методология направлена на получение описаний с локальными особенностями. Имеющиеся выбросы в данных могут провоцировать формирование такой локальной особенности, а также приводить к общему смещению решения, поскольку для его получения, как правило, используется метод наименьших квадратов (МНК). Широко известно, что в случае, когда предположение о гауссовой модели случайных ошибок не выполняется, использование МНК может привести к значительным ошибкам [12]. С целью обеспечения устойчивости оценок неизвестных параметров относительно отклонений случайных ошибок от гауссовой модели разработан ряд робастных статистических методов [12–16]. Термин «робастность» означает устойчивость или слабую чувствительность статистической процедуры к незначительным отклонениям данных от начальных предположений [12]. В последние годы в рамках концепции нечетких систем также стали подниматься вопросы построения робастных решений. Так, в работах [17–19] рассмотрена проблема построения робастных нечетких регрессионных моделей. При этом наблюдения за входом и выходом многофакторной системы производятся в шкале нечетких чисел. В работе [20] обсуждаются вопросы получения робастных гребневых оценок, в работе [21] рассматриваются вопросы анализа эффективности робастного оценивания моделей регрессии. В работе [22] рассмотрена проблема построения моделей типа Такаги–Сугено в условиях имеющихся выбросов в данных. Вопросам построения робастных решений уделяется внимание и в других методологиях построения зависимостей под управлением данными. Например, задачи получения робастных решений в методе опорных векторов решались в работах [23–25].

В данной работе рассматриваются вопросы построения робастных регрессионных моделей на основе размытых правил Такаги–Сугено. Основное внимание уделено сравнению качества получаемых робастных решений при использовании схем глобального и локального оценивания параметров. Ранее подобное сравнение проводилось для случая выборок, не искаженных выбросами [26].

1. ПОСТАНОВКА ЗАДАЧИ

Рассматривается классическая задача построения адекватной регрессионной зависимости отклика исследуемого объекта y от набора регрессоров $x = (x_1, \dots, x_k)^T$. В силу сложности моделируемого объекта влиянию неучтенных факторов не всегда удается однозначно определить вид статистически устойчивой зависимости $y = f(x)$. Часто можно наблюдать, что в различных частях области определения регрессоров более адекватными могут оказываться различные модели. В регрессионном анализе изучаются модели вида $y = f^T(x)\theta + e = \sum_l^m f_l(x)\theta_l + e$, где y – результирующий признак (отклик, случайная зависимая переменная); $f^T(x) = (f_1(x), f_2(x), \dots, f_m(x))$ – заданная вектор-функция от независимой переменной $x = (x_1, \dots, x_k)^T$, которая может изменяться в области \tilde{X} ; $\theta = (\theta_1, \dots, \theta_m)^T$ – неизвестные параметры, которые необходимо определить по результатам экспериментов (измерений); e – ошибка; T – знак транспонирования. Предположим также, что любое число измерений зависимой переменной можно провести при произвольных $x_j \in \tilde{X}$. Оговорим здесь, что будем использовать в одних случаях обозначение x_j как j -ю точку наблюдений, а в других случаях – как j -ю координату вектора переменных x .

Нечеткие модели Такаги–Сугено типа MISO (Multiple Input-Single Output) представляют собой совокупность правил вида [8]:

$$IF \ x_1 \in A_{1i} \ \& \ \dots \ \& \ x_k \in A_{ki} \ THEN \ y = \eta^i(x), \ i = 1, \dots, M, \quad (1)$$

где A_{ji} – нечеткое подмножество для переменной x_j с функцией принадлежности $\mu_{A_{ji}}(x_j)$; M – число правил, $\eta^i(x)$ – функция, определяющая локальную зависимость отклика y от набора регрессоров $x = (x_1, \dots, x_k)^T$. Четкое значение переменной $\hat{y}(x_j)$, полученное с использованием дефаззификации по методу центра тяжести, вычисляется по формуле

$$\hat{y}(x_j) = \frac{\sum_{i=1}^M \mu_i \eta^i(x_j)}{\sum_{i=1}^M \mu_i(x_j)}; \quad \mu_i(x_j) = \prod_{j=1}^k \mu_{A_{ji}}(x_j). \quad (2)$$

Модель в виде (1), (2) будем называть FLR (Fuzzy Logic Regression) регрессионной моделью. Рассмотрим технику построения FLR регрессии для случая построения одномерной зависимости.

Для случая одной переменной x система правил (1) приобретает вид

$$IF \ x \in A_i \ THEN \ y = \eta^i(x), \quad i = 1, \dots, M, \quad (3)$$

где A_i имеют функцию принадлежности $\mu_{A_i}(x)$.

Необходимость нормировки в (2) отпадает, если считать, что функции принадлежности обладают тем свойством, что в любой точке x выполняется условие

$$\sum_{i=1}^M \mu_{A_i}(x) = 1. \quad (4)$$

В случае локальной линейной зависимости отклика от фактора функции $\eta^i(x)$ приобретают вид $\eta^i(x) = \theta_0^i + \theta_1^i x$, $i = 1, \dots, M$. В итоге можно считать, что регрессия y по x подчиняется следующему уравнению наблюдения:

$$y_u = \sum_{i=1}^M (\theta_0^i + \theta_1^i x_u) \mu_{A_i}(x_u) + e_u, \quad u = 1, \dots, n, \quad (5)$$

где e_u – случайная величина, центрированная с конечной дисперсией.

В случае использования метода наименьших квадратов в глобальном его варианте все неизвестные параметры, входящие в (5), оцениваются совместно. При этом в качестве регрессоров используются следующие:

$$\mu_{A_1}(x), \dots, \mu_{A_M}(x), \quad x\mu_{A_1}(x), \dots, x\mu_{A_M}(x). \quad (6)$$

Одной из серьезных проблем построения нечетких TS (Takagi–Sugeno) моделей является быстрый рост числа правил вида (1) как при увеличении числа нечетких партиций при разбиении области определения входных переменных ровно, так и при увеличении числа входных факторов. Например, при двух факторах с разбиением областей их определения на три партиции и линейной правой частью вида $\eta^i(x) = \theta_0^i + \theta_1^i x_1 + \theta_2^i x_2$ нам придется формировать матрицу наблюдений с числом регрессоров, равным 27. При разбиении области определения на пять партиций число регрессоров будет равно 75. При такой размерности информационной матрицы не исключено появление вычислительных проблем, связанных с ее обращением.

В определенной степени снизить остроту данной проблемы возможно, если использовать метод раздельного (локального) оценивания зависимостей $\eta^i(x)$, $i = 1, \dots, M$ по взвешенному МНК. Пусть μ_i из (2) – сила высказывания для i -го правила в (1). Введем в рассмотрение целевую функцию для взвешенного МНК следующего вида:

$$S(\theta^i) = (y - X\theta^i)^T W_i (y - X\theta^i) = y^T W_i y - 2\theta^{iT} X^T W_i y + \theta^{iT} X^T W_i X \theta^i,$$

где $W_i = \text{diag}(\mu_{i1}, \mu_{i2}, \dots, \mu_{in})$, μ_{ij} – значение μ_i в j -й точке.

Первые частные производные $S(\theta^i)$ по параметрам θ^i имеют вид

$$\partial S(\theta^i) / \partial \theta^i = -2X^T W_i y + 2X^T W_i X \theta^i.$$

Приравнивая их к нулю, получаем систему нормальных уравнений

$$X^T W_i X \hat{\theta}^i = X^T W_i y$$

с решением

$$\hat{\theta}^i = (X^T W_i X)^{-1} X^T W_i y.$$

Видим, что параметры локальных моделей в этом случае оцениваются независимо.

Не всегда возможно считать, что помеха в уравнении наблюдения (5) имеет нормальное распределение или какое-то другое унимодальное распределение с постоянной дисперсией. Наиболее часто рассматриваются отклонения от модельных предположений типа Тьюки–Хьюбера [1]. В этом случае помеха e описывается плотностью вероятности вида

$$(1-\lambda)G_1(\mu_1, \sigma_1) + \lambda G_2(\mu_2, \sigma_2), \quad \sigma_1 < \sigma_2, \quad (7)$$

где $G_1(\mu_1, \sigma_1)$ – функция распределения случайных ошибок, обладающая «хорошими» свойствами (как правило, нормальностью); $G_2(\mu_2, \sigma_2)$ – функция распределения засорений, имеющих вид выбросов как по уровню μ_2 , так и по дисперсии σ_2 ; λ – малая величина, такая что $0 \leq \lambda \leq \lambda_+$. Обычно $\lambda_+ = 0,25$. Смысл модели искажений типа Тьюки–Хьюбера состоит в том, что выборочные значения из распределения (7) состоят из «хороших» и «плохих» наблюдений. «Хорошие» наблюдения появляются с вероятностью $(1-\lambda)$, а «плохие» – с вероятностью λ .

При предположениях (7) относительно помехи наблюдений для оценивания параметров регрессионных моделей применяют так называемые робастные процедуры. Среди них на практике широко используются М-оценки [12–16]. Рассмотрим вопрос их получения на примере обычной линейной по параметрам регрессионной модели наблюдения вида

$$y = f^T(x)\theta + e = \sum_{l=1}^m f_l(x)\theta_l + e.$$

Предположим, что ошибки наблюдений имеют засоренное нормальное распределение, т. е. предполагается, что большая часть наблюдений имеет нормальное распределение, а ряд наблюдений – другое (засоряющее) распределение. В качестве засоряющего может рассматриваться, например, нормальное распределение со значительно большей дисперсией по отношению к дисперсии ошибок основной части выборки.

М-оценки вектора параметров θ находятся путем минимизации функции вида

$$Q = \sum_{i=1}^n \rho(r_i(\hat{\theta})/\hat{\sigma}), \quad (8)$$

где $\rho(r_i(\hat{\theta})/\hat{\sigma})$ – функция потерь; $r_i(\hat{\theta}) = y_i - f^T(x_i)\hat{\theta}$ – остаток i -го измерения; $\hat{\sigma} > 0$ – корень квадратный из дисперсии.

Путем выбора подходящей функции потерь можно обеспечить робастность оценки параметров θ . Необходимые условия минимума функции (8) получают, приравнявая к нулю частные производные по параметрам θ_l :

$$\sum_{i=1}^n \psi(r_i(\hat{\theta})/\hat{\sigma}) f_l(x_i) = 0, \quad l = 1, \dots, m, \quad (9)$$

где $\psi(r_i(\hat{\theta})/\hat{\sigma}) = \partial \rho(r_i(\hat{\theta})/\hat{\sigma}) / \partial \theta$.

В данной постановке задачи параметр $\hat{\sigma}$ предполагается известным. Если же это не так, то параметр $\hat{\sigma}$ приходится оценивать. Так, на практике часто предпочитают в качестве устойчивой оценки $\hat{\sigma}$ для нормального распределения использовать медиану абсолютных отклонений (MAD-оценку):

$$\hat{\sigma} = \text{med}_i |r_i(\hat{\theta})| / 0,67449, \quad (10)$$

где деление на константу обеспечивает асимптотическую несмещенность оценки при незасоренном нормальном распределении.

В предположении, что ошибки наблюдений имеют засоренное нормальное распределение, предложено большое число робастных оценок параметров регрессии, относящихся к классу М-оценок и различающихся используемой функцией потерь. Часто такая функция потерь является квадратичной при малых значениях $r_i(\hat{\theta})/\hat{\sigma}$, а при больших значениях $r_i(\hat{\theta})/\hat{\sigma}$ является «менее возрастающей», чем $(r_i(\hat{\theta})/\hat{\sigma})^2$. Перечислим некоторые из них.

Функция потерь Хьюбера

Швейцарский математик П. Хьюбер предложил следующую функцию потерь [1]:

$$\rho(r_i(\hat{\theta})/\hat{\sigma}) = \begin{cases} \frac{1}{2} (r_i(\hat{\theta})/\hat{\sigma})^2, & |r_i(\hat{\theta})/\hat{\sigma}| < c, \\ c |r_i(\hat{\theta})/\hat{\sigma}| - \frac{1}{2} c^2, & |r_i(\hat{\theta})/\hat{\sigma}| \geq c. \end{cases}$$

Оценочная функция в виде первой производной будет иметь вид

$$\psi(r_i(\hat{\theta})/\hat{\sigma}) = \begin{cases} r_i(\hat{\theta})/\hat{\sigma}, & |r_i(\hat{\theta})/\hat{\sigma}| < c, \\ c \operatorname{sign}(r_i(\hat{\theta})/\hat{\sigma}), & |r_i(\hat{\theta})/\hat{\sigma}| \geq c. \end{cases}$$

Для функции потерь Хьюбера константа c должна быть порядка 1,5. Установлено, что для достижения асимптотической 95 %-й эффективности в предположении о нормальном распределении помех измерений необходимо брать $c = 1,345$. Оценка Хьюбера является устойчивой в том случае, когда засоряющее распределение симметрично, в случае же асимметричных отклонений от нормального распределения она считается недостаточно устойчивой. В последнем случае необходимо использовать функции потерь, при увеличении аргумента растущие медленнее, чем функция Хьюбера. Основная часть множества оценок, устойчивых к асимметричным отклонениям, имеет горизонтальную асимптоту.

Функция потерь Эндрюса

Функция потерь Эндрюса будет иметь вид

$$\rho(r_i(\hat{\theta})/\hat{\sigma}) = \begin{cases} c \left(1 - \cos \frac{r_i(\hat{\theta})/\hat{\sigma}}{c} \right), & |r_i(\hat{\theta})/\hat{\sigma}| < \pi c, \\ 2c, & |r_i(\hat{\theta})/\hat{\sigma}| \geq \pi c. \end{cases}$$

Оценочная функция в виде первой производной будет иметь вид

$$\psi(r_i(\hat{\theta})/\hat{\sigma}) = \begin{cases} \sin \left((r_i(\hat{\theta})/\hat{\sigma})/c \right), & |r_i(\hat{\theta})/\hat{\sigma}| < \pi c, \\ 0, & |r_i(\hat{\theta})/\hat{\sigma}| \geq \pi c. \end{cases}$$

Для функция Эндрюса $c = 2,1$. Фактически, если масштаб известен, величина $c = 1,339$ соответствует 5 %-й потере эффективности.

Функция потерь Дж. Тьюки

Функция потерь Дж. Тьюки будет иметь вид

$$\rho(r_i(\hat{\theta})/\hat{\sigma}) = \begin{cases} (r_i(\hat{\theta})/\hat{\sigma})^2 / 2, & |r_i(\hat{\theta})/\hat{\sigma}| < c, \\ c^2 / 2, & |r_i(\hat{\theta})/\hat{\sigma}| \geq c. \end{cases}$$

Соответствующая ей оценочная функция

$$\psi(r_i(\hat{\theta})/\hat{\sigma}) = \begin{cases} (r_i(\hat{\theta})/\hat{\sigma}), & |r_i(\hat{\theta})/\hat{\sigma}| < c, \\ 0, & |r_i(\hat{\theta})/\hat{\sigma}| \geq c. \end{cases}$$

Для функции Тьюки рекомендовано значение $c = 6$. Если масштаб известен, тогда $c = 3,385$, что соответствует 5 %-й потере эффективности.

Биквадратная функция потерь Тьюки

Биквадратная функция Тьюки будет иметь вид

$$\rho(r_i(\hat{\theta})/\hat{\sigma}) = \begin{cases} \frac{(r_i(\hat{\theta})/\hat{\sigma})^6}{6c^4} - \frac{(r_i(\hat{\theta})/\hat{\sigma})^4}{2c^2} + \frac{(r_i(\hat{\theta})/\hat{\sigma})^2}{2}, & |r_i(\hat{\theta})/\hat{\sigma}| < c, \\ \frac{c^2}{6}, & |r_i(\hat{\theta})/\hat{\sigma}| \geq c. \end{cases}$$

Соответствующая ей оценочная функция

$$\psi(r_i(\hat{\theta})/\hat{\sigma}) = \begin{cases} \left(r_i(\hat{\theta})/\hat{\sigma} \left[1 - \left((r_i(\hat{\theta})/\hat{\sigma})/c \right)^2 \right]^2, & |r_i(\hat{\theta})/\hat{\sigma}| < c, \\ 0, & |r_i(\hat{\theta})/\hat{\sigma}| \geq c. \end{cases}$$

Для биквадратной функция Тьюки рекомендовано значение $c = 6,0$. Если масштаб известен, тогда $c = 4,685$, что соответствует 5 %-й потере эффективности.

2. ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Для поиска значений М-оценок можно применять общие методы оптимизации: градиентный, метод Ньютона, метод сопряженных градиентов и т. д. Однако существует более простой метод минимизации – итеративный МНК (ИМНК), опирающийся на обобщенный МНК. Идея его заключается в следующем. Рассмотрим систему уравнений (9):

$$\sum_{i=1}^n \psi(r_i(\hat{\theta})/\hat{\sigma}) f_l(x_i) = 0, \quad l = 1, \dots, m,$$

где $\psi(r_i(\hat{\theta})/\hat{\sigma}) = \rho'(r_i(\hat{\theta})/\hat{\sigma})$ – производная функции потерь по аргументу.

Перепишем ее левую часть в следующем виде:

$$\begin{aligned} \sum_{i=1}^n \frac{\psi(r_i(\hat{\theta})/\hat{\sigma})}{r_i(\hat{\theta})/\hat{\sigma}} f_l(x_i) r_i(\hat{\theta})/\hat{\sigma} &= \sum_{i=1}^n \frac{1}{\hat{\sigma}} \frac{\psi(r_i(\hat{\theta})/\hat{\sigma})}{r_i(\hat{\theta})/\hat{\sigma}} f_l(x_i) (y_i - f_l(x_i)\hat{\theta}) = \\ &= \sum_{i=1}^n \frac{1}{\hat{\sigma}} w(r_i(\hat{\theta})/\hat{\sigma}) f_l(x_i) (y_i - f_l(x_i)\hat{\theta}), \quad l = 1, \dots, m, \end{aligned}$$

где $w(r_i(\hat{\theta})/\hat{\sigma}) = \frac{\psi(r_i(\hat{\theta})/\hat{\sigma})}{r_i(\hat{\theta})/\hat{\sigma}}$. Переходя к матричным обозначениям и опуская константный множитель $\frac{1}{\hat{\sigma}}$, мы можем переписать полученную систему следующим образом:

$$\left[X^T W(\hat{\theta}) X \right] \hat{\theta} = X^T W(\hat{\theta}) y,$$

где $W(\hat{\theta})$ – диагональная матрица, i -й диагональный элемент которой равен $w(r_i(\hat{\theta})/\hat{\sigma})$; X – матрица значений регрессоров. Таким образом, данная система уравнений соответствует схеме взвешенного МНК с весами, зависящими от $\hat{\theta}$. На k -м шаге итерационного процесса решение может быть записано в явном виде:

$$\hat{\theta}^{(k)} = \left[X^T W(\hat{\theta}^{(k-1)}) X \right]^{-1} X^T W(\hat{\theta}^{(k-1)}) y.$$

Веса на каждой итерации «корректируются» на основе оценки вектора параметров, полученной на предыдущей итерации. В случае одновременного оценивания вектора параметров θ и параметра масштаба $\hat{\sigma}$ следует дополнить каждую итерацию ИМНК вычислением оценки $\hat{\sigma}$, например, по формуле (10).

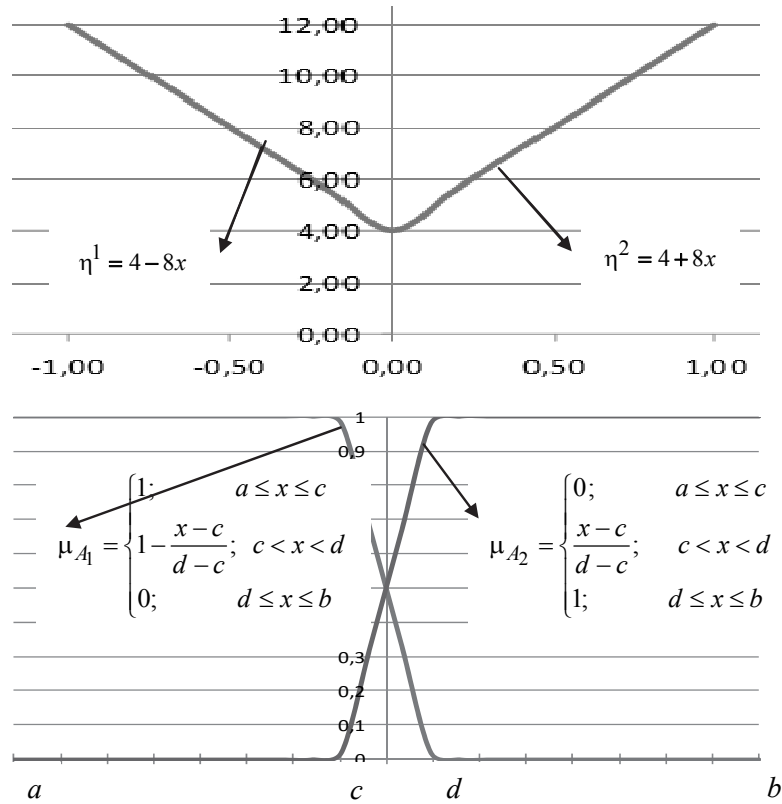
3. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Целью вычислительного эксперимента являлась сравнительное исследование точности локального и глобального оценивания параметров регрессионных моделей при использовании робастных решений по ИМНК на основе функций потерь Хьюбера, Эндрюса, Тьюки и биквадратной функции Тьюки.

В качестве модели, порождающей данные в вычислительном эксперименте, использовалась следующая зависимость:

$$\begin{cases} \text{if } (x \in A_1) & \text{then } \eta^1 = 4 - 8x, \\ \text{if } (x \in A_2) & \text{then } \eta^2 = 4 + 8x, \end{cases}$$

где A_j – нечеткое подмножество для переменной x с функцией принадлежности $\mu_{A_j}(x)$, где $j=1,2$, трапециевидного типа. Вид зависимости представлен на рисунке.



Нечеткая TS-модель с одним входом и одним выходом

Выражения для функций принадлежности:

$$\mu_{A_1} = \begin{cases} 1; & a \leq x \leq c, \\ 1 - \frac{x-c}{d-c}; & c < x < d, \\ 0; & d \leq x \leq b, \end{cases} \quad \mu_{A_2} = \begin{cases} 0; & a \leq x \leq c, \\ \frac{x-c}{d-c}; & c < x < d, \\ 1; & d \leq x \leq b, \end{cases}$$

где $d-c$ – ширина зоны пересечения.

Уровень помехи устанавливался в виде 10 % от мощности незашумленного сигнала. Объем обучаемых выборок варьировался от 15 до 115 наблюдений. Вычисления повторялись по 100 реализациям помехи и далее усреднялись по значению среднеквадратичной ошибки MSE. Ширина зоны пересечения двух партиций варьировалась и выбиралась равной $d-c=0,125$ и $d-c=0,5$.

Ниже в таблицах приведены результаты использования различных функций потерь: Гаусса (МНК), Хьюбера ($c=1,345$), Эндрюса ($c=2,1$), Тьюки ($c=3,385$) и биквадратной Тьюки ($c=4,685$). Для проведения вычислительных экспериментов было разработано программное обеспечение, которое позволяет осуществлять построение кусочно-линейной регрессионной зависимости типа FLR.

Таблица 1

Средние значения MSE при использовании функций принадлежности с шириной области пересечения $d - c = 0,125$ при 10 %-м уровне шума и 0 %-м уровне засорения

№ п/п	МНК		Функция потерь Хьюбера		Функция потерь Эндрюса		Функция потерь Тьюки		Функция потерь биквадратная Тьюки	
	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный
15	0,1590	0,1783	0,1731	0,2032	0,1649	0,1881	0,1673	0,1874	0,1953	0,2425
20	0,1152	0,1321	0,1188	0,1385	0,1161	0,1338	0,1169	0,1351	0,1246	0,1465
25	0,0815	0,0940	0,0894	0,1055	0,0848	0,1003	0,0902	0,1047	0,0937	0,1146
30	0,0687	0,0810	0,0719	0,0869	0,0692	0,0823	0,0697	0,0839	0,0743	0,0903
35	0,0614	0,0679	0,0651	0,0739	0,0621	0,0695	0,0614	0,0685	0,0686	0,0786
40	0,0546	0,0635	0,0578	0,0696	0,0550	0,0649	0,0556	0,0647	0,0589	0,0718
45	0,0515	0,0572	0,0545	0,0606	0,0519	0,0577	0,0521	0,0574	0,0559	0,0618
50	0,0412	0,0479	0,0434	0,0511	0,0414	0,0485	0,0418	0,0485	0,0433	0,0514
55	0,0355	0,0403	0,0373	0,0426	0,0357	0,0408	0,0363	0,0414	0,0379	0,0436
60	0,0374	0,0426	0,0370	0,0427	0,0367	0,0420	0,0368	0,0423	0,0368	0,0427
65	0,0331	0,0392	0,0342	0,0406	0,0331	0,0393	0,0336	0,0398	0,0343	0,0407
70	0,0292	0,0334	0,0302	0,0350	0,0292	0,0337	0,0298	0,0342	0,0307	0,0359
75	0,0265	0,0312	0,0278	0,0330	0,0266	0,0315	0,0269	0,0318	0,0281	0,0337
80	0,0222	0,0255	0,0222	0,0262	0,0219	0,0254	0,0225	0,0262	0,0224	0,0264
85	0,0210	0,0247	0,0228	0,0271	0,0214	0,0252	0,0217	0,0249	0,0228	0,0272
90	0,0230	0,0256	0,0242	0,0274	0,0231	0,0258	0,0233	0,0257	0,0242	0,0275
95	0,0203	0,0236	0,0214	0,0254	0,0205	0,0240	0,0206	0,0240	0,0217	0,0257
100	0,0203	0,0232	0,0219	0,0250	0,0208	0,0237	0,0210	0,0239	0,0222	0,0253
105	0,0202	0,0233	0,0206	0,0240	0,0202	0,0234	0,0205	0,0237	0,0207	0,0242
110	0,0166	0,0192	0,0178	0,0208	0,0170	0,0197	0,0169	0,0197	0,0180	0,0211
115	0,0167	0,0194	0,0175	0,0208	0,0169	0,0198	0,0174	0,0202	0,0178	0,0212

Таблица 2

Средние значения MSE при использовании функций принадлежности с шириной области пересечения $d - c = 0,125$ и функций потерь при 10 %-м уровне шума и 10 %-м уровне засорения симметричного типа

№ п/п	МНК		Функция потерь Хьюбера		Функция потерь Эндрюса		Функция потерь Тьюки		Функция потерь биквадратная Тьюки	
	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный
15	1,0342	1,1460	0,3045	0,3714	0,3665	0,4766	0,5054	0,6670	0,2693	0,3218
20	1,0537	1,2011	0,2398	0,2824	0,2826	0,3508	0,4371	0,5502	0,1641	0,2029
25	0,6319	0,7624	0,1415	0,1723	0,1550	0,1858	0,1593	0,2712	0,1239	0,1516
30	0,4647	0,5474	0,1181	0,1461	0,1257	0,1627	0,1199	0,1656	0,0912	0,1206
35	0,3908	0,4445	0,0876	0,1039	0,0832	0,1037	0,0921	0,1104	0,0682	0,0835
40	0,3353	0,3920	0,0767	0,0899	0,0746	0,0891	0,0757	0,0863	0,0693	0,0791
45	0,2914	0,3411	0,0694	0,0812	0,0707	0,0820	0,0721	0,0844	0,0639	0,0752
50	0,2613	0,3059	0,0513	0,0630	0,0507	0,0630	0,0504	0,0593	0,0472	0,0572
55	0,2074	0,2398	0,0508	0,0597	0,0479	0,0557	0,0484	0,0567	0,0457	0,0545
60	0,2298	0,2659	0,0535	0,0627	0,0529	0,0610	0,0481	0,0577	0,0447	0,0519

Окончание табл. 2

№ п/п	МНК		Функция потерь Хьюбера		Функция потерь Эндрюса		Функция потерь Тьюки		Функция потерь биквадратная Тьюки	
	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный
65	0,1898	0,2131	0,0419	0,0489	0,0427	0,0492	0,0391	0,0457	0,0362	0,0424
70	0,1858	0,2169	0,0352	0,0436	0,0361	0,0439	0,0360	0,0433	0,0309	0,0387
75	0,1845	0,2124	0,0425	0,0494	0,0400	0,0461	0,0388	0,0445	0,0355	0,0421
80	0,1595	0,1893	0,0345	0,0416	0,0355	0,0429	0,0367	0,0430	0,0316	0,0374
85	0,1447	0,1685	0,0346	0,0403	0,0336	0,0386	0,0307	0,0343	0,0292	0,0335
90	0,1452	0,1649	0,0360	0,0401	0,0338	0,0378	0,0338	0,0358	0,0307	0,0339
95	0,1216	0,1420	0,0300	0,0348	0,0276	0,0322	0,0262	0,0306	0,0255	0,0302
100	0,1129	0,1308	0,0269	0,0319	0,0270	0,0321	0,0265	0,0316	0,0231	0,0274
105	0,1310	0,1491	0,0302	0,0352	0,0308	0,0356	0,0299	0,0348	0,0253	0,0299
110	0,1088	0,1268	0,0267	0,0313	0,0269	0,0312	0,0246	0,0285	0,0233	0,0273
115	0,1171	0,1375	0,0236	0,0287	0,0236	0,0285	0,0225	0,0262	0,0209	0,0252

Таблица 3

Средние значения оценивания MSE при использовании функций принадлежности с шириной области пересечения $d - c = 0,5$ при 10 %-м уровне шума и 10 %-м уровне засорения симметричного типа

№ п/п	МНК		Функция потерь Хьюбера		Функция потерь Эндрюса		Функция потерь Тьюки		Функция потерь биквадратная Тьюки	
	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный
15	1,3678	1,9196	0,4254	0,3590	0,4150	0,4576	0,4752	1,0005	0,4317	0,2692
20	1,2173	1,5375	0,4110	0,4567	0,4079	0,4651	0,3858	0,6451	0,3378	0,4284
25	0,9150	1,0924	0,3081	0,1938	0,2836	0,1901	0,2778	0,2039	0,2869	0,1480
30	0,7717	0,9978	0,2664	0,1832	0,2574	0,1672	0,2473	0,1658	0,2589	0,1419
35	0,5626	0,6768	0,2355	0,1155	0,2303	0,1140	0,2221	0,1211	0,2461	0,1045
40	0,5045	0,6322	0,2099	0,1144	0,2023	0,1071	0,1951	0,1055	0,2174	0,0930
45	0,5136	0,5810	0,2244	0,0921	0,2171	0,0893	0,2008	0,0916	0,2263	0,0838
50	0,4052	0,4582	0,2036	0,0958	0,1903	0,0899	0,1862	0,0972	0,2156	0,0826
55	0,3588	0,3338	0,2122	0,0720	0,1939	0,0707	0,1809	0,0703	0,2147	0,0661
60	0,4207	0,4321	0,2016	0,0703	0,1912	0,0683	0,1735	0,0674	0,2022	0,0589
65	0,3095	0,3266	0,1805	0,0561	0,1734	0,0508	0,1634	0,0505	0,1944	0,0489
70	0,3415	0,3359	0,1985	0,0620	0,1921	0,0582	0,1734	0,0548	0,2064	0,0530
75	0,3110	0,2825	0,1919	0,0532	0,1814	0,0506	0,1703	0,0503	0,1976	0,0450
80	0,3125	0,2739	0,1936	0,0546	0,1832	0,0509	0,1672	0,0509	0,1961	0,0462
85	0,3529	0,2923	0,2048	0,0539	0,1905	0,0499	0,1715	0,0462	0,2026	0,0438
90	0,2761	0,2185	0,1864	0,0381	0,1708	0,0376	0,1603	0,0366	0,1899	0,0342
95	0,2867	0,2244	0,1787	0,0462	0,1686	0,0454	0,1498	0,0424	0,1810	0,0393
100	0,2964	0,2277	0,1910	0,0414	0,1784	0,0373	0,1631	0,0358	0,1947	0,0344
105	0,2487	0,2005	0,1727	0,0423	0,1622	0,0395	0,1484	0,0359	0,1789	0,0356
110	0,2357	0,2108	0,1800	0,0416	0,1715	0,0365	0,1559	0,0357	0,1894	0,0345
115	0,2756	0,2043	0,1889	0,0382	0,1754	0,0347	0,1610	0,0330	0,1953	0,0315

В условиях отсутствия выбросов (см. табл. 1) стандартные МНК на основе гауссовой функции потерь показывают лучшие результаты. На втором месте,

как и следовало ожидать, – функция потерь Хьюбера. В условиях симметричного засорения (см. табл. 2, 3) все робастные функции потерь показывают устойчиво лучшие результаты, чем функция потерь Гаусса. Что касается использования локального и глобального метода оценивания, то здесь результаты во многом зависят от совместного расположения нечетких партий. Априори можно было ожидать, что глобальный метод оценивания имеет преимущество над локальным. При достаточно широкой области пересечения нечетких партий ($d - c = 0,5$) в действительности так и происходит (см. табл. 3). В случае, когда область пересечения нечетких партий достаточно узкая, то ситуация меняется в точности наоборот (см. табл. 2). Дополнительно был проведен эксперимент, когда восстанавливались модели по незашумленным данным. В этих условиях глобальный метод оценивания оказался лучше локального. Полученные результаты позволяют говорить, что в условиях использования зашумленных данных методы локального и глобального оценивания конкурентные. И если удается область пересечения нечетких партий выбирать достаточно узкой, то преимущество остается за локальным методом оценивания.

Таблица 4

Средние значения MSE при использовании функции принадлежности с шириной области пересечения $d - c = 0,125$ при 10 %-м уровне шума и 10 %-м уровне засорения несимметричного типа ($\mu_2 = 5$)

№ п/п	МНК		Функция потерь Хьюбера		Функция потерь Эндрюса		Функция потерь Тьюки		Функция потерь биквадратная Тьюки	
	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный
15	2,2953	2,4367	0,5316	0,6017	0,7161	0,8608	1,1257	1,0020	0,2534	0,6356
20	1,2667	1,4507	0,1916	0,2315	0,1710	0,2032	0,3576	0,4062	0,1438	0,1781
25	1,2008	1,3338	0,1791	0,2124	0,1581	0,1885	0,1631	0,1979	0,1285	0,1496
30	0,8423	0,9507	0,1108	0,1375	0,0952	0,1243	0,1016	0,1298	0,0802	0,0992
35	0,7018	0,7995	0,1146	0,1384	0,0892	0,1045	0,0902	0,1069	0,0799	0,0941
40	0,8135	0,9066	0,0941	0,1086	0,0783	0,1050	0,0973	0,1217	0,0661	0,0763
45	0,7854	0,8670	0,0865	0,1018	0,0668	0,0809	0,0828	0,0754	0,0549	0,0640
50	0,6876	0,7606	0,0733	0,0861	0,0547	0,0625	0,0506	0,0604	0,0457	0,0558
55	0,5796	0,6423	0,0624	0,0736	0,0512	0,0611	0,0500	0,0597	0,0470	0,0569
60	0,5181	0,5622	0,0533	0,0620	0,0432	0,0511	0,0427	0,0505	0,0412	0,0488
65	0,6050	0,6436	0,0557	0,0615	0,0407	0,0474	0,0379	0,0445	0,0346	0,0404
70	0,4818	0,5183	0,0530	0,0594	0,0406	0,0458	0,0365	0,0419	0,0347	0,0401
75	0,4557	0,5013	0,0490	0,0578	0,0369	0,0446	0,0353	0,0424	0,0338	0,0411
80	0,5456	0,5801	0,0509	0,0573	0,0369	0,0423	0,0344	0,0393	0,0303	0,0352
85	0,5220	0,5583	0,0439	0,0502	0,0321	0,0367	0,0305	0,0346	0,0290	0,0339
90	0,4861	0,5218	0,0451	0,0514	0,0322	0,0383	0,0307	0,0359	0,0284	0,0338
95	0,4715	0,5024	0,0402	0,0460	0,0291	0,0339	0,0278	0,0325	0,0246	0,0291
100	0,4377	0,4726	0,0371	0,0438	0,0295	0,0352	0,0269	0,0327	0,0250	0,0309
105	0,3845	0,4114	0,0319	0,0373	0,0247	0,0291	0,0225	0,0265	0,0206	0,0242
110	0,5067	0,5345	0,0378	0,0422	0,0264	0,0306	0,0257	0,0303	0,0239	0,0281
115	0,4441	0,4782	0,0377	0,0427	0,0255	0,0297	0,0259	0,0301	0,0223	0,0265

В условиях засорения несимметричного типа явное преимущество имеют функции потерь Эндрюса и Тьюки, имеющие участки со значением оценочной функции, равным нулю (см. табл. 4).

Таблица 5

Средние значения MSE при использовании функций принадлежности с шириной области пересечения $d - c = 0,125$ при 10 %-м уровне шума и 10 %-м уровне засорения симметричного типа, сосредоточенных только в одной партии

№ п/п	МНК		Функция потерь Хьюбера		Функция потерь Эндрюса		Функция потерь Тьюки		Функция потерь биквадратная Тьюки	
	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный
15	0,4734	0,5464	0,1882	0,2533	0,2297	0,3172	0,2306	0,3749	0,1788	0,3457
20	0,4451	0,4970	0,1401	0,1678	0,1322	0,1639	0,1372	0,1820	0,1299	0,1642
25	0,3108	0,3768	0,0781	0,0985	0,0901	0,1100	0,0883	0,1037	0,0808	0,1000
30	0,2035	0,2401	0,0748	0,0920	0,0685	0,0852	0,0729	0,0864	0,0687	0,0835
35	0,2260	0,2552	0,0634	0,0739	0,0630	0,0732	0,0615	0,0725	0,0582	0,0687
40	0,1691	0,2010	0,0587	0,0695	0,0567	0,0670	0,0593	0,0726	0,0574	0,0672
45	0,1338	0,1553	0,0494	0,0583	0,0477	0,0554	0,0487	0,0560	0,0459	0,0544
50	0,1327	0,1556	0,0470	0,0551	0,0468	0,0534	0,0465	0,0531	0,0452	0,0532
55	0,1062	0,1250	0,0410	0,0488	0,0394	0,0464	0,0400	0,0469	0,0422	0,0498
60	0,0847	0,0980	0,0379	0,0439	0,0373	0,0427	0,0375	0,0429	0,0369	0,0431
65	0,1025	0,1198	0,0336	0,0399	0,0326	0,0383	0,0341	0,0406	0,0333	0,0397
70	0,1046	0,1206	0,0306	0,0356	0,0300	0,0348	0,0313	0,0360	0,0287	0,0334
75	0,0821	0,0927	0,0262	0,0308	0,0248	0,0290	0,0252	0,0289	0,0252	0,0296
80	0,0917	0,1055	0,0278	0,0341	0,0276	0,0337	0,0284	0,0340	0,0263	0,0330
85	0,0696	0,0796	0,0228	0,0269	0,0219	0,0256	0,0217	0,0251	0,0210	0,0250
90	0,0605	0,0683	0,0223	0,0266	0,0209	0,0249	0,0209	0,0251	0,0213	0,0257
95	0,0767	0,0909	0,0223	0,0267	0,0200	0,0238	0,0194	0,0227	0,0187	0,0225
100	0,0680	0,0799	0,0230	0,0270	0,0211	0,0251	0,0207	0,0241	0,0206	0,0245
105	0,0645	0,0759	0,0184	0,0227	0,0182	0,0220	0,0188	0,0227	0,0174	0,0215
110	0,0539	0,0615	0,0206	0,0241	0,0197	0,0229	0,0203	0,0235	0,0199	0,0232
115	0,0434	0,0499	0,0178	0,0207	0,0176	0,0200	0,0176	0,0199	0,0171	0,0197

Таблица 6

Средние значения MSE при использовании функций принадлежности с шириной области пересечения $d - c = 0,125$ и функций потерь при 10 %-м уровне шума и 10 %-м уровне засорения несимметричного типа ($\mu_2 = 5$), сосредоточенных только в одной партии

№ п/п	МНК		Функция потерь Хьюбера		Функция потерь Эндрюса		Функция потерь Тьюки		Функция потерь биквадратная Тьюки	
	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный
15	0,9649	1,1007	0,2971	0,4200	0,3626	0,4617	0,3512	0,4862	0,2198	0,3953
20	0,8936	1,0363	0,1406	0,1796	0,1071	0,1286	0,1135	0,1629	0,1126	0,1377
25	0,6483	0,7438	0,1265	0,1592	0,1005	0,1210	0,0981	0,1226	0,0928	0,1132
30	0,5236	0,6016	0,0860	0,1033	0,0754	0,0921	0,0822	0,0918	0,0727	0,0871
35	0,3720	0,4045	0,0777	0,0898	0,0706	0,0809	0,0715	0,0805	0,0671	0,0784
40	0,4489	0,5095	0,0664	0,0774	0,0556	0,0643	0,0578	0,0653	0,0531	0,0624
45	0,2829	0,3172	0,0508	0,0577	0,0482	0,0545	0,0471	0,0520	0,0495	0,0585
50	0,3177	0,3570	0,0497	0,0590	0,0428	0,0507	0,0432	0,0519	0,0424	0,0504

Окончание табл. 6

№ п/п	МНК		Функция потерь Хьюбера		Функция потерь Эндрюса		Функция потерь Тьюки		Функция потерь биквадратная Тьюки	
	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный	Локальный	Глобальный
55	0,2807	0,3197	0,0432	0,0515	0,0367	0,0451	0,0382	0,0462	0,0374	0,0462
60	0,2882	0,3180	0,0463	0,0538	0,0413	0,0474	0,0417	0,0481	0,0388	0,0453
65	0,2590	0,2860	0,0398	0,0469	0,0344	0,0408	0,0329	0,0403	0,0304	0,0364
70	0,2477	0,2828	0,0359	0,0441	0,0302	0,0367	0,0295	0,0357	0,0288	0,0357
75	0,2872	0,3108	0,0330	0,0374	0,0272	0,0313	0,0285	0,0322	0,0272	0,0311
80	0,1857	0,2040	0,0306	0,0347	0,0259	0,0295	0,0251	0,0289	0,0256	0,0293
85	0,2695	0,2979	0,0309	0,0368	0,0251	0,0300	0,0250	0,0300	0,0250	0,0303
90	0,2205	0,2448	0,0249	0,0293	0,0213	0,0249	0,0217	0,0253	0,0214	0,0253
95	0,1774	0,1994	0,0261	0,0309	0,0224	0,0264	0,0224	0,0262	0,0224	0,0264
100	0,1889	0,2104	0,0253	0,0286	0,0229	0,0251	0,0233	0,0259	0,0224	0,0250
105	0,1988	0,2159	0,0211	0,0254	0,0180	0,0216	0,0182	0,0219	0,0176	0,0212
110	0,1739	0,1897	0,0218	0,0251	0,0192	0,0221	0,0203	0,0232	0,0187	0,0216
115	0,2130	0,2320	0,0217	0,0257	0,0169	0,0200	0,0175	0,0208	0,0162	0,0194

Представленные в табл. 5, 6 результаты отражают ситуацию, когда засорение имеет пространственную привязку. В данном случае засоренные наблюдения располагались только в области одной из нечетких партиций. Отмеченные выше закономерности (см. табл. 1–4) здесь также подтверждаются.

ЗАКЛЮЧЕНИЕ

Проведенные исследования показали, что в практическом плане для построения устойчивых моделей TS типа по загрязненным выборкам можно использовать любые из рассмотренных вариантов М-оценок. В условиях загрязнений несимметричного типа предпочтение следует отдавать функциям потерь Эндрюса и Тьюки. Рассмотренные локальный и глобальный методы оценивания параметров показали близкие результаты. Преимущество локального метода оценивания может проявиться в случае, когда области пересечения нечетких партиций узкие.

СПИСОК ЛИТЕРАТУРЫ

1. Takagi T., Sugeno M. Fuzzy identification of systems and its applications to modeling and control // IEEE Transactions on Systems, Man and Cybernetics. – 1985. – Vol. 15, N 1. – P. 116–132.
2. Babuška R., Verbruggen H.B. Constructing fuzzy models by product space clustering // Fuzzy model identification: selected approaches / ed. by H. Hellendoorn, D. Driankov. – Berlin: Springer, 1997. – P. 53–90.
3. Babuška R. Fuzzy modeling for control. – London; Boston: Kluwer Academic Publishers, 1998. – 257 p.
4. Lilly J.H. Fuzzy control and identification. – Hoboken, NJ: Wiley, 2010. – 231 p.
5. Понов А.А. Регрессионное моделирование на основе нечетких правил // Сборник научных трудов НГТУ. – 2000. – № 2 (19). – С. 49–57.
6. Понов А.А. Оптимальное планирование эксперимента в задачах структурной и параметрической идентификации моделей многофакторных систем: монография. – Новосибирск: Изд-во НГТУ, 2013. – 296 с.

7. Kosko B. Fuzzy systems as universal approximators // IEEE Transactions on Computers. – 1994. – Vol. 43, iss. 11. – P. 1329–1333.
8. Hao Ying. General SISO Takagi-Sugeno fuzzy systems with linear rule consequent are universal approximators // Transactions on Fuzzy Systems. – 1998. – Vol. 6, N 4. – P. 582–587.
9. Popov A.A., Bykhanov K.V. Modeling volatility of time series using fuzzy GARCH models // KORUS-2005. The 9 Russian-Korean International Symposium on Science and Technology: Proceedings, Novosibirsk, 26 June – 2 July 2005. – Novosibirsk, 2005. – Vol. 1. – P. 687–692.
10. Пегат А. Нечеткое моделирование и управление: пер. с англ. – 2-е изд. – М.: Бином, 2013. – 798 с.
11. Попов А.А. Конструирование дискретных и непрерывно-дискретных моделей регрессионного типа // Сборник научных трудов НГТУ. – 1996. – № 1. – С. 21–30.
12. Хьюбер П. Робастность в статистике. – М.: Мир, 1984. – 304 с.
13. Мудров В.И., Кушко В.Л. Методы обработки измерений: квазиравдоподобные оценки. – 2-е изд., перераб. и доп. – М.: Радио и связь, 1983. – 304 с.
14. Себер Дж. Линейный регрессионный анализ. – М.: Мир, 1980. – 456 с.
15. Theory and applications of recent robust methods, statistics for industry and technology / ed. by M. Hubert, G. Pison, A. Struyf, S. Van Aelst. – Basel: Birkhauser, 2004. – 400 p.
16. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: исследование зависимостей. – М.: Финансы и статистика, 1985. – 487 с.
17. Sohn B.-Y. Robust fuzzy linear regression based on M-estimators // Journal of Applied Mathematics and Computing. – 2005. – Vol. 18, N 1–2. – P. 591–601.
18. Kula K., Tank F., Dalkılıç T. A study on fuzzy robust regression and its application to insurance // Mathematical and Computational Applications. – 2012. – Vol. 17, N 3. – P. 223–234.
19. Грицюк В.И. Нечеткий робастный метод оценивания регрессионной модели в страховании // Вестник ХНАДУ. – 2016. – Вып. 72. – С. 88–93.
20. Грицюк В.И. Улучшенные робастные гребневые оценки регрессии // Восточно-Европейский журнал передовых технологий. – 2015. – Т. 1, вып. 9. – С. 53–57.
21. Антонов В.А., Шамша Б.В. Оценка эффективности алгоритмов робастного оценивания // Радиоэлектроника и информатика. – 2000. – № 2 (11). – С. 93–97.
22. Chuang C.-C., Su S.-F., Chen S.-S. Robust TSK fuzzy modeling for function approximation with outliers // IEEE Transactions on Fuzzy Systems. – 2001. – Vol. 9, iss. 6. – P. 810–821.
23. Гулятьева Т.А., Попов А.А., Саутин А.С. Методы статистического обучения в задачах регрессии и классификации: монография. – Новосибирск: Изд-во НГТУ, 2016. – 322 с.
24. Попов А.А., Саутин С.А. Построение регрессионных зависимостей с использованием алгоритма опорных векторов с адаптивными функциями потерь // Научный вестник НГТУ. – 2011. – № 1 (42). – С. 17–26.
25. Попов А.А., Саутин А.С. Использование робастных функций потерь в алгоритме опорных векторов при решении задачи построения регрессии // Научный вестник НГТУ. – 2009. – № 4 (37). – С. 45–56.
26. Попов А.А., Холдонов А.А. Глобальное и локальное оценивание параметров регрессионных моделей при использовании концепции нечетких систем // Сборник научных трудов НГТУ. – 2015. – № 4 (82). – С. 56–66.

Попов Александр Александрович, доктор технических наук, профессор кафедры теоретической и прикладной информатики Новосибирского государственного технического университета. Основное направление научных исследований – статистические методы анализа данных, оптимальное планирование экспериментов. Имеет более 150 публикаций, в том числе 3 монографии. E-mail: a.popov@corp.nstu.ru

Холдонов Абдурахмон Абдуллоевич, аспирант кафедры теоретической и прикладной информатики Новосибирского государственного технического университета. Основное направление исследований – статистические методы анализа данных. Имеет 2 публикации. E-mail: firuz_530_11_29@mail.ru

A Comparative Study of the Accuracy of Local and Global Estimation of Regression Model Parameters Using Various Types of M-estimators *

A.A. POPOV¹, A.A. Kholdonov²

¹ Novosibirsk State Technical University, 20, K. Marx Prospekt, Novosibirsk, 630073, Russian Federation, D. Sc. (Eng.), professor. E-mail: a.popov@corp.nstu.ru

² Novosibirsk State Technical University, 20, K. Marx Prospekt, Novosibirsk, 630073, Russian Federation, post-graduate student. E-mail: firuz_530_11_29@mail.ru

Local and global estimation of regression model parameters when making robust solutions are described in the paper. Robust solutions are based on M-estimators by using the Iterative Method of Least Squares (IMLS). Several types of M-estimators based on the Huber, Andrews, and Tukey loss functions are considered. The purpose of this study is to provide practical recommendations on local and global parameter estimation of regression models using various types of M-estimators. When using the local method of least squares, parameters of individual linear models within the rule-based system are evaluated separately. In the global version of least squares estimation is carried out jointly for the entire set of unknown parameters. The Takagi-Sugeno model was used as a rule-based system. When splitting the domain of the input factor estimation, trapezoidal membership functions were used. To evaluate the accuracy of the solutions the root mean square error (MSE) was used. To carry out computational experiments we developed appropriate software. A computational experiment was conducted on simulated data. Linear dependence on the input factor was used as a data generating model. Symmetric and asymmetric data sample noises are considered in the paper. The dispersion of noise (noise level) was determined as a percentage of the signal power. The results of the studies are presented in tables. It is concluded that any of the above M-estimators can be used to construct stable models by noised samples. The local and global methods of parameter estimation considered in the paper showed similar results. The advantage of the local method of parameter estimation can become apparent in the case when the intersection domains of membership functions are narrow.

Keywords: fuzzy system, regression model, method of least squares, Takagi-Sugeno model, M-estimators, membership function, mean square error (MSE), symmetric and asymmetric noise

DOI: 10.17212/1814-1196-2017-4-67-84

REFERENCES

1. Takagi T., Sugeno M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 1985, vol. 15, no. 1, pp. 116–132.
2. Babuška R., Verbruggen H.B. Constructing fuzzy models by product space clustering. *Fuzzy model identification: selected approaches*. Ed. by H. Hellendoorn, D. Driankov. Berlin, Springer, 1997, pp. 53–90.
3. Babuška R. *Fuzzy modeling for control*. London, Boston, Kluwer Academic Publishers, 1998. 257 p.
4. Lilly J.H. *Fuzzy control and identification*. Hoboken, NJ, Wiley, 2010. 231 p.
5. Popov A.A. Regressionnoe modelirovanie na osnove nechetkikh pravil [Regression modeling based on fuzzy rules]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2000, no. 2 (19), pp. 49–57.
6. Popov A.A. *Optimal'noe planirovanie eksperimenta v zadachakh strukturnoi i parametricheskoi identifikatsii modelei mnogofaktornykh sistem* [Optimal planning of the experiment in problems of structural and parametric identification of models of multifactor systems]. Novosibirsk, NSTU Publ., 2013. 296 p.
7. Kosko B. Fuzzy systems as universal approximators. *IEEE Transactions on Computers*, 1994, vol. 43, iss. 11, 1329–1333.
8. Hao Ying. General SISO Takagi-Sugeno fuzzy systems with linear rule consequent are universal approximators. *Transactions on Fuzzy Systems*, 1998, vol. 6, no. 4, pp. 582–587.
9. Popov A.A., Bykhanov K.V. Modeling volatility of time series using fuzzy GARCH models. *KORUS-2005. The 9 Russian-Korean International Symposium on Science and Technology: Proceedings*, Novosibirsk, 26 June – 2 July 2005, vol. 1, pp. 687–692.
10. Piegat A. *Fuzzy modeling and control*. Heidelberg, Physica-Verlag, 2001 (Russ. ed.: Pegat A. *Nechetkoe modelirovanie i upravlenie*. 2nd ed. Moscow, Binom Publ., 2013. 798 p.).
11. Popov A.A. Konstruirovaniye diskretnykh i nepreryvno-diskretnykh modelei regressionnogo tipa [Construction of discrete and continuous-discrete models such as regression]. *Sbornik nauchnykh*

* Received 29 September 2017.

trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university, 1996, no. 1, pp. 21–30.

12. Huber P.J. *Robust statistics*. New York, Wiley and Sons, 1981 (Russ. ed.: Kh'yuber P. *Robastnost' v statistike*. Moscow, Mir Publ., 1984. 304 p.).

13. Mudrov V.I., Kushko V.L. *Metody obrabotki izmerenii: kvazipravdopodobnye otsenki* [Methods of measurement processing: quasi plausible estimates]. 2nd ed. Moscow, Radio i svyaz' Publ., 1983. 304 p.

14. Seber G.A.F. *Linear regression analysis*. New York, Wiley and Sons, 1977 (Russ. ed.: Seber Dzh. *Lineinyi regressionnyi analiz*. Moscow, Mir Publ., 1980. 456 p.).

15. Hubert M., Pison G., Struyf A., Aelst S. van, eds. *Theory and applications of recent robust methods, statistics for industry and technology*. Basel, Birkhauser, 2004. 400 p.

16. Aivazyan S.A., Enyukov I.S., Meshalkin L.D. *Prikladnaya statistika: issledovanie zavisimosti* [Applied statistics: research dependencies]. Moscow, Finansy i statistika Publ., 1985. 487 p.

17. Sohn B.-Y. Robust fuzzy linear regression based on M-estimators. *Journal of Applied Mathematics and Computing*, 2005, vol. 18, no. 1–2, pp. 591–601.

18. Kula K., Tank F., Dalkılıç T. A study on fuzzy robust regression and its application to insurance. *Mathematical and Computational Applications*, 2012, vol. 17, no. 3, pp. 223–234.

19. Gritsyuk V.I. Nechetkii robstnyy metod otsenivaniya regressionnoi modeli v strakhovanii [Fuzzy robust method for regression model estimation in insurance]. *Vestnik Khar'kovskogo natsional'nogo avtomobil'no-dorozhnogo universiteta – Bulletin of Kharkov National Automobile and Highway University*, 2016, iss. 72, pp. 88–93.

20. Gritsyuk V.I. Uluchshennyye robstnyye grebnevyye otsenki regressii [Improved robust ridge regression estimation]. *Vostochnoo-Evropeiskii zhurnal peredovykh tekhnologii – Eastern-European Journal of Enterprise Technologies*, 2015, vol. 1, iss. 9, pp. 53–57.

21. Antonov V.A., Shamsha B.V. Otsenka effektivnosti algoritmov robstnogo otsenivaniya [Evaluating the effectiveness of robust estimation algorithms]. *Radioelektronika i informatika – Radioelectronics and Informatics*, 2000, no. 2 (11), pp. 93–97.

22. Chuang C.-C., Su S.-F., Chen S.-S. Robust TSK fuzzy modeling for function approximation with outliers. *IEEE Transactions on Fuzzy Systems*, 2001, vol. 9, iss. 6, pp. 810–821.

23. Gul'tyaeva T.A., Popov A.A., Sautin A.S. *Metody statisticheskogo obucheniya v zadachakh regressii i klassifikatsii* [Statistical learning methods in regression and classification problems]. Novosibirsk, NSTU Publ., 2016. 322 p.

24. Popov A.A., Sautin S.A. Postroenie regressionnykh zavisimosti s ispol'zovaniem algoritma opornykh vektorov s adaptivnymi funktsiyami poter' [Adaptive loss functions in support vector regression]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2011, no. 1 (42), pp. 17–26.

25. Popov A.A., Sautin A.S. Ispol'zovanie robstnykh funktsii poter' v algoritme opornykh vektorov pri reshenii zadachi postroeniya regressii [Using robust loss functions in support vector machine algorithm for solving the problem of constructing a regression]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2009, no. 4 (37), pp. 45–56.

26. Popov A.A., Holdonov A.A. Global'noe i lokal'noe otsenivanie parametrov regressionnykh modelei pri ispol'zovanii kontseptsii nechetkikh sistem [Global and local parameter estimation of regression models using fuzzy systems concept]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2015, no. 4 (82), pp. 56–66.

Для цитирования:

Попов А.А., Холдонов А.А. Сравнительное исследование точности локального и глобального оценивания параметров регрессионных моделей при использовании различных типов М-оценок // Научный вестник НГТУ. – 2017. – № 4 (69). – С. 67–84. – doi: 10.17212/1814-1196-2017-4-67-84.

For citation:

Popov A.A., Kholdonov A.A. Sravnitel'noe issledovanie tochnosti lokal'nogo i global'nogo otsenivaniya parametrov regressionnykh modelei pri ispol'zovanii razlichnykh tipov M-otsenok [A Comparative Study of the Accuracy of Local and Global Estimation of Regression Model Parameters Using Various Types of M-estimators]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2017, no. 4 (69), pp. 67–84. doi: 10.17212/1814-1196-2017-4-67-84.