

Нечеткая линейная регрессия в задачах оценки

Аннотация. Нечеткая линейная регрессия применяется как альтернатива классическим эконометрическим методам для моделирования экономических явлений в случае, когда использование вероятностных и регрессионных моделей выглядит недостаточно обоснованным. Рассмотрены три задачи: оценка стоимости квартир в Москве, оценка стоимости коттеджей в Московской области и оценка стоимости колготок. Общим для всех трех задач является то, что неопределенность, связанная с объясняемой величиной, обусловлена факторами, теоретико-вероятностная природа которых неочевидна. В этой ситуации подход, основанный на методах теории нечетких множеств, применение нечеткой линейной регрессии выглядят достаточно обоснованными и приводят к экономически осмысленным результатам.

Ключевые слова: нечеткая линейная регрессия; оценка недвижимости; ценообразование.

Abstract. Fuzzy linear regression is considered as an alternative to conventional econometric methods for modeling of economic phenomena where the probability and regression models do not look sufficiently justified. Three problems are studied: assessment of apartments in Moscow and houses in the Moscow region, and the valuation of tights. Common to all three problems is the following feature: uncertainty is caused by factors of not obvious probabilistic nature. In such a situation, an approach based on fuzzy methods including fuzzy linear regression seems quite reasonable and lead to an economically meaningful results.

Keywords: fuzzy linear regression; real estate; pricing.



Вишнякова Е.В.,

студентка магистратуры
Финансового университета
✉ katushka353@mail.ru



Колодяжная Ю.А.,

студентка магистратуры
Финансового университета
✉ yulya.shine@gmail.com



Иванова Е.В.,

студентка магистратуры
Финансового университета
✉ IvanovaEvgeniyaVl@gmail.com



Хамидуллина Л.Ф.,

студентка магистратуры
Финансового университета
✉ liliya_hamid@mail.ru



Камалов С.М.,

студент магистратуры
Финансового университета
✉ kamalov_stanislav@mail.ru

Введение

При принятии решений каждый субъект экономической деятельности неизбежно сталкивается с неопределенностью. Данные, собираемые для осуществления дальнейшего их анализа, характеризуются неточностью, приводящей в ряде

Научный руководитель: **Гисин В.Б.**, кандидат физико-математических наук, профессор, заведующий кафедрой «Математика».

случаев к серьезному искажению получаемых результатов.

При этом как исходные данные, так и параметры моделей, а вместе с тем моделируемые объекты и явления не всегда могут быть оценены с использованием абсолютно точных количественных характеристик [1]. Отклонения от возможных прогнозируемых значений естественным образом возникают в финансово-экономической среде. Они обусловлены тем, что неопределенность возникает не как результат воздействия большого числа случайных факторов, а вследствие объективной неточности и недостаточности имеющейся информации. Например, в работе [2] нечеткие величины использовались для оценки эффективности инвестиционных проектов, в статье [3] приводятся оценки цены уточнения информации в случае, когда финансовые вычисления проводятся в условиях нечеткости.

В статье нечеткая линейная регрессия рассматривается как альтернатива традиционным эконометрическим подходам. В основе вероятностных моделей лежат достаточно убедительные предположения, что зачастую вызывает сомнения в их обоснованности. Применение нечеткого подхода отличается большей гибкостью и позволяет учитывать в модели тот факт, что исходные переменные являются не совсем точными.

Одним из наиболее распространенных способов прогнозирования величины является построение линейной множественной регрессии как функции от экономически обоснованного набора факторов. Однако в случае, когда прогнозируемая величина имеет признаки нечеткости, линейная регрессия может оказаться недостаточно адекватной моделью, поскольку прогноз в такой ситуации есть вероятность попадания величины в прогнозный интервал.

При построении нечеткой линейной регрессии коэффициенты экзогенных переменных искали в виде нечетких чисел. В итоге эндогенная переменная (цена на квартиру, коттедж или колготки) также представлена в виде нечеткого числа, а прогнозный интервал – нечеткого множества. Нечеткая линейная регрессия использовалась в самых различных задачах [4]. В том числе тех, где нечеткая линейная регрессия применялась для оценки недвижимости и построения моделей роста [5, 6].

Представленное исследование организовано следующим образом. Чтобы сделать изложение замкнутым, в разделе 1 приводятся необходимые сведения о нечеткой линейной регрессии. В разделе 2 нечеткая линейная регрессия используется для оценки стоимости коттеджей в Московской области, в разделе 3 – для оценки стоимости квартир в Москве, в разделе 4 – для определения цены колготок.

В терминологии, связанной с нечеткими множествами, мы следуем [7] и [8]. В методике оценки недвижимости мы следуем [10].

1. Треугольные нечеткие числа и нечеткая линейная регрессия

Нечеткая величина A задается функцией принадлежности $\mu_A: \mathbf{R} \rightarrow [0; 1]$. Число a , для которого $\mu_A(a) = 1$, называется модальным значением величины A . Замыкание множества $\{x | \mu_A(x) > 0\}$ называется носителем нечеткой величины A и обозначается $\text{supp } A$. «Обычное» четкое число a можно рассматривать как нечеткое с носителем, состоящим из единственной точки a , которая является модальным значением. Таким образом, нечеткое число оказывается как бы «размазанным» по своему носителю. Число $\mu_A(x)$ указывает на возможность того, что x является значением величины A .

Нечеткие величины, описываемые выражениями типа «примерно a », обычно представляют так называемыми треугольными нечеткими числами. Треугольное нечеткое число A задается тройкой чисел $(a^L; a; a^R)$, где $a^L \leq a \leq a^R$. Отрезок $[a^L; a^R]$ является носителем множества A , число a – модальным значением. Для обозначения треугольного нечеткого числа используется задающая его тройка $(a^L; a; a^R)$. С учетом этого соглашения тройка $(k; k; k)$ задает четкое число k . Треугольное нечеткое число $A = (a^L; a; a^R)$ называется симметричным, если $a - a^L = a^R - a$. Симметричное треугольное нечеткое число задается тройкой вида $(a - d; a; a + d)$, $d \geq 0$; число d называют мерой нечеткости треугольного симметричного нечеткого числа (иногда говорят об интервале толерантности). Функция принадлежности симметричного треугольного нечеткого числа $(a - d; a; a + d)$ при $d > 0$ имеет вид

$$\mu_A(x) = \max\left(0, 1 - \frac{|x - a|}{d}\right).$$

Сумма треугольных нечетких чисел $A = (a^L; a; a^R)$ и $B = (b^L; b; b^R)$ – это треугольное нечеткое число:

$$A + B = (a^L + b^L; a + b; a^R + b^R).$$

Произведение треугольного нечеткого числа $A = (a - d; a; a + d)$ на четкое число k – это треугольное нечеткое число:

$$kA = (ka - |k|d; ka; ka + |k|d).$$

Опишем общую схему нечеткой линейной регрессии.

В общем случае задача нечеткой линейной регрессии может быть поставлена следующим образом. Имея m результатов наблюдений $(y_j, x_j), j = 1, 2, \dots, m$, требуется оптимальным образом определить, вообще говоря, нечеткие коэффициенты A_0, A_1, \dots, A_n . Оптимальность выражается двумя условиями:

(R1) для каждого j число y_j принадлежит носителю нечеткой величины $Y_j = A_0 + A_1 x_{1j} + A_2 x_{2j} + \dots + A_n x_{nj}, j = 1, 2, \dots, m$;

(R2) суммарная мера нечеткости величин Y_j минимальна.

Иногда первое условие заменяют более сильным требованием:

(R1_h) для каждого j выполняется неравенство

$$\mu_{Y_j}(y_j) \geq h,$$

где h – некоторое заданное наперед пороговое значение.

При такой постановке говорят о задаче нечеткой линейной регрессии с пороговым значением h .

Если речь идет о поиске коэффициентов в виде симметричных треугольных нечетких чисел, задача нечеткой линейной регрессии сводится к задаче линейного программирования.

Будем искать коэффициенты A_j в виде:

$$A_i = \langle a_i - d_i, a_i, a_i + d_i \rangle.$$

Тогда Y_j имеет следующий вид:

$$Y_j = \langle z_j - r_j, z_j, z_j + r_j \rangle,$$

где

$$z_j = a_0 + \sum_{i=1}^n a_i x_{ij}, \quad r_j = d_0 + \sum_{i=1}^n d_i |x_{ij}|.$$

Суммарная мера нечеткости вычисляется по формуле

$$r = \sum_{j=1}^m r_j = md_0 + \sum_{j=1}^m \sum_{i=1}^n d_i |x_{ij}|.$$

Задача нечеткой линейной регрессии с ограничениями (R1_h) сводится к следующей задаче линейного программирования:

$$r = \sum_{j=1}^m r_j = md_0 + \sum_{j=1}^m \sum_{i=1}^n d_i |x_{ij}| \rightarrow \min;$$

$$y_j \geq \sum_{i=1}^n a_i x_{ij} - (1-h) \sum_{i=1}^n d_i |x_{ij}|, \quad j = 1, 2, \dots, m;$$

$$y_j \leq \sum_{i=1}^n a_i x_{ij} + (1-h) \sum_{i=1}^n d_i |x_{ij}|, \quad j = 1, 2, \dots, m;$$

$$d_i \geq 0, \quad i = 1, 2, \dots, n.$$

При $h = 0$ получается решение задачи нечеткой линейной регрессии с ограничениями (R1).

2. Оценка стоимости коттеджей

В рамках исследования были рассмотрены и проанализированы сведения о стоимости коттеджей в Московской области по Киевскому направлению (по информации строительной компании «Стройсервис», осень 1997 г.). Исходными данными стали 50 наблюдений [9], содержащие следующую информацию:

- **Price** – цена, тыс. долл. (*объясняемая переменная*);
- **Dist** – расстояние от МКАД, км;
- **House** – площадь дома, кв. м;
- **Area** – площадь участка, сотки;
- **Eco** – наличие озера в окрестностях (двуми-переменная, 1 – да, 0 – нет).

2.1. Линейная регрессия

Прежде всего был использован классический подход. Применяя МНК к исходным данным, получим:

$$Price = 8,412 - 0,738 * Dist + 0,175 * House + 3,462 * Area + 42,324 * Eco.$$

При этом коэффициент детерминации оказывается невысоким: $R^2 = 0,632$.

Следовательно, модель в принципе является приемлемой, однако недостаточно хорошо описывает исследуемую зависимость. На рис. 1 представлен график (наблюдения для нагляд-

ности отсортированы по возрастанию значений стоимости).

Мы видим, что линейная регрессия не позволяет моделировать зависимость стоимости коттеджа от рассматриваемых параметров с достаточной точностью. Целью работы стало изучение нечеткой линейной регрессии.

2.2. Нечеткая линейная регрессия

Допустим, что коэффициенты модели являются симметричными треугольными нечеткими числами $A_0, A_1(Dist), A_2(House), A_3(Area), A_4(Eco)$, вида $A_i = (a_i - p_i; a_i; a_i + p_i)$.

Регрессия в этом случае принимает вид:

$$Y = A_0 + A_1 * Dist + A_2 * House + A_3 * Area + A_4 * Eco.$$

При этом стоимость коттеджа также станет нечеткой величиной:

$$Y = (y_{down}; y_{mod}; y_{up}),$$

где

$$y_{down} = a_0 - p_0 + x_1(a_1 - p_1) + x_2(a_2 - p_2) + x_3(a_3 - p_3) + x_4(a_4 - p_4);$$

$$y_{mod} = a_0 + x_1(a_1) + x_2(a_2) + x_3(a_3) + x_4(a_4);$$

$$y_{up} = a_0 + p_0 + x_1(a_1 + p_1) + x_2(a_2 + p_2) + x_3(a_3 + p_3) + x_4(a_4 + p_4).$$

По-другому границы стоимости можно записать, выделив отклонения от модального значения, а именно:

$$y_{down} = a_0 + \sum_{i=1}^4 a_i x_i - \left[p_0 + \sum_{i=1}^4 p_i x_i \right];$$

$$y_{mod} = a_0 + \sum_{i=1}^4 a_i x_i;$$

$$y_{up} = a_0 + \sum_{i=1}^4 a_i x_i + \left[p_0 + \sum_{i=1}^4 p_i x_i \right].$$

Для каждого из 50 наблюдений определена нечеткая величина:

$$y_j = (y_{j\ down}; y_{j\ mod}; y_{j\ up}),$$

а также величина отклонения от модального значения:

$$p_j = p_0 + \sum_{i=1}^4 p_i x_{ij}.$$

Если просуммировать все такие отклонения, то можно определить суммарную нечеткость регрессии и получить тем самым задачу линейного программирования, в которой минимизации подвергается следующая величина:

$$p = \sum_{j=1}^{50} p_j = \sum_{j=1}^{50} p_0 + \sum_{i=1}^4 p_i x_{ij} = 50 p_0 + \sum_{j=1}^{50} \sum_{i=1}^4 p_i x_{ij} \rightarrow \min.$$

Наконец, для каждой величины Y_j должно выполняться условие:

$$\mu_{Y_j}(y_j) \geq h.$$



Рис. 1. Стоимость коттеджей (тыс. долл. США)

Это сводится к следующим ограничениям:

$$y_j \geq a_0 + \sum_{i=1}^4 a_i x_{ij} - (1-h) \left[p_0 + \sum_{i=1}^4 p_i x_{ij} \right];$$

$$y_j \leq a_0 + \sum_{i=1}^4 a_i x_{ij} + (1-h) \left[p_0 + \sum_{i=1}^4 p_i x_{ij} \right];$$

$$p_i \geq 0;$$

$$j = 1, \dots, m; \quad i = 1, \dots, n.$$

Или в соответствии с принятыми ранее обозначениями:

$$y_j \geq y_{j \text{ down}}; \quad y_j \leq y_{j \text{ up}},$$

Подставляя имеющиеся значения наблюдений, получаем целевую функцию:

$$p = 45p_0 + 2077,5p_1 + 8002p_2 + 603,5p_3 + 22p_4 \rightarrow \min.$$

Из 50 наблюдений 5 отнесены к контрольной выборке, остальные 45 – к объясняющей.

В качестве примера приведем ограничения для первых трех наблюдений:

$$300 \geq a_0 + 20a_1 + 400a_2 + 22a_3 + a_4 - (1-h)[p_0 + 20p_1 + 400p_2 + 22p_3 + p_4];$$

$$300 \leq a_0 + 20a_1 + 400a_2 + 22a_3 + a_4 + (1-h)[p_0 + 20p_1 + 400p_2 + 22p_3 + p_4];$$

$$60 \geq a_0 + 18a_1 + 170a_2 + 6a_3 - (1-h)[p_0 + 18p_1 + 170p_2 + 6p_3];$$

$$60 \leq a_0 + 18a_1 + 170a_2 + 6a_3 + (1-h)[p_0 + 18p_1 + 170p_2 + 6p_3];$$

$$14 \geq a_0 + 90a_1 + 60a_2 + 11a_3 + a_4 - (1-h)[p_0 + 90p_1 + 60p_2 + 11p_3 + p_4];$$

$$14 \leq a_0 + 90a_1 + 60a_2 + 11a_3 + a_4 + (1-h)[p_0 + 90p_1 + 60p_2 + 11p_3 + p_4].$$

Учитывая, что $p_j \geq 0$ для всех j , получаем задачу линейного программирования, содержащую 10 неизвестных и 95 ограничений.

Дальнейшие расчеты были произведены с использованием MS Excel. В зависимости от заданного значения h найдены различные решения: чем больше h , тем большую уверенность мы хотим получить, а следовательно, значение нечет-

кой объясняемой переменной регрессии будет более размытым (см. табл. 1).

Таблица 1

Оценка нечетких коэффициентов

<i>h</i>	0,2	0,5	0,75
<i>p</i>	3949,77	6319,62	12639,23
<i>a</i> ₀	-14,23	-14,27	-14,27
<i>a</i> ₁	-0,0484	-0,0484	-0,0483
<i>a</i> ₂	0,3130	0,3129	0,1329
<i>a</i> ₃	1,083	1,088	1,088
<i>a</i> ₄	90,894	90,883	90,883
<i>p</i> ₀	0	0	0
<i>p</i> ₁	0	0	0
<i>p</i> ₂	0,204	0,327	0,653
<i>p</i> ₃	0	0	0
<i>p</i> ₄	105,25	168,45	336,90

Приведем формулы для вычисления цены:

$h = 0,2$:

$$Price = -14,23 - 0,048 * Dist + < 0,109; 0,313; 0,517 > * House + 1,083 * Area + < -14; 91; 196 > * Eco;$$

$h = 0,5$:

$$Price = -14,27 - 0,048 * Dist + < -0,014; 0,313; 0,640 > * House + 1,088 * Area + < -78; 91; 259 > * Eco;$$

$h = 0,75$:

$$Price = -14,27 - 0,048 * Dist + < -0,34; 0,313; 0,966 > * House + 1,088 * Area + < -246; 91; 428 > * Eco.$$

Наиболее весомый вклад в регрессию при изменении порогового значения h вносит коэффициент при переменной *Eco*, которая является фиктивной и принимает значения 0 и 1. Приведем результаты расчетов, которые получаются, если считать, что коэффициент при *Eco* не является нечеткой величиной (зададим $p_4 = 0$) (см. табл. 2).

$h = 0,2$:

$$Price = -22,55 - 0,412 * Dist + < 0,016; 0,545; 1,074 > * House + 3,411 * Area + 1,2 * Eco;$$

Оценка нечетких коэффициентов, $p_4 = 0$

h	0,2	0,5	0,75
p	4231,99	6771,18	13542,36
a_0	-22,55	-22,55	-22,55
a_1	-0,412	-0,412	-0,412
a_2	0,545	0,545	0,545
a_3	3,411	3,411	3,411
a_4	1,2	1,2	1,2
p_0	0	0	0
p_1	0	0	0
p_2	0,529	0,846	1,629
p_3	0	0	0
p_4	0	0	0

$h = 0,5$:

$$Price = -22,55 - 0,412 * Dist + < 0,301; 0,545; 1,391 > * House + 3,411 * Area + 1,2 * Eco;$$

$h = 0,75$:

$$Price = -22,55 - 0,412 * Dist + < -1,148; 0,545; 2,237 > * House + 3,411 * Area + 1,2 * Eco.$$

Наконец, проанализируем полученные варианты нечеткой линейной регрессии на графиках. Будем рассматривать $h = 0,5$.

В том случае, когда все коэффициенты предполагаются нечеткими числами, имеем (см. рис. 2):

$$Pr1 = -14,27 - 0,048 * Dist + < -0,014; 0,313; 0,640 > * House + 1,088 * Area + < -78; 91; 259 > * Eco.$$

Если же ввести ограничение на коэффициент при фиктивной переменной – сделать его заведомо четким, то (см. рис. 3):

$$Pr2 = -22,55 - 0,412 * Dist + < -0,301; 0,545; 1,391 > * House + 3,411 * Area + 1,2 * Eco.$$

Можно заметить, что вторая модель регрессии лучше описывает изучаемую зависимость. Во-первых, мы видим это на представленных графиках. Во-вторых, интуитивно неверно придавать столь большой вес переменной, принимающей значения 0 и 1, ведь в результате оптимизации нечеткий коэффициент при *Eco* получил значения в диапазоне от -90 до 260 (для $h = 0,5$), в то время как максимальная цена в наблюдениях оставила всего 320.

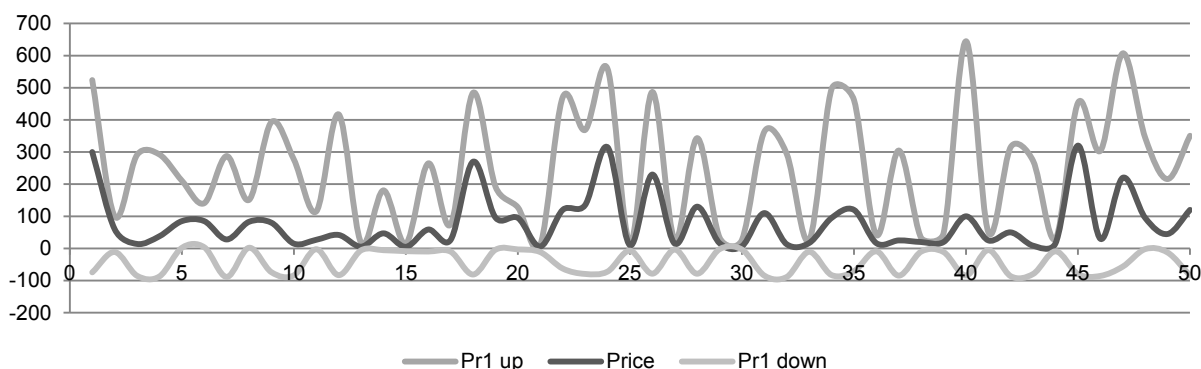


Рис. 2. Нечеткая цена Pr1

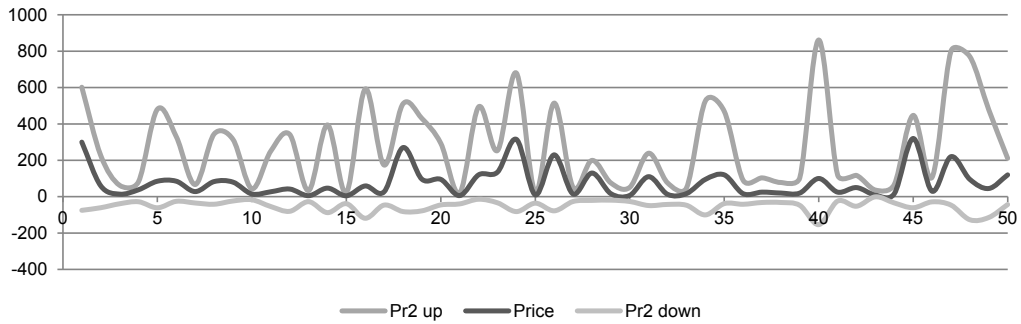


Рис. 3. Нечеткая цена *Pr2*

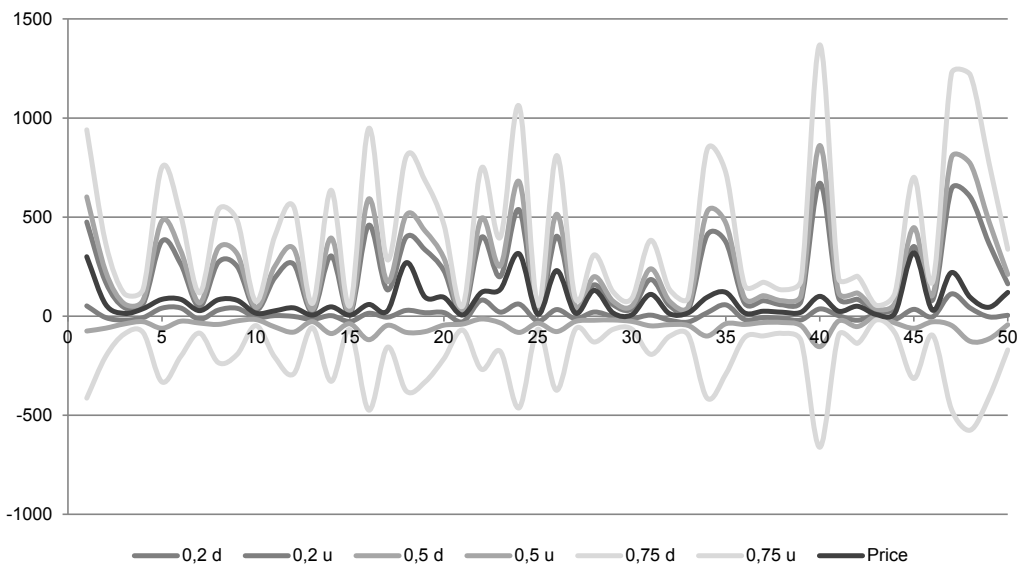


Рис. 4. Нечеткая цена *Pr2* для различных пороговых значений уверенности

Также стоит отметить, что, учитывая содержание задачи, нижняя граница не имеет содержательного смысла, поскольку цена не может быть отрицательной.

В заключение продемонстрируем, как изменение порога h влияет на величину коридора для *Price*. Приведены результаты вычислений для функции *Pr2* (см. рис. 4).

Как было отмечено ранее, желание повысить уверенность в том, что результат оценки стоимости окажется верным, приводит к ситуации, когда получаемая нечеткая оценка становится чрезмерно размытой. Адекватными и имеющими практическую ценность можно считать расчеты для $h \leq 0,5$, но в любом случае многое зависит от того, какие цели преследует лицо, принимающее решение.

2.3. Выводы

Основной задачей исследования стало изучение зависимости стоимости коттеджей в

Московской области от ряда объясняющих переменных.

Коэффициенты, полученные при построении линейной регрессии, позволили сделать весьма естественные выводы о том, что стоимость находится в обратной зависимости от расстояния от МКАД и в прямой зависимости от площадей дома и участка, а также наличия озера в окрестностях. При этом мы выяснили, что в целом линейная регрессия недостаточно точно описывает наблюдаемую зависимость, однако коэффициент детерминации является приемлемым (0,63).

Затем была построена и проанализирована нечеткая линейная регрессия. Предположение о том, что все коэффициенты могут быть нечеткими величинами, привело к ситуации, в которой чрезмерно большая часть стоимости объяснялась нечеткой фиктивной переменной – нали-

чием озера. Такой результат не отражает наших естественных представлений о том, что должно в первую очередь лежать в основе стоимости загородной недвижимости. Поэтому было принято решение сделать эту переменную четкой. В полученной нечеткой регрессии она по-прежнему имела немалый вес, однако не достигала неоправданно больших значений, как при введении условия о возможном отклонении.

Наконец, отметим, что в результате построения нечеткой линейной регрессии только один из коэффициентов при объясняющих переменных оказался нечетким – площадь дома. Остальные коэффициенты приняли четкие значения.

3. Оценка стоимости квартир в Москве

Рассмотрим построение нечеткой регрессии при моделировании цены на квартиры. В *таблице 3* приведена часть наблюдений, используемых в качестве обучающей выборки.

Для рассматриваемой задачи нечеткая регрессия примет вид:

$$\left\{ \begin{array}{l} P_j = A_0 + A_1 S_j + A_2 K_j + \dots + A_7 U_j \\ A_i = \langle a_i - d_i, a_i, a_i + d_i \rangle \\ P_j = \langle z_j - r_j, z_j, z_j + r_j \rangle; z_j = a_0 + a_1 S_j + \dots + a_7 U_j; r_j = d_0 + d_1 S_j + \dots + d_7 U_j \end{array} \right.$$

Оценки, полученные в результате решения этой задачи, приведены в *таблице 4*.

Проверим нашу нечеткую регрессию на адекватность. Для этого сравним цены на квартиры, рассчитанные при помощи регрессии с наблюдаемыми значениями из контролирующей выборки (*табл. 5*).

Нетрудно видеть, что фактическая цена попадает в интервал, образованный нечеткой оценкой цены на квартиру для обоих уравнений наблюдений из контролирующей выборки. При том центр нечеткой оценки цены лежит достаточно близко к ее фактическому значению. Поэтому будем считать, что наша нечеткая регрессия адекватно описывает процесс ценообразования на квартиры.

Сравним полученные результаты с линейной регрессией, оцененной по методу наименьших квадратов. В *табл. 6* представлена сравнительная характеристика значений коэффициентов для нечеткой и классической линейной регрессии.

В *табл. 7* представлена сравнительная характеристика оценок цены для контролирующей выборки в случае нечеткой и линейной регрессии. Заметим, что оценки, полученные методом построения нечеткой регрессии, точнее. Более того, для второго уравнения наблюдения из контролирующей выборки фактическая цена вообще не попала в прогнозный интервал. Таким образом, делаем вывод, что нечеткая регрессия в рассмотренной задаче адекватнее описывает ценообразование на квартиры.

Разумеется, при решении задачи не стоит ограничиваться только лишь применением метода нечеткой регрессии или метода построения линейной регрессии. Лучшим выбором будет комбинирование двух подходов с целью достижения оптимального результата, ведь невозможно сразу определить природу изучаемого явления.

Таблица 3

Фрагмент обучающей выборки

Цена	Площадь	Площадь кухни	Расстояние до метро	Этаж: 1 – если первый или последний, иначе – 0	Дом: 1 – кирпичный, 0 – нет	Телефон: 1 – есть, 0 – нет	Лифт: 1 – есть, 0 – нет
<i>P</i>	<i>S</i>	<i>K</i>	<i>D</i>	<i>F</i>	<i>C</i>	<i>T</i>	<i>U</i>
31	37,6	6	45	1	0	1	1
31	41	5,7	30	1	1	1	0
31	43,3	6,8	7	1	0	1	0
31	44	10	60	0	0	1	1

Нечеткие коэффициенты линейной регрессии

Центры коэффициентов a_i		Отклонение нечетких коэффициентов d_i	
a_0	21,89	d_0	-4,06
a_1	0,25	d_1	0,17
a_2	0,32	d_2	-0,43
a_3	-0,13	d_3	0,00
a_4	-0,75	d_4	1,10
a_5	4,29	d_5	0,51
a_6	1,22	d_6	2,13
a_7	1,23	d_7	-1,24

Таблица 5

Нечеткая цена

Фактическая цена	Центр нечеткой оценки цены	Отклонение нечеткой границы цены
36	35,42	2,72
32	33,62	3,64

4. Оценка стоимости колготок

Целью исследования является построение нечетких моделей линейной регрессии цены колготок в московских оптовых фирмах, а также сравнение полученных результатов с прогнозными ценами, полученными при помощи моделей линейной регрессии, построенных классическими эконометрическими методами. Описание переменных модели представлено в табл. 8.

Линейную регрессию будем искать в виде

$$Y = A_0 + A_1x_1 + A_2x_2 + A_3x_3 + A_4x_4 + A_5x_5 + A_6x_6,$$

где

$x_1 = \text{DEN}$,

$x_2 = \text{firm}$,

$x_3 = \text{polyamid}$,

$x_4 = \text{lykra}$,

$x_5 = \text{cotton}$,

$x_6 = \text{wool}$.

Коэффициенты мы ищем как треугольные нечеткие числа вида

$$A_i = (a_i - d_i, a_i, a_i + d_i),$$

где a_i – центральное (модальное) значение, а d_i – интервал неопределенности. Значения результирующей переменной должны попасть на

носитель нечеткого числа Y , при этом неопределенность должна быть минимальной.

Дополнительно будем предполагать, что правая и левая часть уравнения регрессии неотрицательны.

Для решения задачи будем применять пакет Excel и встроенную в него процедуру поиска решения. В качестве исходных будем использовать данные о ценах колготок в московских оптовых фирмах осенью 1997 г. (выборка в объеме 74 наблюдения) [9, с. 147].

Наблюдения №№ 70–74 будем использовать в качестве контролирующей выборки, процедуру построения модели будем проводить дважды: отдельно для наблюдений №№ 1–30 (набор А) и для наблюдений №№ 40–69 (набор Б).

1. Строим регрессию, которая содержит все имеющиеся объясняющие переменные $x_1 - x_6$ для каждого их двух наборов данных.

А). Решение задачи выглядит следующим образом (см. табл. 9).

Видим, что свободный член a_0 принимает отрицательное значение, велики коэффициенты при x_2 (название фирмы) и при x_3 (содержание хлопка).

Что касается контролирующей выборки, заметим, что из пяти наблюдений только одно (№ 2) не попадает в полученный интервал, однако для данного наблюдения расстояние до ближайшего конца интервала достаточно велико (см. табл. 10).

Таблица 6

Сравнение коэффициентов

Центры коэффициентов a_i в нечеткой регрессии		Значения коэффициентов a_i в линейной регрессии	
a_0	21,89	b_0	17,27
a_1	0,25	b_1	0,33
a_2	0,32	b_2	0,41
a_3	-0,13	b_3	-0,07
a_4	-0,75	b_4	-0,42
a_5	4,29	b_5	3,15
a_6	1,22	b_6	2,16
a_7	1,23	b_7	1,71

Таблица 7

Сравнение оценок

Фактическая цена	Центр нечеткой оценки цены	Отклонение нечеткой границы цены	МНК – оценка цены	Стандартное отклонение
36	35,42	2,72	37,10	2,05
32	33,62	3,64	34,93	2,05

Таблица 8

Переменные (факторы)

Переменная	Описание
N	Номер по порядку
Price	Цена колготок в рублях, 1997 г.
DEN	Плотность в DEN
firm	Фирма-производитель, 0 – Levante, 1 – Golden Lady
polyamid	Содержание полиамида
lykra	Содержание лайкры
cotton	% хлопка
wool	% шерсти

Б). Решение задачи выглядит следующим образом (см. табл. 11).

Из табл. 11 видно, что свободный член a_0 принимает сравнительно большое (причем отрицательное) значение, также велик коэффициент при переменной x_2 (название фирмы). Коэффициент при переменной x_5 (содержание хлопка) для данной выборки положительный.

В этом случае контролирующая выборка полностью попадает в полученные интервалы (табл. 12).

Решим данную задачу с помощью четкого метода эконометрического моделирования.

Спецификация модели стоимости готового изделия имеет вид:

$$Y = A_0 + A_1x_1 + A_2x_2 + A_3x_3 + A_4x_4 + A_5x_5 + A_6x_6 + u_i,$$

где влияние неучтенных факторов на стоимость готового изделия отражено включением в поведенческое уравнение случайного остатка u_i .

Применим метод наименьших квадратов для оценки параметров регрессионной модели. Напомним, что из 74 наборов данных первые

Таблица 9

Нечеткие коэффициенты (1.А)

A		D	
a_0	-269.31	d_0	—
a_1	255.92	d_1	155.92
a_2	1 833.33	d_2	—
a_3	58.30	d_3	49.94
a_4	375.03	d_4	—
a_5	1 841.69	d_5	—
a_6	271.69	d_6	—

Таблица 10

Оценка результатов (1.А)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Delta	Принадлежность Y контролирующему интервалу
1.	17 000.00	13 366.67	33 431.34	20 064.68	+
2.	31 800.00	132 666.67	166 547.83	33 881.16	—
3.	5 100.00	3 900.00	18 565.71	14 665.71	+
4.	19 100.00	11 600.00	42 318.44	30 718.44	+
5.	23 300.00	16 766.67	33 712.90	16 946.24	+

Таблица 11

Нечеткие коэффициенты (1.Б)

A		D	
a_0	-36 061.87	d_0	—
a_1	315.47	d_1	168.54
a_2	4 404.65	d_2	—
a_3	387.08	d_3	43.89
a_4	680.98	d_4	—
a_5	432.48	d_5	—
a_6	252.35	d_6	—

69 составляют обучающую выборку, последние 5 – контролирующую.

После оценки модель имеет следующий вид:

что из численных значений параметров, входящих в состав регрессионной модели, значимым является только значение *DEN* в готовом изделии. Результаты вычислений представлены в *табл. 13*.

$$\left\{ \begin{array}{l} Y = 8\,542 + 181,5 x_1 + 1\,460 x_2 - 37,1 x_3 + 262,4 x_4 - 24,5 x_5 + 202,9 x_6 + u_i \\ (66\,133) \quad (42,9) \quad (1\,519) \quad (670) \quad (605) \quad (657) \quad (1\,292) \\ R^2 = 0,46 \end{array} \right. .$$

Процедура проверки значимости предопределенных переменных с помощью Т-теста показала,

Проверим модель на адекватность. Для этого используем контролирующую выборку. Интер-

Таблица 12

Оценка результатов (1.Б)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Delta	Принадлежность Y контролирующему интервалу
1.	17 000.00	12 240.93	32 395.53	20 154.60	+
2.	31 800.00	21 455.99	57 533.52	36 077.53	+
3.	5 100.00	4 865.35	18 699.96	13 834.61	+
4.	19 100.00	12 257.56	43 665.56	31 408.00	+
5.	23 300.00	10 026.13	26 809.99	16 783.86	+

Таблица 13

Исследование результатов классической линейной регрессии (1)

t_6	t_5	t_4	t_3	t_2	t_1	t_0
0,16	0,04	0,43	0,06	0,96	4,21	0,13
Уровень значимости		0,05	$t_{кр} < t_1 = >$		x_1 значащая	
Кол-во степеней свободы		62				
$t_{кр}$		1,99897				

вальный прогноз стоимости готового изделия представлен в *табл. 14*.

Ошибки прогноза достаточно велики. Следовательно, модель не может быть использована для адекватного прогнозирования. Ни в случае нечеткой линейной регрессии, ни в случае применения метода наименьших квадратов адекватную модель получить не удалось.

2. Удалим переменную x_2 (название фирмы), поскольку коэффициент (в случае линейной нечеткой регрессии) при ней получается слиш-

ком велик, и x_6 (содержание хлопка), поскольку выборка содержит подавляющее количество нулевых данных по этой переменной.

А). Решение задачи приведено в *табл. 15*.

Видно, что коэффициенты $a_3 - a_5$ принимают большие (по модулю) значения, причем два из них отрицательные. Коэффициент a_0 в несколько раз превышает остальные коэффициенты.

В контролирующей выборке одно наблюдение не попало в прогнозный интервал (*табл. 16*).

Б). Решение задачи приведено в *табл. 17*.

Таблица 14

Интервальный прогноз (1)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Принадлежность Y контролирующему интервалу
1	17 000	18 278	20 202	+
2	31 800	24 800	27 410	-
3	5 100	8 547	9 446	-
4	19 100	19 725	21 801	-
5	23 300	15 745	17 402	-

Таблица 15

Нечеткие коэффициенты (2.A)

A		D	
a_0	200 054.84	d_0	–
a_1	241.81	d_1	147.08
a_3	–1 912.90	d_3	61.02
a_4	–1 918.69	d_4	66.81
a_5	1 330.29	d_5	–

Таблица 16

Оценка результатов (2.A)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Delta	Принадлежность Y контролирующему интервалу
1.	17 000.00	6 174.74	30 422.22	24 247.49	+
2.	31 800.00	226 817.89	260 596.96	33 779.06	–
3.	5 100.00	4 084.21	20 700.00	16 615.79	+
4.	19 100.00	9 167.37	42 088.89	32 921.52	+
5.	23 300.00	15 174.74	36 079.86	20 905.12	+

Таблица 17

Нечеткие коэффициенты (2.Б)

A		D	
a_0	20 925.71	d_0	–0.00
a_1	282.62	d_1	69.62
a_3	–131.48	d_3	64.37
a_4	–25.13	d_4	108.22
a_5	–169.61	d_5	–

Видим, что свободный член a_0 принимает большое значение, а коэффициенты $a_3 – a_5$ отрицательны для данного набора данных.

Контролирующая выборка полностью попала в прогнозные интервалы (табл. 18).

Линейная регрессия, полученная с помощью метода эконометрического моделирования, выглядит следующим образом:

Процедура проверки предопределенных переменных с помощью Т-теста снова показала, что значимым является только показатель DEN в готовом изделии. Результаты вычислений представлены в табл. 19.

Интервальный прогноз стоимости готового изделия для данной модели приведен в табл. 20.

$$\left\{ \begin{array}{l} Y = 19761 + 172,6x_1 - 140,5x_3 + 262,4x_4 - 24,5x_5 + u, \\ (9951) \quad (41,3) \quad (91,7) \quad (124,2) \quad (116). \\ R^2 = 0,449 \end{array} \right.$$

Для второго набора переменных ни один из способов не годится для построения адекватной модели прогнозирования цены готового изделия.

3. Процедура проверки значимости предопределенных переменных с помощью Т-теста в модели линейной регрессии, полученной с помощью метода наименьших квадратов, показала, что из числовых величин регрессионной модели значимой является только показатель DEN в готовом изделии. Учитывая это, построим нечеткую регрессионную модель, используя только свободный член и переменную x_1 (показатель DEN).

А). Решение задачи выглядит следующим образом (табл. 21):

Видим, что свободный член принимает достаточно большое значение по сравнению с коэффициентом a_1 , контролирующая выборка полностью попадает в наши интервалы (табл. 22).

Б). Решение задачи приведено в табл. 23.

Видим, что свободный член снова получился большим.

Контролирующая выборка полностью попадает в интервалы (табл. 24).

Таблица 18

Оценка результатов (2.Б)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Delta	Принадлежность Y контролирующему интервалу
1.	17 000.00	11 360.00	31 909.10	20 549.10	+
2.	31 800.00	24 846.25	43 977.62	19 131.37	+
3.	5 100.00	4 535.00	19 498.39	14 963.39	+
4.	19 100.00	16 937.50	40 523.56	23 586.06	+
5.	23 300.00	9 121.25	27 628.65	18 507.40	+

Таблица 19

Исследование результатов классической линейной регрессии (2)

t_5	t_4	t_3	t_1	t_0	
1,06	1,32	1,53	4,18	1,99	
Уровень значимости		0,05	$t_{кр} < t_1 = >$		x_1 значащая
Кол-во степеней свободы		64			
$t_{кр}$		1,99773			

Таблица 20

Интервальный прогноз (2)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Принадлежность Y контролирующему интервалу
1.	17 000	18 918.7	20 910.1	—
2.	31 800	25 222.1	27 877.1	—
3.	5 100	78 83.84	8 713.72	—
4.	19 100	20 082.8	22 196.8	—
5.	23 300	16 461.8	18 194.7	—

Нечеткие коэффициенты (3.A)

A		D	
a_0	18 350.00	d_0	12 350.00
a_1	25.00	d_1	100.00

Таблица 22

Оценка результатов (3.A)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Delta	Принадлежность Y контролирующему интервалу
1.	17 000.00	3 000.00	35 700.00	32 700.00	+
2.	31 800.00	-1 500.00	43 200.00	44 700.00	+
3.	5 100.00	4 875.00	32 575.00	27 700.00	+
4.	19 100.00	750.00	39 450.00	38 700.00	+
5.	23 300.00	3 750.00	34 450.00	30 700.00	+

Таблица 23

Нечеткие коэффициенты (3.Б)

A		D	
a_0	14 250.00	d_0	11 150.00
a_1	125.00	d_1	-

Линейная регрессия, полученная эконометрическим методом:

$$\begin{cases} Y = 10\,891 + 154,7 x_1 + u_t \\ (1472) \quad (40,1) \\ R^2 = 0,18 \end{cases}$$

T-тест указывает на значимость переменных. Однако следует отметить низкое значение коэффициента детерминации для этой модели.

Интервальный прогноз стоимости готового изделия для третьей модели представлен в табл. 25.

Для набора переменных, в который входит показатель DEN и свободный член, модель, построенная методом нечеткой регрессии, является адекватной для прогнозирования.

4. Построим нечеткую регрессионную модель,

в составе которой только переменная x_1 (показатель DEN).

При учете всех необходимых ограничений на коэффициенты процедура «поиск решения» не привела к решению задачи ни для набора (A), ни для набора (Б).

В случае обыкновенной линейной регрессии модель принимает вид:

$$\begin{cases} Y = 401,4 x_1 + u_t \\ (30) \\ R^2 = 0,72 \end{cases}$$

T-тест указал на значимость переменной, входящей в модель.

Для четвертого набора переменных ни в случае линейной нечеткой регрессии, ни в случае применения метода наименьших квадратов адекватную модель получить не удалось.

Таблица 24

Оценка результатов (3.Б)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Delta	Принадлежность Y контролирующему интервалу
1.	17 000.00	8 100.00	30 400.00	22 300.00	+
2.	31 800.00	15 600.00	37 900.00	22 300.00	+
3.	5 100.00	4 975.00	27 275.00	22 300.00	+
4.	19 100.00	11 850.00	34 150.00	22 300.00	+
5.	23 300.00	6 850.00	29 150.00	22 300.00	+

Таблица 25

Интервальный прогноз (3)

№ п/п	Y контролирующее	Ограничение слева	Ограничение справа	Принадлежность Y контролирующему интервалу
1.	17 000	16 224	17 932	+
2.	31 800	25 041	27 677	–
3.	5 100	12 551	13 872	–
4.	19 100	20 633	22 050	–
5.	23 100	14 055	16 008	–

Заключение

В результате исследования построены нечеткие модели линейной регрессии и проведено сравнение полученных результатов с прогнозными ценами, полученными при помощи моделей линейной регрессии, построенных классическими эконометрическими методами. Исследование показало, что оба метода приводят к согласованным в определенной степени результатам. В то же время применение нечеткой линейной регрессии должно сопровождаться экономическим анализом получаемых результатов, без которого модель может потерять адекватность.

Литература

1. Михалькевич И.С. Повышение достоверности слабоформализованных данных // Научные записки молодых исследователей. 2014. № 2. С. 17–21.
2. Волкова Е.С., Гисин В.Б. Меры возможности и внутренняя норма доходности инвестиционных проектов с нечетко определенными платежами // Вестник Финансового университета. 2014. № 3 (81). С. 93–104.
3. Волкова Е.С., Гисин В.Б. Цена информации в задачах оценки инвестиционных проектов с нечетко определенными денежными потоками // Экономические науки. № 10 (107), 2013. С. 105–112.
4. Dymowa L. Soft computing in economics and finance. Springer, 2011. С. 295.
5. Волкова Е.С., Гисин В.Б. Нечеткая линейная регрессия в оценке недвижимости // Вопросы оценки. 2015. № 1 (79). С. 26–33.
6. Волкова Е.С., Гисин В.Б. Нечеткая линейная регрессия в модели роста технологических знаний // Вестник Финансового университета. № 5 (89), 2015. С. 97–104.
7. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб: БХВ-Петербург, 2005. 736 с.
8. Волкова Е.С., Гисин В.Б. Нечеткие множества и мягкие вычисления в экономике и финансах. М.: Финансовый университет, 2015. 184 с.
9. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. М.: Дело, 2004. 504 с.
10. Грибовский С.В., Сивец С.А. Математические методы оценки стоимости недвижимого имущества. М.: Финансы и статистика, 2008. 368 с.