УДК 514.765

Нечеткое моделирование напряженности на рынке труда[1]

Пономарев И.В.
Алтайский государственный университет
igorpon@mail.ru

Аннотация

В статье рассмотрена одна из возможных моделей нечеткого регрессионного анализа рынка труда. Проводится исследование имеющейся выборки на предмет выбросов.

1. Модель нечеткой регрессии

Пусть R^m – m-мерное арифметическое евклидово пространство. Пусть Ω конечное подмножество точек:

$$\Omega = \{(x_{i1}, x_{i2}, \dots, x_{im}) : i = 1, \dots, n\},$$

которое можно рассматривать как результат n экспериментов. В приложениях часто возникает вопрос о существовании функциональной зависимости между переменными $\{x_1, x_2, \cdots, x_m\}$. Наиболее исследованный вид зависимости – линейный. В статистике разработаны мощные методы для анализа множества Ω на линейную зависимость основанные на Евклидовой норме [1–5]. В данной работе в качестве основы берется Чебышевская норма равномерного отклонения [6, 7].

Определение 1. *Минимальной шириной множества Ω вдоль переменной x_j, $j = 1, \dots, m$ назовем число*

$$\alpha_\infty(\Omega, x_j) = 2 \cdot \min_{k_s, s\neq j\,; b} \left\{ \max_{i=1,\dots,n} |x_{i,j} - \sum_{s\neq j}^{m} k_s x_{i,s} - b| \right\}. \tag{1}$$

С геометрической точки зрения величина $\alpha_\infty(\Omega, x_j)$ равна минимуму ширины "полосы" ограниченной двумя параллельными гиперплоскостями и содержащей множество Ω, ширина берется вдоль оси x_j в R^m (т.е. длина пересечения полосы с осью x_j).

Уравнение гиперплоскости на котором достигается (1) назовем уравнением L_∞ регрессии на переменную x_j:

$$x_j = \sum_{s\neq j}^{m} k_s^0 x_s - b^0, \tag{2}$$

или уравнением регрессии на переменную x_j относительно Чебышевской нормы.

Замечание 1. *Аналогичные определения справедливы в случае произвольного выпуклого подмножества $\Omega \subset R^m$. Величины $\alpha_\infty(\Omega, x_j)$ тесно связаны с такими понятиями из выпуклой геометрии, как ширина выпуклого множества в данном направлении и широта выпуклого множества.*

[1]Работа выполнена при поддержке РФФИ (гранты: №16–01–00336А, №16–31–00048мол_а), Минобрнауки РФ в рамках базовой части государственного задания в сфере научной деятельности ФГБОУ ВПО «Алтайский государственный университет» (код проекта: 1148).

Определение 2. *Определим коэффициент корреляции $korr_\infty(x,y)$ для L_∞-регрессии формулой*

$$korr_\infty(x,y) = k_\infty \cdot \overline{k}_\infty,$$

где k_∞, $\overline{k}_\infty$ угловые коэффициенты прямых L_∞-регрессий одномерных множеств y на x и x на y соответственно.

Теорема 1. *Справедливо неравенство:*

$$-1 \le korr_\infty(x,y) \le 1.$$

2. Выбросы в L_∞ регрессии

Стандартные методы оценки классической регрессионной модели достаточно чувствительны к наличию в выборочных данных аномальных, сильно отличающихся от остальных наблюдений — выбросов. Присутствие в данных даже одного выброса может в значительной мере испортить модель, сделать ее непригодной к дальнейшему использованию. Проблеме исследования данных на наличие выбросов посвящены работы С. Вайсберга [2], Р. Кука [4], Д. Эндрюса и Д. Прегибона [5].

Сформулируем задачу о выбросах следующим образом: требуется из данного экспериментального множества данных Ω отбросить фиксированный процент данных так, чтобы оставшиеся данные Ω_0 имели наименьшую величину разброса $\alpha_p(\Omega_0)$, т.е.

$$\alpha_p(\Omega_0) = \min\left\{\alpha_\infty(\Omega') : \Omega' \subset \Omega,\ \#\left[\Omega'\right] = n_0\right\}, \tag{3}$$

где $\#\left[\Omega'\right]$ – число элементов в множестве Ω', $n_0 < n$, $n - n_0 = m_0$ – число выбросов.

Реализация данного алгоритма основано на преобразовании Лежандра. В работе [8] определяется и исследуется обобщенное преобразование Лежандра для произвольного конечного подмножества евклидова пространства.

Определение 3. *Пусть Ω – конечное множество точек, и дана пара натуральных чисел $1 \le r, s \le N$. Обозначим через*

$$[MAX_r]\left[\{c_i\}_{i=1}^n\right] = c_{i_{r+1}}, \quad [MIN_s]\left[\{c_i\}_{i=1}^n\right] = c_{i_{n-s}},$$

где $\{c_{i_k}\}_{k=1}^n$ – перестановка последовательности $\{c_i\}_{i=1}^n$ в порядке убывания:

$$c_{i_1} \ge c_{i_2} \ge \ldots \ge c_{i_k} \ge \ldots \ge c_{i_n}.$$

Таким образом:

$$[MAX_0]\left[\{c_i\}_{i=1}^n\right] = \max\left[\{c_i\}_{i=1}^n\right]$$
$$[MIN_0]\left[\{c_i\}_{i=1}^n\right] = \min\left[\{c_i\}_{i=1}^n\right].$$

Назовем обобщенным преобразованием Лежандра множества пару функций:

$$f_r^+\left(k_t^0\right) = [MAX_r]\left\{\sum_{t\ne j}^{m} k_t^0 x_{is} - x_{ij} : i = 1,\ldots,n\right\},$$
$$f_s^-\left(k_t^0\right) = [MIN_s]\left\{\sum_{t\ne j}^{m} k_t^0 x_{is} - x_{ij} : i = 1,\ldots,n\right\}.$$

Следствие 1. *Функции f_r^+, f_s^- определения – выпуклые вниз и вверх соответственно, разность $f_r^+(k_t^0) - f_s^-(k_t^0)$ – неотрицательная функция, выпуклая вниз.*

Теорема 2. *Справедливо равенство*

$$\min\left\{\alpha_\infty(\Omega') : \Omega' \subset \Omega,\ \#\left[\Omega'\right] = n_0\right\} = \min_{k_t^0} \min_{0 \le r \le m_0} \left[f_r^+\left(k_t^0\right) - f_{m_0-r}^-\left(k_t^0\right)\right] \tag{4}$$

3. Модель напряженности на рынке труда

Изучению рынка труда и механизмов его функционирования в современных условиях посвящены работы [9–15].

Для построения нечеткой модели нами были выбраны следующие социально-экономические показатели:
y – уровень безработицы, %.
x_1 – коэффициенты миграционного прироста на 10000 человек населения.
x_2 – валовой региональный продукт, млн. руб.
x_3 – объем бытовых услуг населению, млн. руб.
x_4 – инвестиции в основной капитал, млн. руб.
x_5 – численность населения, тыс. чел.

Составим матрицу R – аналогом корреляционной матрицы, состоящей из элементов $korr_\infty(X,Y)$. Эти коэффициенты будут показывать силу связи между двумя факторами.

$$R = \begin{pmatrix} 1.0 & 0.03284 & 0.1397 & 0.009550 & 0.07163 & 0.01472 \\ 0.03284 & 1.0 & 0.3382 & 0.4329 & 0.5466 & 1.0 \\ 0.1397 & 0.3382 & 1.0 & 0.05248 & 1.0 & 1.0 \\ 0.009550 & 0.4329 & 0.05248 & 1.0 & 0.1810 & -0.001235 \\ 0.07163 & 0.5466 & 1.0 & 0.1810 & 1.0 & 0.1398 \\ 0.01472 & 1.0 & 1.0 & -0.001235 & 0.1398 & 1.0 \end{pmatrix}$$

Анализ матрицы позволяет сделать вывод о том, что между некоторыми регрессорами существует сильная взаимозависимость. Отрицательный знак является следствием не равнозначности переменных, т.е. характер связи (прямая или обратная) зависит от того какая из переменных является зависимой. Например, увеличение численности населения ведет к увеличению объема бытовых услуг, но в тоже время увеличение количества бытовых услуг населению будет сигнализировать об уменьшении численности населения.

Уравнение нечеткой модели для оценки напряженности на рынке труда будет иметь вид

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5.$$

Конкретные значения параметров a_i приведены в таблице 1:

Таблица 1

Коэффициенты регрессионной модели

	Модальное значение	Нижняя граница	Верхняя граница
a_0	16.28	4.19	28.37
a_1	0.02801	0.02801	0.02801
a_2	0.1308	0.1308	0.1308
a_3	- 0.3144	- 0.3144	- 0.3144
a_4	- 0.01568	- 0.01568	- 0.01568
a_5	0.1794	0.1794	0.1794

Анализ данной модели показывает, что предполагаемый нулевой уровень безработицы рынке труда составляет 16,28% и может колебаться в интервале от 4,19% до 28,37%. Из представленных факторов на рост безработицы оказывают воздействие x_1, x_2 и x_5, т.е. с ростом этих факторов напряженность на рынке труда возрастает. С увеличением остальных факторов индекс напряженности будет уменьшаться.

Исследование первичных данных на выбросы позволяет сделать заключение, что наблюдения с номерами 36, 40, 62 могут быть признаны аномальными, так как они влекут за собой максимальное увеличение функционала качества (рисунок 1).

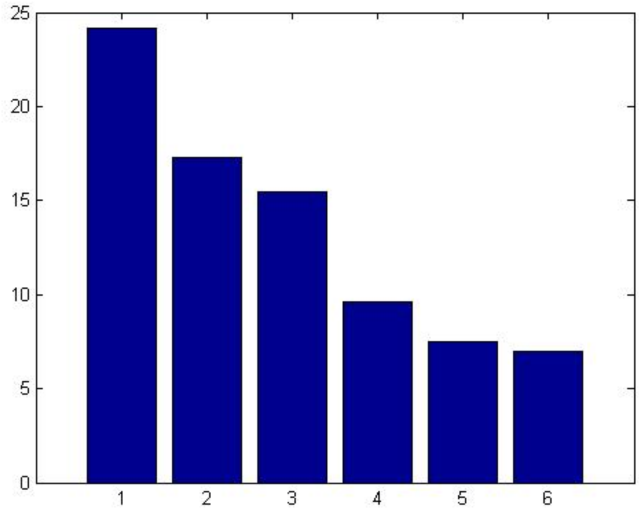

Рисунок 1. Диаграмма изменения функционала качества модели

Если удалить эти наблюдения из первоначальной выборки и провести повторное исследование модели, то коэффициенты полученной модель будет иметь следующие характер

Таблица 2

Коэффициенты регрессионной модели после удаления выбросов

	Модальное значение	Нижняя граница	Верхняя граница
a_0	6.811	3.302	10.32
a_1	0.001913	0.001913	0.001913
a_2	- 0.05902	- 0.05902	- 0.05902
a_3	- 0.1276	- 0.1276	- 0.1276
a_4	- 0.0007034	- 0.0007034	- 0.0007034
a_5	0.6716	0.6716	0.6716

Теперь модель выглядит более реалистичной (после удаления выбросов произошло снижение модального значения безработицы почти на 10%). Так же обратим внимание на то, что фактор x_2 изменил направление своего воздействия на противоположное.

В заключении стоит отметить, что все указанные алгоритмы и процедуры были объединены в программный комплекс, реализованный в программе MatLab. Это позволяет очень быстро изменять конструкцию модели и оперативно реагировать на изменение факторов.

Список литературы

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: в 2-х кн. — 2-е, перераб. и доп. изд. — М. : Финансы и статистика, 1986.

2. Weisberg S. Applied linear regression. — 3^{th} edition. — Jonh Wiley & Sans, Inc., 2005.

3. Green W.H. Econometrics Analysis. — 5^{th} edition. — New Jersey : Prentice-Hall, Upper Saddle River, 1995.

4. Cook R.D. Detection of Influential Observation in Linear Regression // Technometrics. — 1977. — Vol. 19, no. 1. — P. 15–18.

5. Andrews D.F., Pregibon D. Finding the outliers that matter // Journal of the Royal Statistical Society. — 1978. — Vol. 40. — P. 84–93.

6. Ponomarev I.V., Slavsky V.V. Uniformly fuzzy model of linear regression // Journal of Mathematical Sciences. — 2012. — Vol. 186, no. 3. — P. 478–494.

7. Пономарев И.В., Родионов Е.Д., Родионова Л.В., и др. Математическое моделирование объектов науки: монография. — Барнаул : Концепт, 2010. — 160 с.

8. Куркина М.В., Пономарев И.В. Система нечетких отношений равенств в банаховом пространстве // Дифференциальные уравнения. Функциональные пространства. Теория приближений. — Новосибирск : Изд-во Института математики, 2008. — С. 513.

9. Гуров А.В., Пономарев И.В. Моделирование численности вакансий на рынке труда Алтайского края // Известия Алтайского государственного университета. — 2014. — № 1/2. — С. 81–85.

10. Перекаренкова Ю.А., Пономарев И.В., Родионов Е.Д., Родионова Л.В. Региональный рынок труда: анализ, моделирование, прогноз // Вестник Алтайской науки. — 2014. — № 1(19). — С. 57–65.

11. Перекаренкова Ю.А., Родионова Л.В. Социально-трудовые права сельского населения: формальные нормы, реальные практики и перспективы, социологические исследования // Социологические исследования. — 2014. — № 5. — С. 88–96.

12. Пономарев И.В., Родионов Е.Д., Родионова Л.В., и др. Применение пакетов символьных вычислений к решению задач теории и практики: монография. — Барнаул : ИП Колмогоров И.А., 2014.

13. Пономарев И.В. Нечеткие временные ряды и их применение к моделированию социально-экономических процессов // Сборник статей международной конференции “Ломоносовские чтения на Алтае: фундаментальные проблемы науки и образования”. — Барнаул : Изд-во Алт. ун-та, 2014. — С. 511–513.

14. Пономарев И.В., Родионов Е.Д., Родионова Л.В., Славский В.В. Комплекс моделей для построения и оценки вариантов развития регионального рынка труда // Вестник Алтайской науки. — 2013. — № 1. — С. 86–88.

15. Родионова Л.В., Перекаренкова Ю.А. Социально-трудовые аспекты устойчивого развития сельских территорий: монография. — Барнаул : Изд-во Алт. ун-та, 2013.