МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра дифференциальных уравнений и
математической экономики

НЕЧЕТКАЯ РЕГРЕССИОННАЯ МОДЕЛЬ РОСТА ТЕХНОЛОГИЧЕСКИХ ЗНАНИЙ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 451 группы
направления 38.03.05 — Бизнес-информатика

механико-математического факультета
Цветковой Анны Николаевны

Научный руководитель
доцент, к. ф.-м. н., доцент ______________ В. В. Новиков

Заведующий кафедрой
зав.каф., д. ф.-м. н., профессор ______________ С. И. Дудов

Саратов 2022

ВВЕДЕНИЕ

При принятии решений каждый субъект экономической деятельности неизбежно сталкивается с неопределенностью. Данные, собираемые для осуществления дальнейшего их анализа, характеризуются неточностью, приводящей в ряде случаев к серьезному искажению получаемых результатов, поэтому задачи нечеткой линейной регрессии являются актуальными.

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. 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.

С возникновением теории нечетких множеств ее модели и методы получили широкое распространение и нашли применение в самых разных областях, включая эконометрику, экономику, финансы. В рамках этой теории в качестве альтернативы классическому вероятностному подходу был разработан метод оценки параметров моделей, основанный на нечеткой линейной регрессии. Была показана его применимость для анализа как микро-, так и макроэкономических моделей. Применение нечеткой линейной регрессии оказалось оправданным в тех ситуациях, когда нет оснований считать, что ошибки в конкретных наблюдениях подчиняются вероятностным закономерностям. Вскоре метод нечеткой линейной регрессии был усовершенствован и приобрел ту форму, в которой применяется в исследованиях и в настоящее время.

Целью данной работы является применение метода нечеткой линейной регрессии для оценки параметров на реальных данных.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- Изучить теоретическую базу;
- Собрать необходимые данные;

- Провести оценку параметров нечеткой линейной регрессии;
- Проанализировать полученные результаты.

Основное содержание работы

В первом разделе приводится уравнение модели множественной линейной регрессии, метод наименьших квадратов (МНК), теорема Гаусса-Маркова, статистические свойства МНК-оценок и коэффициент детерминации R^2.

Во втором разделе содержатся необходимые сведения о нечетких числах - функция принадлежности нечеткого числа, треугольные нечеткие числа и арифметические операции с ними.

В третьем разделе рассказывается про построение регрессии на нечетких данных, в нем определена в общем случае задача нечеткой линейной регрессии и ее сведение к задаче линейного программирования.

В четвертом разделе рассказывается про модель роста технологических знаний Ботацци-Пери.

В пятом разделе реализовано построение регрессии на нечетких данных.

Необходимые сведения о нечетких числах. Нечеткое подмножество A базового множества X задается своей функцией принадлежности $\mu A : X \to [0; 1]$. Под нечеткой величиной обычно понимают нечеткое подмножество A множества действительных чисел R. Значения функции принадлежности указывают возможность того, что нечеткая величина принимает соответствующее значение.

Нечеткие величины, описываемые выражениями типа «примерно a», обычно представляют так называемыми треугольными нечеткими числами. Треугольное нечеткое число A задается тройкой чисел $(a^L; a; a^R)$, такой что $(a^L \leq a \leq a^R)$

В нечеткой линейной регрессии, как правило, используются симметричные треугольные нечеткие числа, для которых $a^L = a - d$, $a^R = a + d$, $d \geq 0$. Для симметричных треугольных нечетких чисел мы будем использовать обозначение $A = (a; d)$ и называть их для краткости СТН - числами. При $d > 0$ функция принадлежности СТН - числа $A = (a; d)$ имеет вид

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

Если $d = 0$, то $\mu_A(x)$ — характеристическая функция одноточечного множества a:

$$\mu_A(x) = \begin{cases} 0, & \text{если } x \neq a; \\ 1, & \text{если } x = a. \end{cases} \tag{2}$$

В этом случае естественно считать, что $A = a$ — обычное (четкое) число.

Число d называется коэффициентом нечеткости симметричного треугольного нечеткого числа $A = (a; d)$.

Арифметические операции с нечеткими величинами определяются на основе принципа обобщения. Сумма СТН - чисел $(a_1; d_1)$ и $(a_2; d_2)$ — это СТН - число $(a_1 + a_2; d_1 + d_2)$. Произведение СТН - числа $(a; d)$ на скаляр k — это СТН - число $(ka; |k|d)$.

Для сравнения нечетких величин используются различные методы, как правило, основанные на мерах возможности и необходимости. В общем случае возможность отношения $A \leq B$ оценивается числом

$$Pos(A \leq B) = \sup\{\min(\mu_A(x), \mu_B(y))|x \leq y; x, y \in R\}.$$

В частности, если $A = (a^L < a < a^R)$ — треугольное нечеткое число, а $B = b$ четкое число, имеем:

$$Pos(A \leq B) = \begin{cases} 1, & \text{если } a \leq b; \\ \mu_A(b), & \text{если } b \leq a. \end{cases}$$

Отсюда

$$Pos(A = B) = \min(Pos(A \leq b), Pos(A \geq b)) = \mu_A(b).$$

Таким образом, возможность того, что нечеткая величина принимает значение b, оценивается значением функции принадлежности для этого значения.

Далее рассмотрим **задачу нечеткой линейной регрессии и задачу линейного программирования.**

В общем случае задача нечеткой линейной регрессии может быть поставлена следующим образом. Имея m результатов наблюдений (y_j, x_j), $j = 1, \ldots, m$, требуется оптимальным образом определить вектор нечетких коэффициентов $A = (A_0, A_1, \ldots, A_n)$. Оптимальность выражается двумя условиями:

1. (R1) для каждого j число y_j принадлежит носителю нечеткой величины $Y_j = A_0 + A_1 x_{1j} + A_2 x_{2j} + \ldots + A_n x_{nj}$, $j = 1, 2, \ldots, m$;
2. (R2) суммарная мера нечеткости величин Y_j минимальна.

Иногда первое условие заменяют более сильным требованием:

(R1h) для каждого j выполняется неравенство

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

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_i в виде:

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

Тогда Y_j имеет следующий вид:

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

Суммарная мера нечеткости вычисляется по формуле

$$r = \sum_{j=1}^{m} r_j = md_0 + \sum_{j=1}^{m} \sum_{i=1}^{n} d_i |x_{ij}|. \tag{3}$$

Задача нечеткой линейной регрессии с ограничениями (R1h) сводится к следующей задаче линейного программирования:

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

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

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

$$d_i \geq 0, \quad i + 1, 2, \ldots, n.$$

Наиболее часто задача линейной регрессии в такой постановке решается при $h = 0$, т.е. условие $\mu_{Yj}(y_j) \geq h$ фактически заменяется требованием, чтобы y_j принадлежал носителю множества Y_j.

Нечеткая линейная регрессия в модели Ботацци-Пери. Модель Ботацци-Пери — это модель роста технологических знаний.

Для вычислений используются данные за период с 1973 по 1999 г. для каждой страны j из списка: Австралия, Великобритания, Дания, Ирландия, Испания, Италия, Канада, Нидерланды, Норвегия, ФРГ, Финляндия, Франция, Швеция, США, Япония, указаны:

- A_{jt} — запас технологических знаний в стране j к началу года t,
- $R\&D_{jt}$ — число полных ставок в секторе научно-исследовательских и опытно-конструкторских работ (НИОКР) в стране j в год t.

Величина A_{ROWjt} вычисляется по формуле

$$A_{ROWjt} = \sum_{k \neq j} A_{jt}.$$

Положим

$$Y_{jt} = S + M\ln(R\&D_{jt}) + R\ln(A_{ROWjt}).$$

В определении коэффициентов из классических уравнений регрессии присутствует неопределенность, которая связана не только и не столько со случайными ошибками измерений, но и с неполнотой и нечеткостью информации о процессе накопления знаний и его особенностями в отдельно взятых странах. В этой ситуации оценка коэффициентов, как нечетких величин представляется достаточно мотивированной.

Регрессия строится для каждой страны с условием, что для всех j число $\ln A_j$ попадает на носитель СНТ - числа Y_{jt}, и при этом суммарная нечеткость должна быть минимальной. Коэффициенты M, R, S вычисляются как СНТ - числа.

$$10,5258606 \geq a_0 + 12,62424994a_1 + 13,068098303a_2 -$$
$$-(1-h)[d_0 + 12,62424994d_1 + 13,068098303d_2];$$

$$10,5258606 \leq a_0 + 12,62424994a_1 + 13,068098303a_2 +$$
$$+(1-h)[d_0 + 12,62424994d_1 + 13,068098303d_2];$$

$$10,61783025 \geq a_0 + 12,62313689a_1 + 13,10852685a_2 -$$
$$-(1-h)[d_0 + 12,62313689d_1 + 13,10852685d_2];$$

$$10,61783025 \leq a_0 + 12,62313689a_1 + 13,10852685a_2 +$$
$$+(1-h)[d_0 + 12,62313689d_1 + 13,10852685d_2].$$

Учитывая, что $d_j \geq 0$ для всех j, получаем задачу линейного программирования, содержащую 6 неизвестных и 57 ограничений. Порог надежности задается $h = 0$.

Дальнейшие вычисления производятся на языке программирования Python с использованием встроенной библиотеки Scipy. Библиотека включает компоненты для решения дифференциальных уравнений, интегрирования, обработки изображений, визуализации, оптимизации и многих других задач. В частности, для решения задач линейного программирования Scipy предлагает функцию linprog, входящую в компонент optimize.

```
import os
import numpy as np
from openpyxl import load_workbook
from scipy.optimize import linprog
```

Вызов функции linprog представлен в следующей части кода программы,

```
res = linprog(c, A_ub=A, b_ub=b,
              bounds=[a0_bounds, a1_bounds, a2_bounds,
                      d0_bounds, d1_bounds, d2_bounds])
print(res)
```

где c – коэффициенты целевой функции; A_ub – двумерный массив (или список списков) с коэффициентами ограничений верхних границ ($\leq$); b_ub – правая часть ограничений – верхних границ; bounds – ограничения на значения переменных.

При оценке нечеткие коэффициентов динамика роста запаса знаний в большей степени объясняется эффективностью функционирования сектора НИОКР в Германии, Японии, Норвегии и США в остальных странах динамика роста запаса знаний в большей степени объясняется значением межстранового перетока знаний. Если оценивать четкие коэффициенты, то рост запаса знаний в большей степени объясняется эффективностью функционирования сектора НИОКР в Германии, Японии, Норвегии, Дании, Франции, Италии и Швеции, в остальных странах динамика роста запаса знаний в большей степени объясняется значением межстранового перетока знаний.

Если сравнивать два подхода, то можно заметить, что влияние независимых переменных распределено по разному, при этом оценки коэффициентов близки друг к другу. Можно сделать вывод, что лучше использовать комбинирование регрессионных подходов, это позволит получить более точные результаты.

Далее произведем вычисления для набора стран, в котором будет содержаться: Российская Федерация, США, Мексика, Южная Корея, Япония, Китай, Индия за период с 2000 по 2017 года. Регрессия строится аналогичным образом для каждой страны из списка.

Подставляя в формулу целевой функции имеющиеся значения наблюдений, получаем целевую функцию для данных Мексики:

$$r = 18d_0 + 188,7666d_1 + 255,7018d_2.$$

Учитывая все ограничения, получаем задачу линейного программирования, содержащую 6 неизвестных и 39 ограничений. Порог надежности аналогично задается $h = 0$.

Переток знаний между странами оказывает наибольшее влияние на запас знаний в Китае, нежели в других странах. Так же в Китае показатель вклада сотрудников работающих в области НИОКР наибольший. Все это говорит о стремительном развитие и техническом лидерстве Китая. При оценке нечетких коэффициентов переток знаний вносит наибольший вклад в запас знаний Индии, России и Китая, чем сотрудники, работающие в этой отрасли, если анализировать четкие коэффициенты в этот список добавляется Мексика.

Так же можно заметить, что в развитых странах показатель вклада сотрудников в запасы знаний близок к единице нежели в развивающихся странах. В целом результаты практически совпадают, можно сделать аналогичный вывод как и для первого набора данных, что для прогнозирования и наблюдения за данными необходимо применять комбинирование регрессионных подходов, это позволит получить более точные результаты и при этом проанализировать возможную неопределенность в данных.

Благодаря проведенному анализу можно оценить количественное влияние международных знаний на технологическое развитие страны в краткосрочной и долгосрочной перспективе. Научно-технологические знания являются одним из важных факторов экономического развития. В долгосрочной перспективе они влияют на технологический рост. В краткосрочной и среднесрочной перспективе отклонения от нормы в процессе накопления знаний обуславливают быстрый рост производительности, либо технологический спад.

ЗАКЛЮЧЕНИЕ

В работе рассматривались некоторые аспекты регрессионного моделирования, связанные с неопределенностью при задании коэффициентов соответствующей модели. При этом основное внимание уделено методу, при котором неопределенность моделируется не на основе вероятностного подхода, а рассматривается в рамках теории нечетких множеств. По итогам можно сделать вывод, что нечеткая линейная регрессия, примененная для оценки параметров модели роста технологических знаний Ботацци-Пери, позволяет получить результаты, имеющие содержательную экономическую интерпретацию. В практической части работы рассматривалась регрессионное моделирование роста технологических знаний с использованием нечеткой модели на основе реальных данных.