Санкт-Петербургский политехнический университет имени Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Интервальный анализ
Отчёт по курсовой работе

"Нелинейная интервальная регрессия"

Выполнил:

Студент: Аникин Александр
Группа: 3630102/80201

Принял:

к. ф.-м. н., доцент
Баженов Александр Николаевич

2022 г.

Содержание

Список иллюстраций

Список таблиц

1 Постановка задачи

Для набора данных с интервальной неопределённостью построить нелинейную регрессию методом центра неопределённости и методом распознающего функционала. Сравнить результаты.

2 Теория

Пусть величина y является функцией от независимых аргументов $x_1, x_2, ..., x_p$.

$$y = f(x, \beta) \tag{1}$$

где $x = (x_1, ..., x_n)$ - вектор независимых переменных, $\beta = (\beta_1, ..., \beta_k)$ - вектор параметров функции. Имеется набор $x = x_1, ..., x_m, y = y_1, ..., y_m$, необходимо найти оценку вектора параметров β, который соответствует конкретной функции из семейства 1. Будем рассматривать задачу, в которой x - точечный вектор, а y - интервальный вектор.

Будем рассматривать функцию вида:

$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 \tag{2}$$

2.1 Метод центра неопределённости

Для удобства отсортируем все входные параметры по x по возрастанию, и сместим значения так, чтобы $x_0 = 0$. x_i – число, $y_i = [h_{Hi}, h_{Bi}]$ – интервал $\forall i = 1, ..., m$.

Вычисление коэффициента β_0

Рассмотрим тройку наблюдений y_i, y_j, y_k с упорядоченными значениями $x_i < x_j < x_k$. Для каждой тройки индексов $i < j < k$ вычислим допустимые значения коэффициента β_0:

$$\beta_{0H0jk} = h_{H0}, \text{если } i = 0 \tag{3}$$

$$\beta_{0Hijk} = \frac{x_i x_j x_k}{x_j - x_i}\left(\frac{h_{Hi}}{x_i(x_k - x_i)} - \frac{h_{Bj}}{x_j(x_k - x_j)} + \frac{h_{Hk}(x_j - x_i)}{x_k(x_k - x_i)(x_k - x_j)}\right) \tag{4}$$

$$\beta_{0B0jk} = h_{B0}, \text{если } i = 0 \tag{5}$$

$$\beta_{0Bijk} = \frac{x_i x_j x_k}{x_j - x_i}(\frac{h_{Bi}}{x_i(x_k - x_i)} - \frac{h_{Hj}}{x_j(x_k - x_j)} + \frac{h_{Bk}(x_j - x_i)}{x_k(x_k - x_i)(x_k - x_j)}) \quad (6)$$

Далее вычисляем:

$$\beta_{0min} = \max_{ijk}(\beta_{0Hijk}) \quad (7)$$

$$\beta_{0max} = \min_{ijk}(\beta_{0Bijk}) \quad (8)$$

Вычисление коэффициентов β_1
Аналогично вычислению β_0 вычислим допустимые значения коэффициента β_1:

$$\beta_{1Hijk} = \frac{(x_i + x_k)(x_j + x_k)}{x_j - x_i}(-\frac{h_{Bi}}{x_k^2 - x_i^2} + \frac{h_{Hj}}{(x_k^2 - x_j^2)} - \frac{h_{Bk}(x_j^2 - x_i^2)}{(x_k^2 - x_j^2)(x_k^2 - x_i^2)}) \quad (9)$$

$$\beta_{1Bijk} = \frac{(x_i + x_k)(x_j + x_k)}{x_j - x_i}(-\frac{h_{Hi}}{x_k^2 - x_i^2} + \frac{h_{Bj}}{(x_k^2 - x_j^2)} - \frac{h_{Hk}(x_j^2 - x_i^2)}{(x_k^2 - x_j^2)(x_k^2 - x_i^2)}) \quad (10)$$

Далее вычислим:

$$\beta_{1min} = \max_{ijk}(\beta_{1Hijk}) \quad (11)$$

$$\beta_{1max} = \min_{ijk}(\beta_{1Bijk}) \quad (12)$$

Вычисление коэффициента β_2
Аналогично вычислению β_0 и β_1 вычислим допустимые значения коэффициета β_2:

$$\beta_{2Hijk} = \frac{\frac{h_{Hi}}{(x_k - x_j)} - \frac{x_{Bj}}{(x_k - x_j)} + \frac{h_{Hk}(x_j - x_i)}{(x_k - x_j)(x_k - x_i)}}{(x_j - x_i)} \quad (13)$$

$$\beta_{2Bijk} = \frac{\frac{h_{Bi}}{(x_k - x_j)} - \frac{x_{Hj}}{(x_k - x_j)} + \frac{h_{Bk}(x_j - x_i)}{(x_k - x_j)(x_k - x_i)}}{(x_j - x_i)} \quad (14)$$

Далее вычислим:

$$\beta_{2min} = \max_{ijk}(\beta_{2Hijk}) \quad (15)$$

$$\beta_{2max} = \min_{ijk}(\beta_{2Bijk}) \quad (16)$$

Оценка параметров зависимости Далее для всех пар $\beta_{imin}, \beta_{imax}, i = 1, 2, 3$ проведём следующее сравнение.

- Если $\beta_{imin} \leq \beta_{imax}$, то выборка совместна и значение коэффициента β_i лежит в интервале $[\beta_{imin}, \beta_{imax}]$.
- Иначе выборка несовместна.

И в качестве точечной оценки коэффициента β_i возьмём $\hat{\beta}_i = \frac{\beta_{imin}+\beta_{imax}}{2}$.

2.2 Метод распознающего функционала

Перепишем исходную задачу в матричном виде:

$$X\beta = y \tag{17}$$

где x, y - исходный набор, β - искомый вектор параметров,

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ ... & ... & ... \\ 1 & x_m & x_m^2 \end{pmatrix} \tag{18}$$

Введём распознающий функционал $\text{Tol}(\beta, X, y)$.

$$\text{Tol}(\beta, X, y) = \min_{1 \leq i \leq m} (\text{rad} y_i - |\text{mid} y_i - \sum_{j=1}^{n} X_{ij}\beta_j|) \tag{19}$$

Тогда оценкой ветора параметров будет $\arg\max_\beta \text{Tol}(\beta, X, y)$

3 Реализация

Язык программирования: Python. Среда разработки: Visual Studio Code.

4 Результаты

4.1 Генерация данных

Рассмотрим модель 1. Для заданного набора входных значений $x = x_1, ..., x_m$ посчитаем точечный вектор выходных значений $y = y_1, ..., y_m$. Затем построим интервальный вектор выходных значений следующим образом: вместо каждого точечного значения y_i возьмём интервал $[y_i - |\delta_{i1}|, y_i + |\delta_{i2}|]$, где δ_{i1}, δ_{i2} - случайные величины, $\delta_{i1}, \delta_{i2} \in N(0, 5), \forall i = 1, ..., m$.

4.2 Результаты

Рассмотрим две модели:

$$y = 10 + 5x - x^2 \tag{20}$$

$$y = -50 - 20x + 3x^2 \tag{21}$$

Модель 20

Сначала сгенерируем совместную выборку для модели 20 размера 25. Исходный набор данных на рисунке ниже.

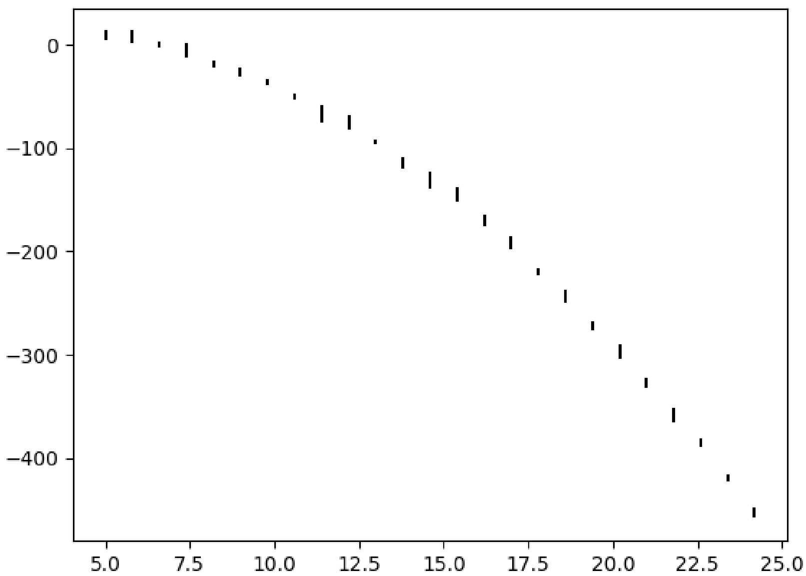

Рис. 1: Совместная выборка для модели 20

Метод центра неопределённости дал следующие результаты:

	β_{imin}	β_{imax}	$\hat{\beta}$
β_0	8.883	20.186	14.53
β_1	2.138	5.247	3.693
β_2	-1.010	-0.911	-0.961

Таблица 1: Оценки коэффициентов зависимости методом центра неопределённости для выборки по модели 20

Тогда найденная зависимость будет иметь следующий вид:

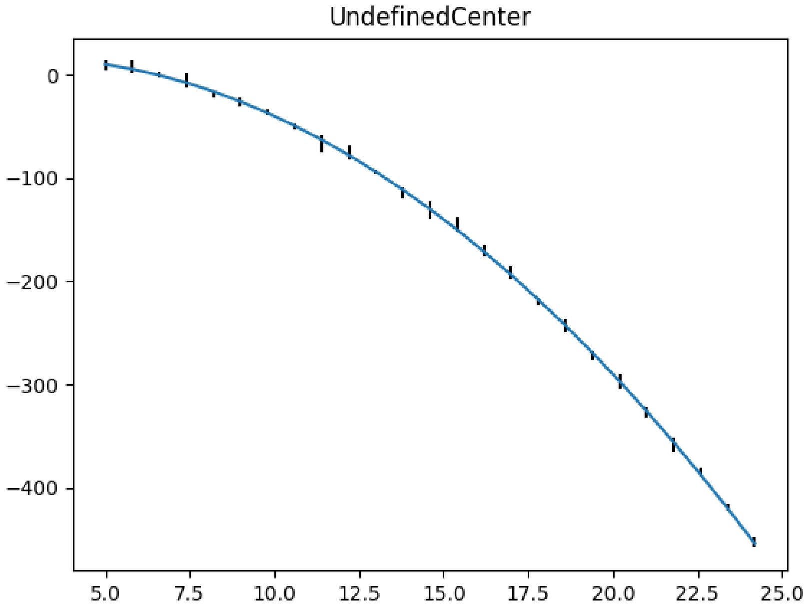

Рис. 2: Метод центра неопределённости для выборки по модели 20

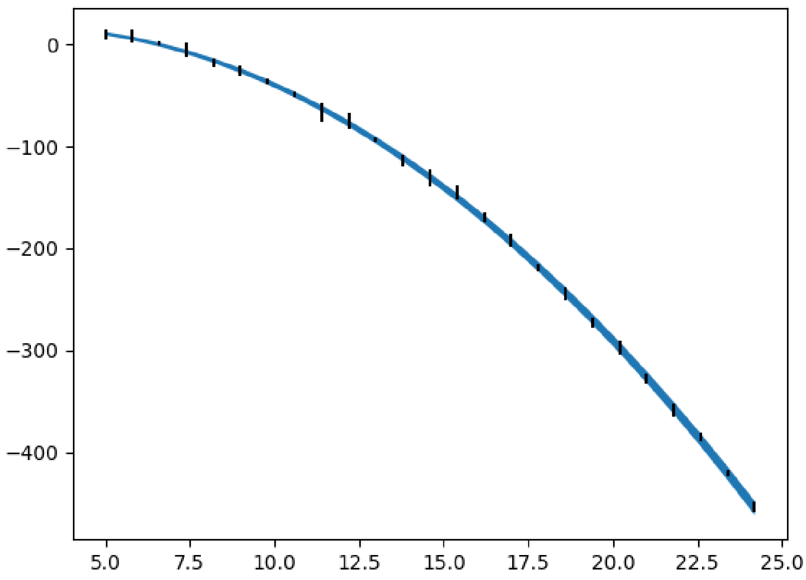

Рис. 3: Коридор совместных зависимостей

Для наглядности также рассмотрим выборку на том же промежутке, но с большим числом элементов (150), и посмотрим на полученный коридор совместных зависимостей.

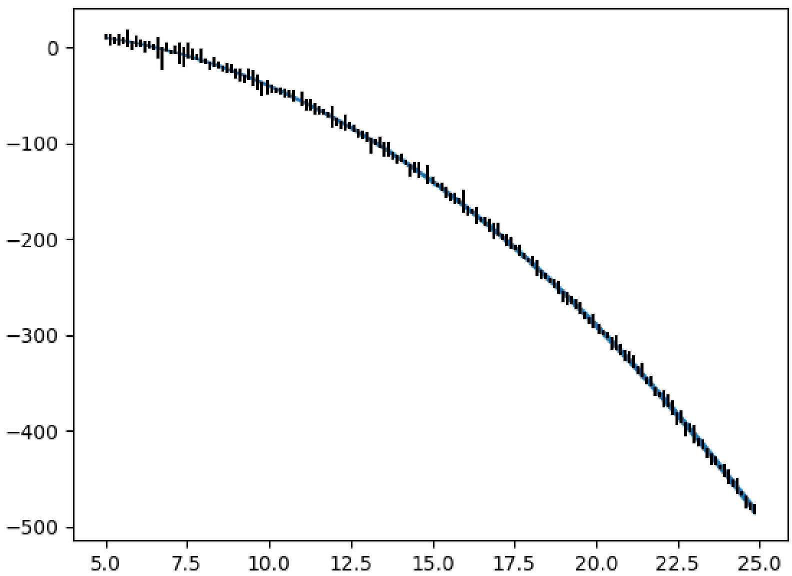

Рис. 4: Коридор совместных зависимостей для выборки большего размера

В свою очередь метод распознающего функционала дал следующую оценку коэффициентов: $\hat{\beta}_0 = 9.999, \hat{\beta}_1 = 5.025, \hat{\beta}_2 = -1.0003$. Причём значение распознающего функционала на этом наборе положительно: $\text{Tol}(\hat{\beta}) = 0.279$.
И график построенной зависимости:

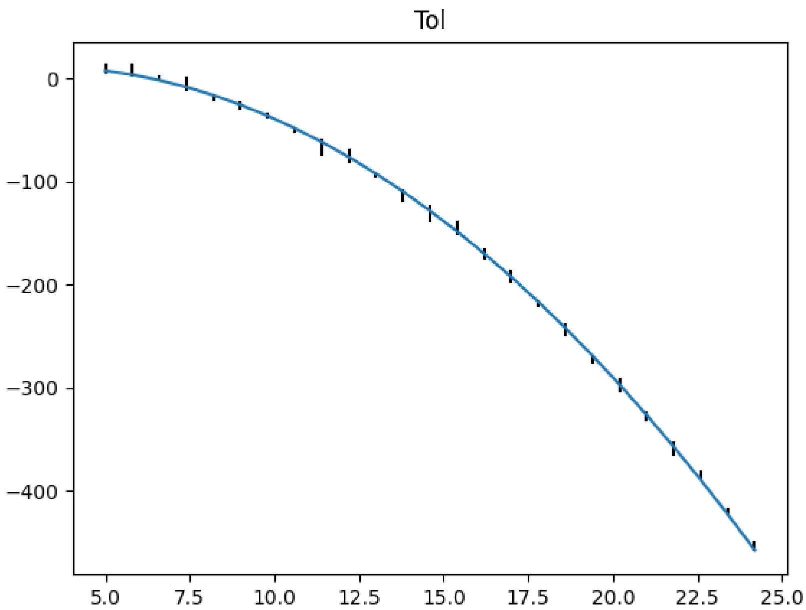

Рис. 5: Метод распознающего функционала для выборки по модели 20

Теперь внесём пять случайных малых изменений в исходную выборку, сделав её несовместной. Новая выборка будет иметь следующий вид:

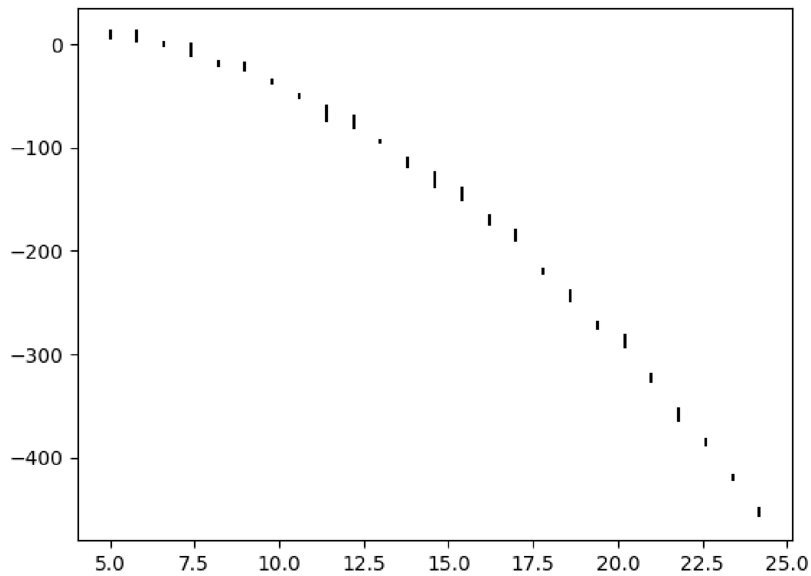

Рис. 6: Несовместная выборка для модели 20

Методом центра неопределённости получим следующие результаты:

	β_{imin}	β_{imax}	$\hat{\beta}$
β_0	68.601	-32.227	-
β_1	16.724	-3.990	-
β_2	-0.67	-1.726	-

Таблица 2: Оценки коэффициентов зависимости методом центра неопределённости для несовместной выборки по модели 20

В свою очередь метод распознающего функционала дал следующую оценку коэффициентов: $\hat{\beta}_0 = 9.999, \hat{\beta}_1 = 5.532, \hat{\beta}_2 = -1.029$. И в этом случае значение распознающего функционала на этом наборе меньше нуля: $\mathrm{Tol}(\hat{\beta}) = -1.986$.
И график построенной зависимости:

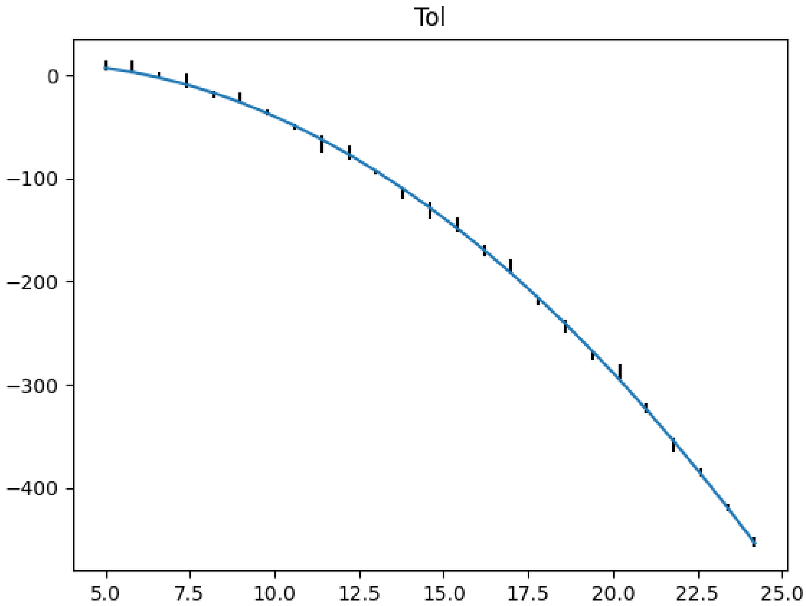

Рис. 7: Метод распознающего функционала для несовместной выборки по модели 20

Модель 21

Теперь сгенерируем совместную выборку для модели 21 размера 50 и проделаем с ней аналогичную работу. Исходная выборка имеет следующий вид:

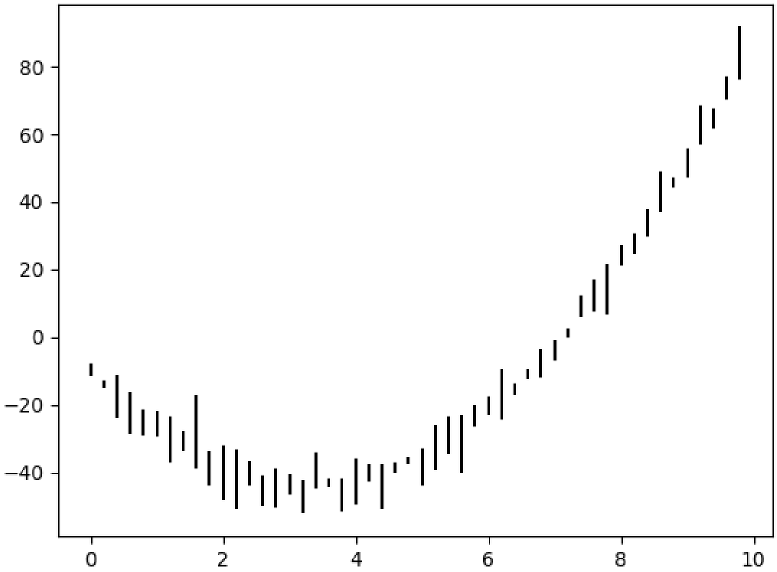

Рис. 8: Совместная выборка для модели 21

Метод центра неопределённости дал следующие результаты:

	β_{imin}	β_{imax}	$\hat{\beta}$
β_0	-10.959	-8.697	-9.828
β_1	-20.453	-19.599	-20.026
β_2	2.963	3.036	3.0002

Таблица 3: Оценки коэффициентов зависимости методом центра неопределённости для выборки по модели 21

Тогда график построенной зависимости будет иметь следующий вид:

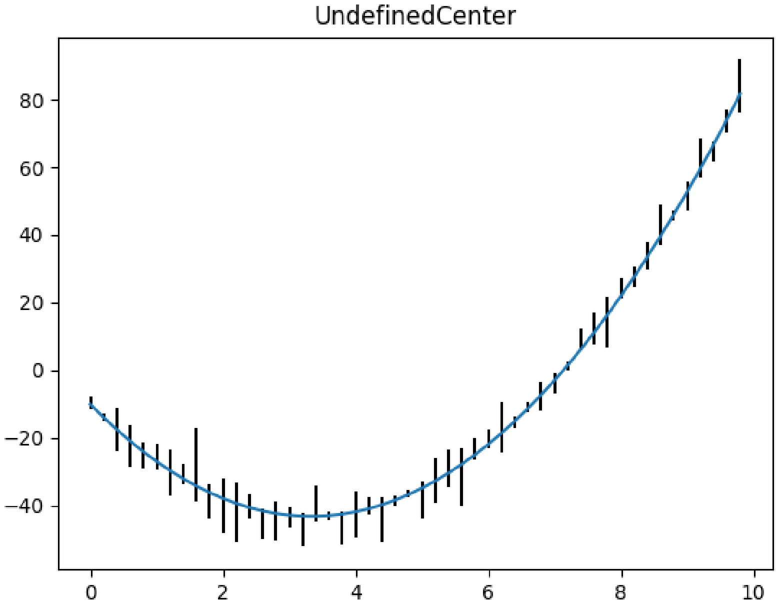

Рис. 9: Метод центра неопределённости для выборки по модели 21

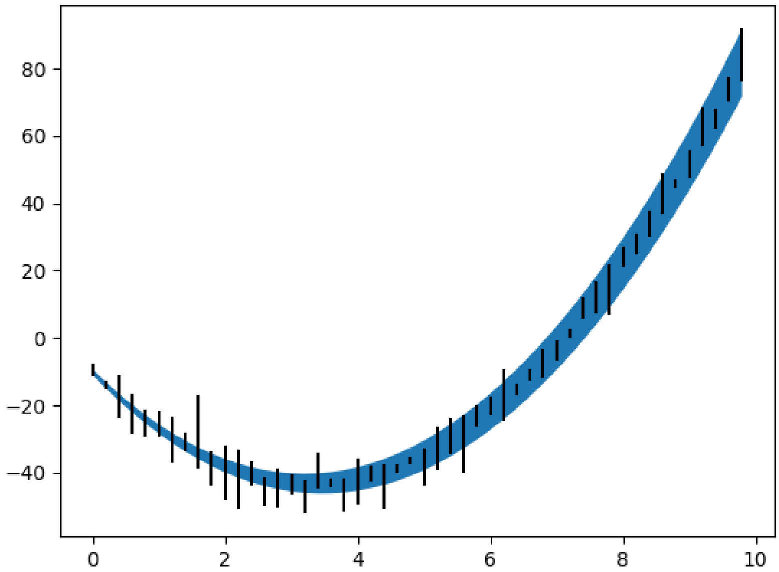

Рис. 10: Коридор совместных зависимостей

Метод распознающего функционала для следующую оценку коэффициентов зависимости: $\hat{\beta}_0 = -5.566, \hat{\beta}_1 = -10.0, \hat{\beta}_2 = 1.676$. Причём значение распознающего функционала на этом значении меньше нуля: $\mathrm{Tol}(\hat{\beta}) = -22.834$.

И тогда график найденной зависимости будет иметь следующий вид:

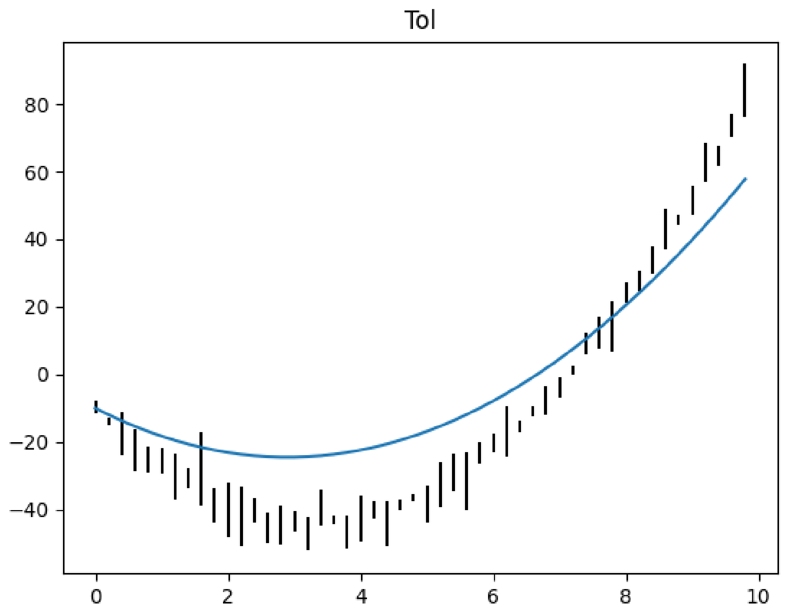

Рис. 11: Метод распознающего функционала для выборки по модели 21

Теперь внесём пять малых случайных изменений в выборку так, чтобы она стала несовместной. Тогда получим выборку:

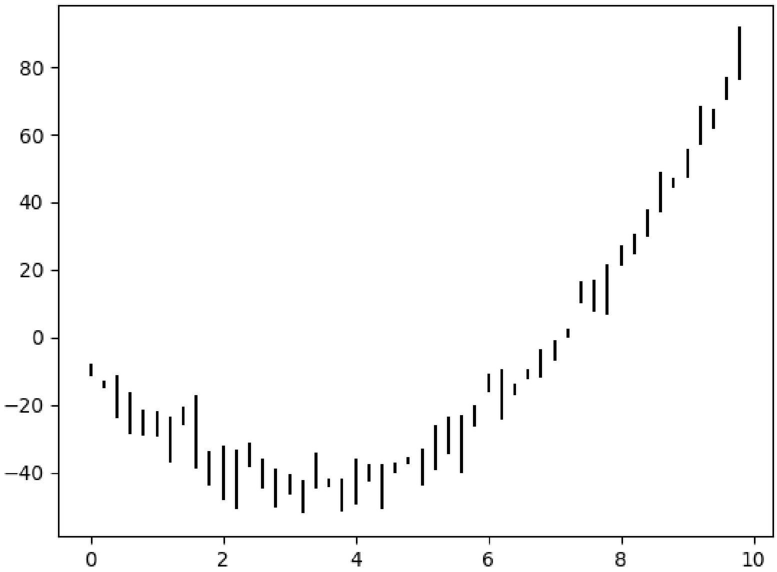

Рис. 12: Несовместная выборка для модели 21

Метод центра неопределённости дал следующие результаты:

	β_{imin}	β_{imax}	$\hat{\beta}$
β_0	1806.275	-1923.715	-
β_1	441.179	-416.402	-
β_2	24.593	-24.692	-

Таблица 4: Оценки коэффициентов зависимости методом центра неопределённости для несовместной выборки по модели 21

А в свою очередь метод распзнающего функционала дал следующую оценку коэффициентов: $\hat{\beta_0} = -9.99\hat{\beta_1} = -9.99, \hat{\beta_2} = 1.82$. Распознающий функционал на этом наборе имеет отрицательное значение: $\mathrm{Tol}(\hat{\beta}) = -20.7$.
График построенной зависимости имеет следующий вид:

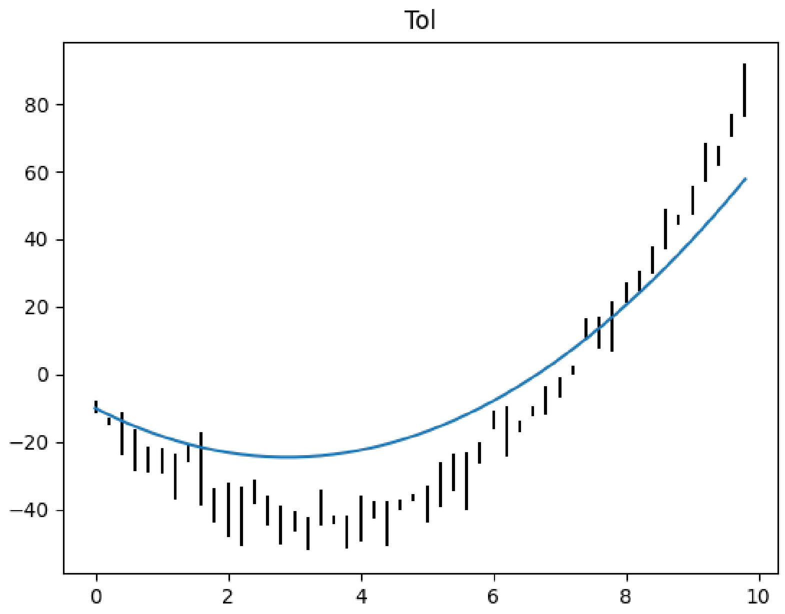

Рис. 13: Метод распознающего функционала для несовместной выборки по модели 21

5 Обсуждение

Из полученных результатов можно сделать следующие наблюдения. Для совместной выборки метод центра неопределённости более точно оценивает коэффициенты параметров зависимости. Особенно заметно лучшие оценки коэффициентов метод центра неопределённости строит для второй выборки, что видно на рисунках 2, 9, 5. Причём для обеих выборок построенная методом центра неопределённости зависимость проходит через все интервалы, что видно на рисунках 2, 9.
Но в свою очередь метод центра неопределённости оказывается сильно чувствительным к выбросам, что можно заметить из результатов в таблицах 2, 4. Даже при совсем небольших отклонениях от совместных данных метод центра неопределённости даёт некорректные результаты. Метод распознающего функционала является куда менее чувствительным к выбросам, что видно на рисунках 5, 7, хотя всё же на выборке с выбросами максимальное значение распознающего функционала становиться отрицательным, что также говорит о несовместности системы.
Также стоит отметить, что даже на совместной выборке метод распо-

знающего функционала может давать плохие результаты, что видно на рисунке 11. Причём максимальное значение распознающего функционала в этом случае сильно меньше нуля.

Из рисунков 3, 10, 4 можно сделать предположение, что ширина коридора зависит от числа элементов в выборке. Это особенно наглядно видно, на рисунках 3, 4.

Также стоит отметить, что метод центра неопределённости даёт наименьшую неопределённость для коэффициента β_2, а наибольшую для коэффициента β_0, что видно на таблицах 1 - 4.