

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

<b>1</b>	<b>Постановка задачи</b>	<b>4</b>
<b>2</b>	<b>Теория</b>	<b>4</b>
2.1	Метод центра неопределённости . . . . .	4
2.2	Метод распознающего функционала . . . . .	6
<b>3</b>	<b>Реализация</b>	<b>6</b>
<b>4</b>	<b>Результаты</b>	<b>6</b>
4.1	Генерация данных . . . . .	6
4.2	Результаты . . . . .	7
<b>5</b>	<b>Обсуждение</b>	<b>28</b>

## Список иллюстраций

1	Совместная выборка для модели 20 . . . . .	7
2	Метод центра неопределённости для выборки по модели 20	8
3	Коридор совместных зависимостей . . . . .	9
4	Коридор совместных зависимостей для выборки большего размера . . . . .	10
5	Метод распознающего функционала для выборки по моде- ли 20 . . . . .	11
6	Несовместная выборка для модели 20 . . . . .	12
7	Метод распознающего функционала для несовместной вы- борки по модели 20 . . . . .	13
8	Совместная выборка для модели 21 . . . . .	14
9	Метод центра неопределённости для выборки по модели 21	15
10	Коридор совместных зависимостей . . . . .	16
11	Метод распознающего функционала для выборки по моде- ли 21 . . . . .	17
12	Несовместная выборка для модели 21 . . . . .	18
13	Метод распознающего функционала для несовместной вы- борки по модели 21 . . . . .	19
14	35 примеров на промежутке от 0 до 30 . . . . .	20
15	Результат метода центра неопределённости за пределами данных . . . . .	21
16	Коридор метода центра неопределённости за пределами данных . . . . .	22
17	Результат метод распознающего функционала за пределами данных . . . . .	23
18	60 примеров на промежутке от -5 до 15 . . . . .	24
19	Результат метода центра неопределённости за пределами данных . . . . .	25
20	Коридор метода центра неопределённости за пределами данных . . . . .	26
21	Результат метод распознающего функционала за пределами данных . . . . .	27

## Список таблиц

1	Оценки коэффициентов зависимости методом центра неопределённости для выборки по модели 20 . . . . .	7
2	Оценки коэффициентов зависимости методом центра неопределённости для несовместной выборки по модели 20 . . . . .	12
3	Оценки коэффициентов зависимости методом центра неопределённости для выборки по модели 21 . . . . .	14
4	Оценки коэффициентов зависимости методом центра неопределённости для несовместной выборки по модели 21 . . . . .	18

## 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  $x_1, x_2, \dots, x_p$ .

$$y = f(x, \beta) \quad (1)$$

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

Будем рассматривать функцию вида:

$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 \quad (2)$$

### 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  $\forall i = 1, \dots, m$ .

**Вычисление коэффициента  $\beta_0$**

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  $\beta_0$ :

$$\beta_{0H0jk} = h_{H0}, \text{ если } i = 0 \quad (3)$$

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

$$\beta_{0B0jk} = h_{B0}, \text{ если } i = 0 \quad (5)$$

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

Далее вычисляем:

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

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

### Вычисление коэффициентов $\beta_1$

Аналогично вычислению  $\beta_0$  вычислим допустимые значения коэффициента  $\beta_1$ :

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

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

Далее вычислим:

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

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

### Вычисление коэффициента $\beta_2$

Аналогично вычислению  $\beta_0$  и  $\beta_1$  вычислим допустимые значения коэффициента  $\beta_2$ :

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

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

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  $\beta_i$  лежит в интервале  $[\beta_{imin}, \beta_{imax}]$ .
- Иначе выборка несовместна.

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

## 2.2 Метод распознающего функционала

Перепишем исходную задачу в матричном виде:

$$X\beta = y \quad (17)$$

где  $x, y$  - исходный набор,  $\beta$  - искомый вектор параметров,

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \dots & \dots & \dots \\ 1 & x_m & x_m^2 \end{pmatrix} \quad (18)$$

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

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

Тогда оценкой вектора 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  $x = x_1, \dots, x_m$  посчитаем точечный вектор выходных значений  $y = y_1, \dots, y_m$ . 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  $\delta_{i1}, \delta_{i2}$  - случайные величины,  $\delta_{i1}, \delta_{i2} \in N(0, 5), \forall i = 1, \dots, m$ .

## 4.2 Результаты

Рассмотрим две модели:

$$y = 10 + 5x - x^2 \quad (20)$$

$$y = -50 - 20x + 3x^2 \quad (21)$$

### Модель 20

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

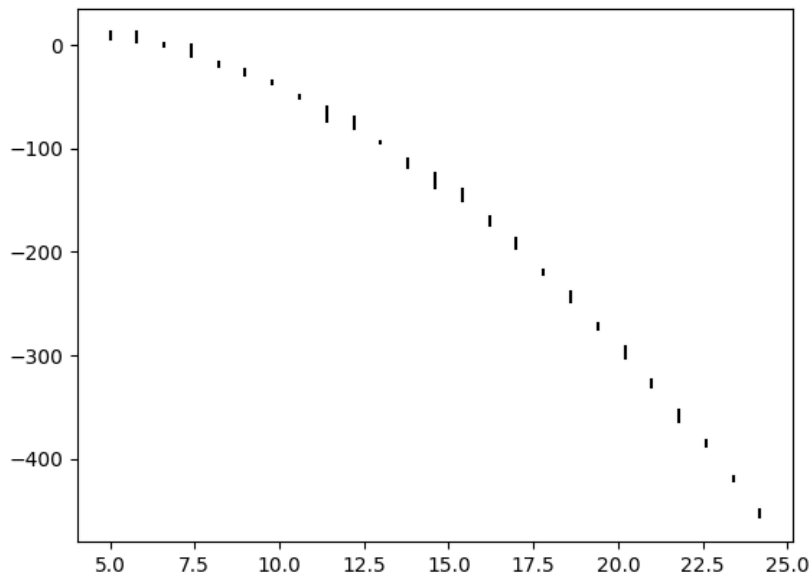


Рис. 1: Совместная выборка для модели 20

Метод центра неопределённости дал следующие результаты:

	$\beta_{imin}$	$\beta_{imax}$	$\hat{\beta}$
$\beta_0$	8.883	20.186	14.53
$\beta_1$	2.138	5.247	3.693
$\beta_2$	-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:

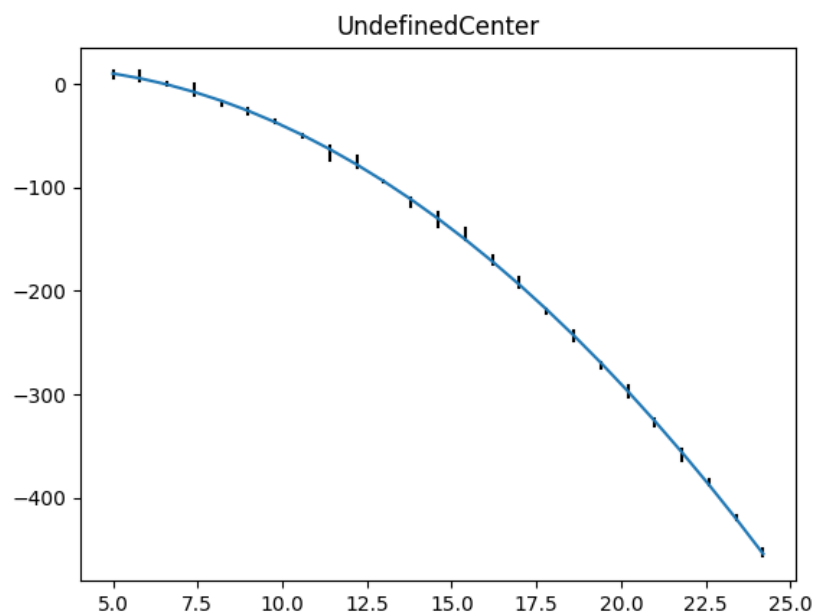


Рис. 2: Метод центра неопределённости для выборки по модели 20

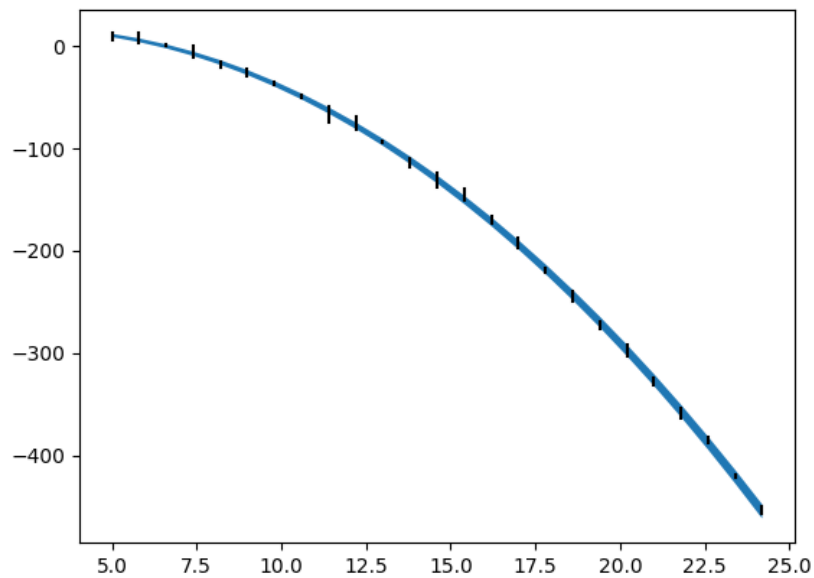


Рис. 3: Коридор совместных зависимостей

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

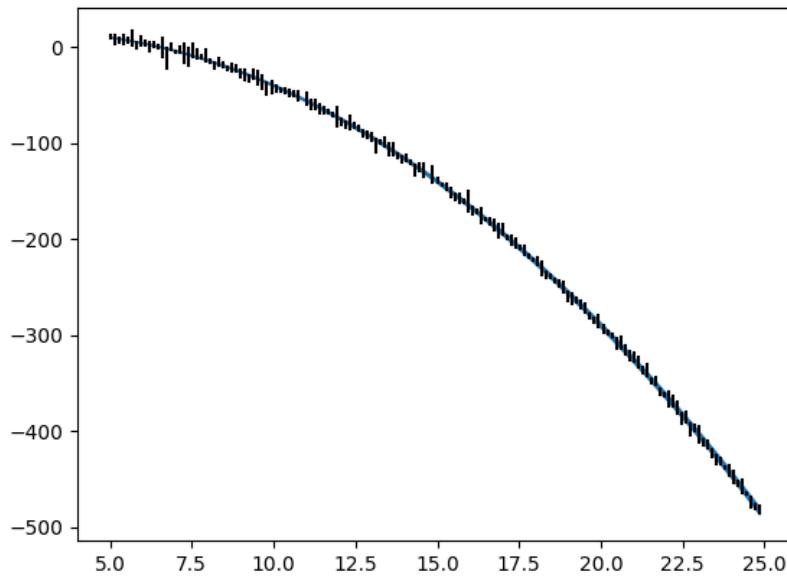


Рис. 4: Коридор совместных зависимостей для выборки большего размера

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:

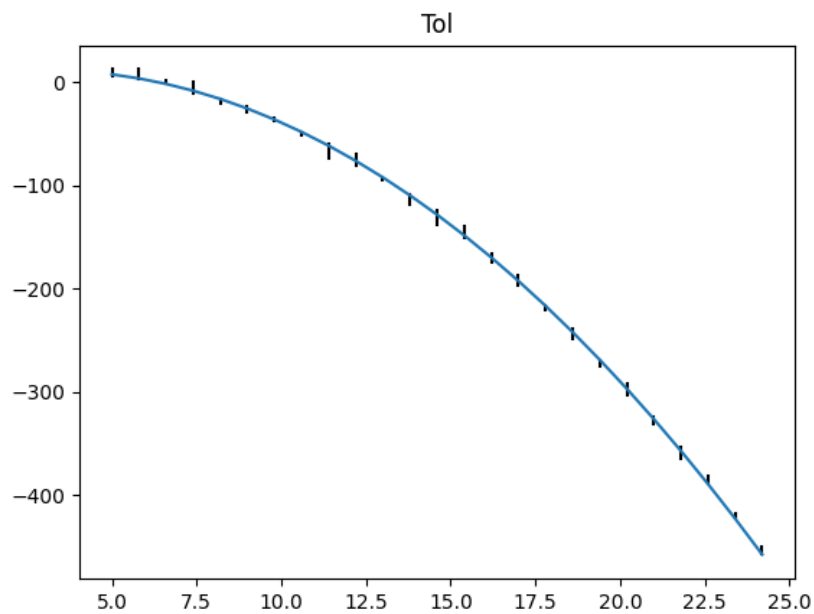


Рис. 5: Метод распознающего функционала для выборки по модели 20

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

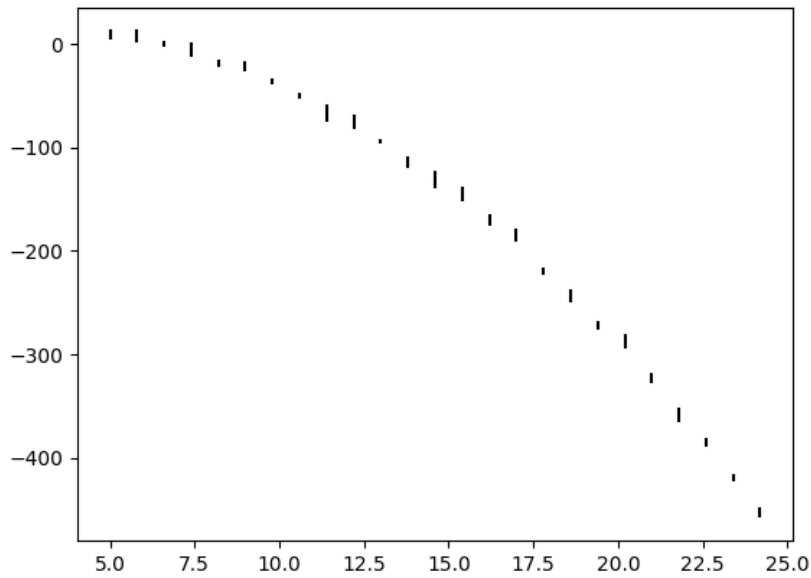


Рис. 6: Несовместная выборка для модели 20

Методом центра неопределённости получим следующие результаты:

	$\beta_{imin}$	$\beta_{imax}$	$\hat{\beta}$
$\beta_0$	68.601	-32.227	-
$\beta_1$	16.724	-3.990	-
$\beta_2$	-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:  $\text{Tol}(\hat{\beta}) = -1.986$ .

И график построенной зависимости:

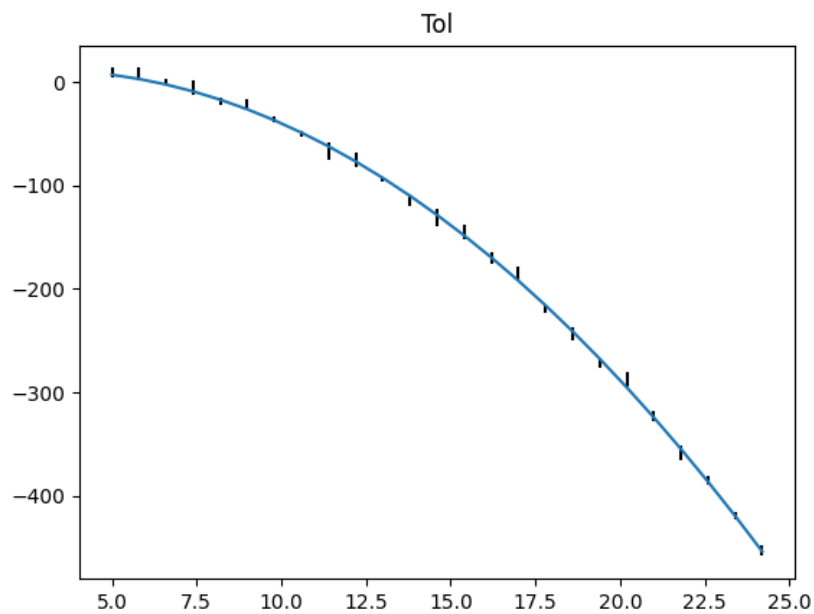


Рис. 7: Метод распознающего функционала для несовместной выборки по модели 20

### Модель 21

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

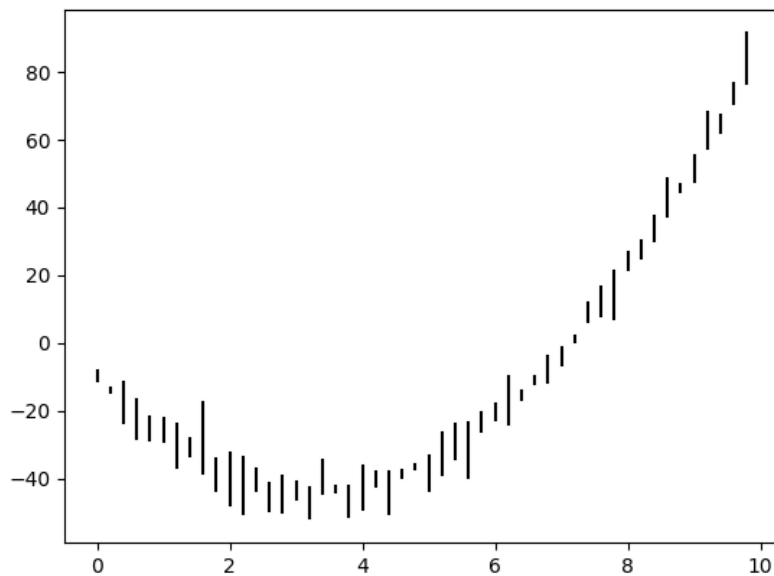


Рис. 8: Совместная выборка для модели 21

Метод центра неопределённости дал следующие результаты:

	$\beta_{imin}$	$\beta_{imax}$	$\hat{\beta}$
$\beta_0$	-10.959	-8.697	-9.828
$\beta_1$	-20.453	-19.599	-20.026
$\beta_2$	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:

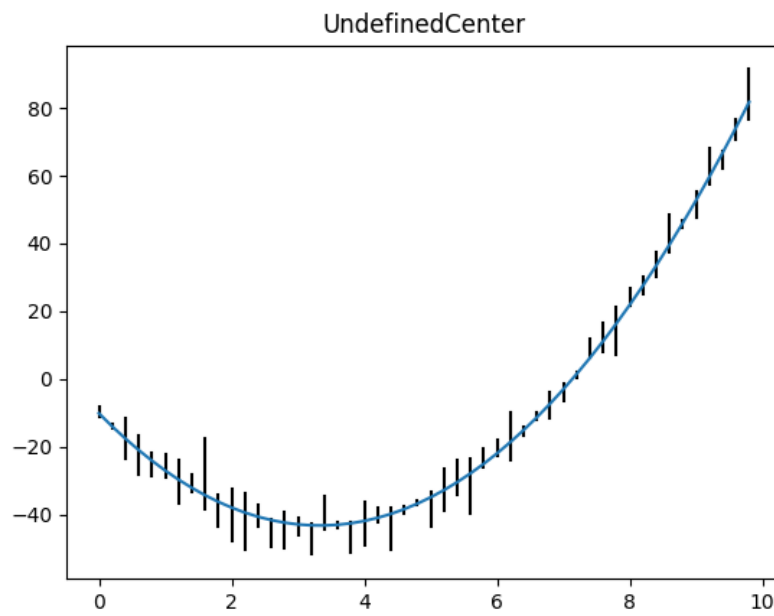


Рис. 9: Метод центра неопределённости для выборки по модели 21



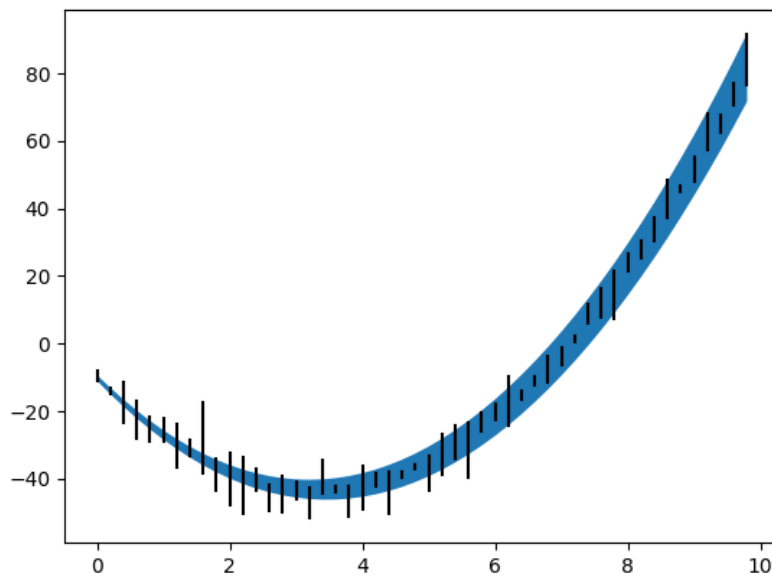


Рис. 10: Коридор совместных зависимостей

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:  $\text{Tol}(\hat{\beta}) = -22.834$ .

И тогда график найденной зависимости будет иметь следующий вид:

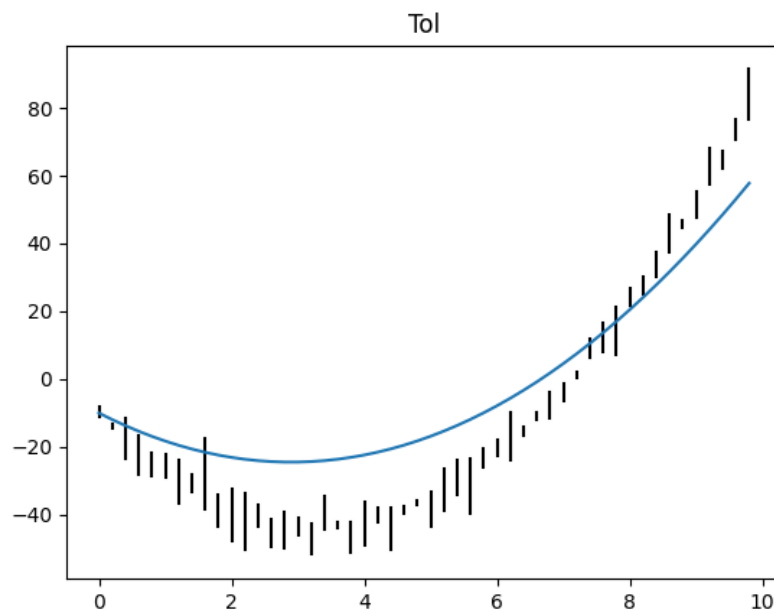


Рис. 11: Метод распознающего функционала для выборки по модели 21

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

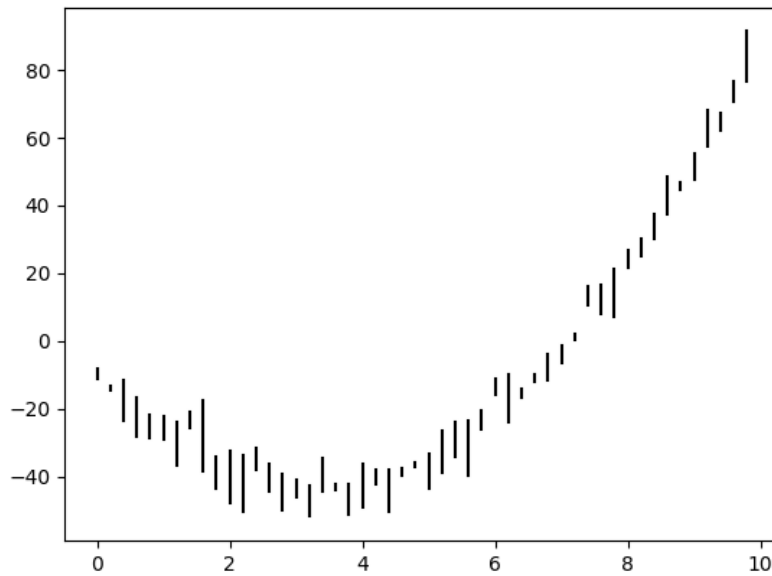


Рис. 12: Несовместная выборка для модели 21

Метод центра неопределённости дал следующие результаты:

	$\beta_{imin}$	$\beta_{imax}$	$\hat{\beta}$
$\beta_0$	1806.275	-1923.715	-
$\beta_1$	441.179	-416.402	-
$\beta_2$	24.593	-24.692	-

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

А в свою очередь метод распознающего 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:  $\text{Tol}(\hat{\beta}) = -20.7$ .

График построенной зависимости имеет следующий вид:

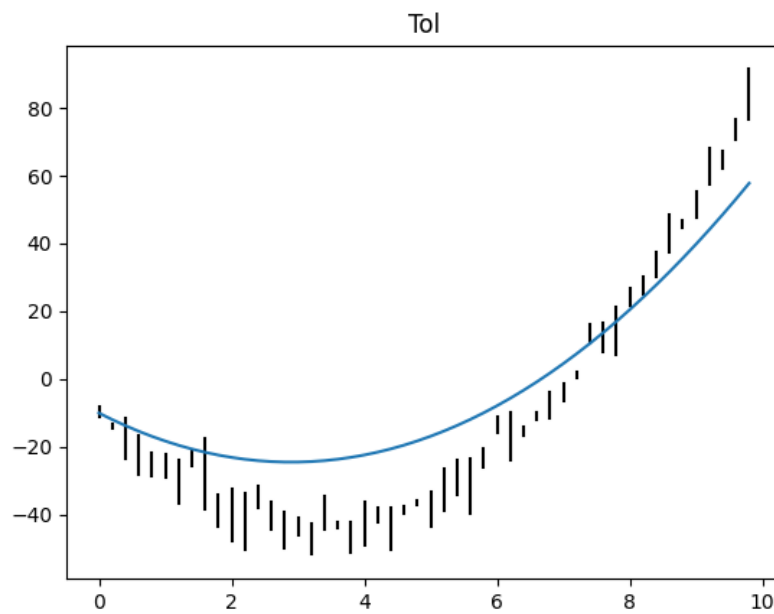


Рис. 13: Метод распознающего функционала для несовместной выборки по модели 21

### Предсказания за пределами данных

Рассмотрим поведение методов за пределами данных на которых строится модель. Для модели 20 сгенерируем выборку в промежутке от 0 до 30 из 35 примеров, после чего выберем из неё примеры с 5-ого по 25-ый и построим на них регрессию. Также для модели 21 сгенерируем выборку из 60 примеров в промежутке от 0 до 20

#### Модель 20

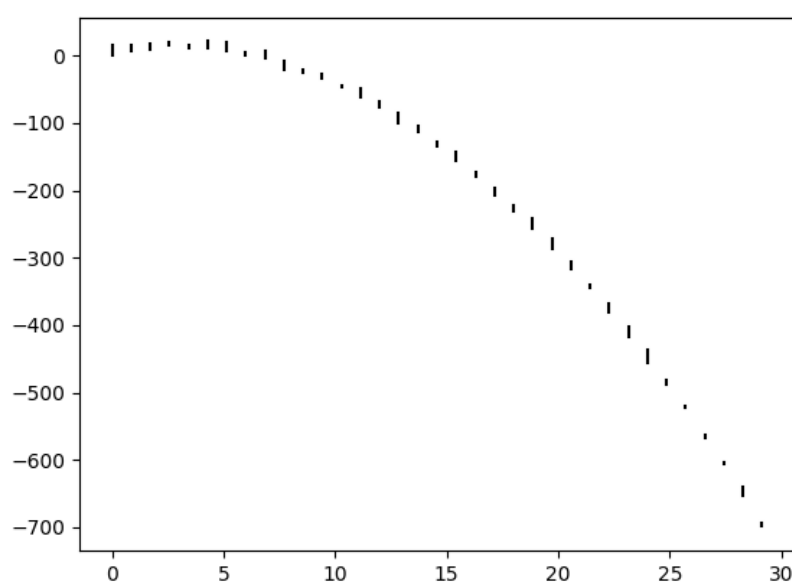


Рис. 14: 35 примеров на промежутке от 0 до 30

Метод центра неопределённости дал следующие результаты

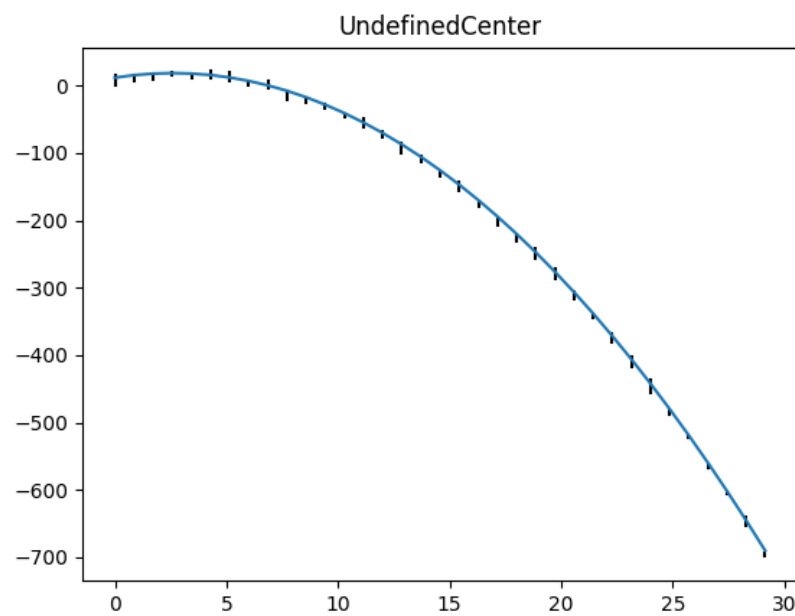


Рис. 15: Результат метода центра неопределённости за пределами данных

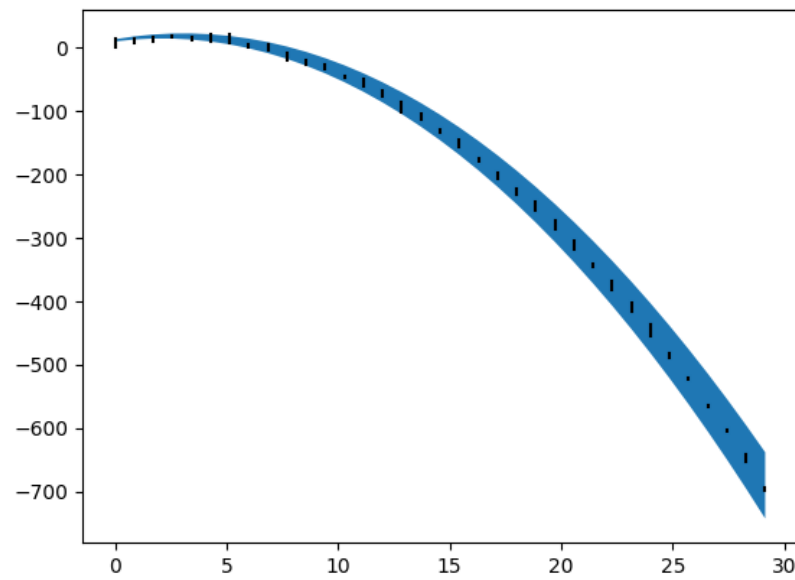


Рис. 16: Коридор метода центра неопределённости за пределами данных

Метод распознающего функционала показал следующие результаты.

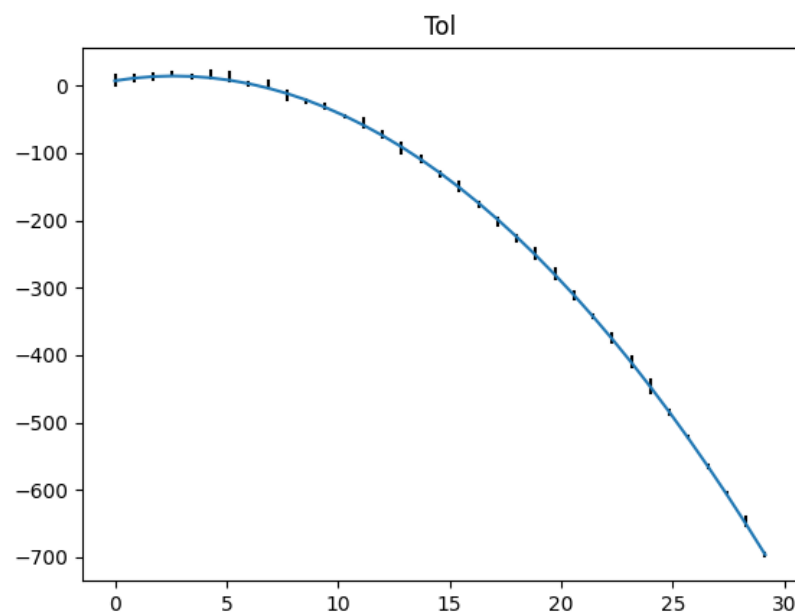


Рис. 17: Результат метод распознающего функционала за пределами данных



## Модель 21

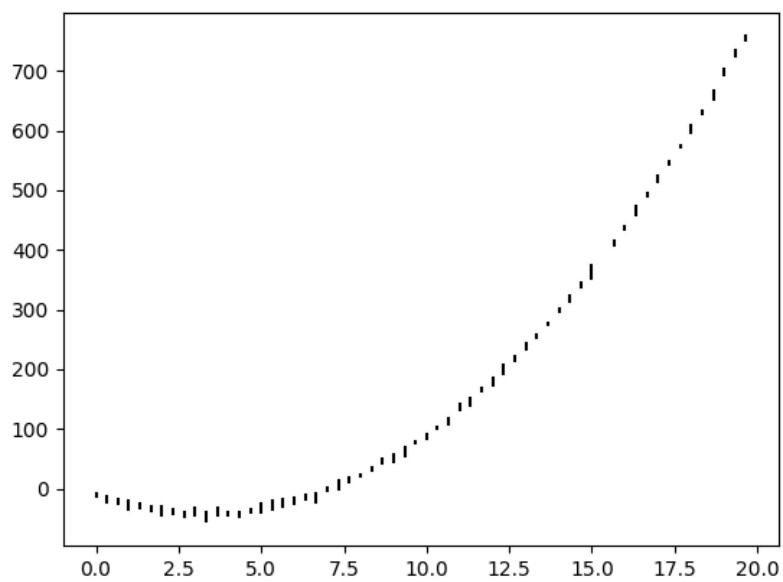


Рис. 18: 60 примеров на промежутке от 0 до 20

Метод центра неопределённости даёт следующие результаты

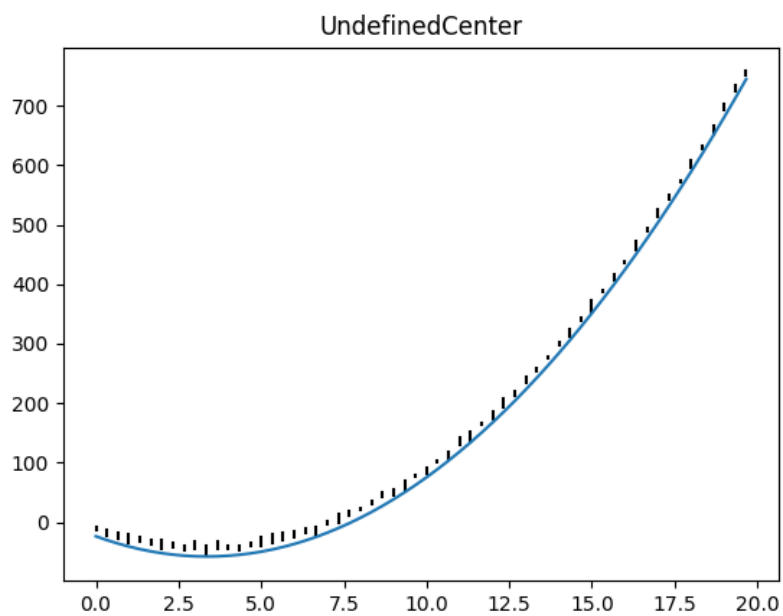


Рис. 19: Результат метода центра неопределённости за пределами данных

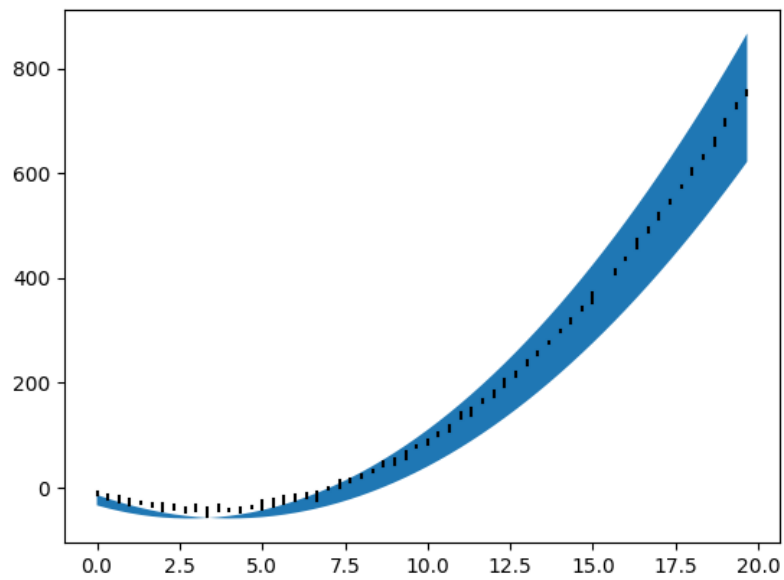


Рис. 20: Коридор метода центра неопределённости за пределами данных

Метод распознающего функционала показал следующие результаты.

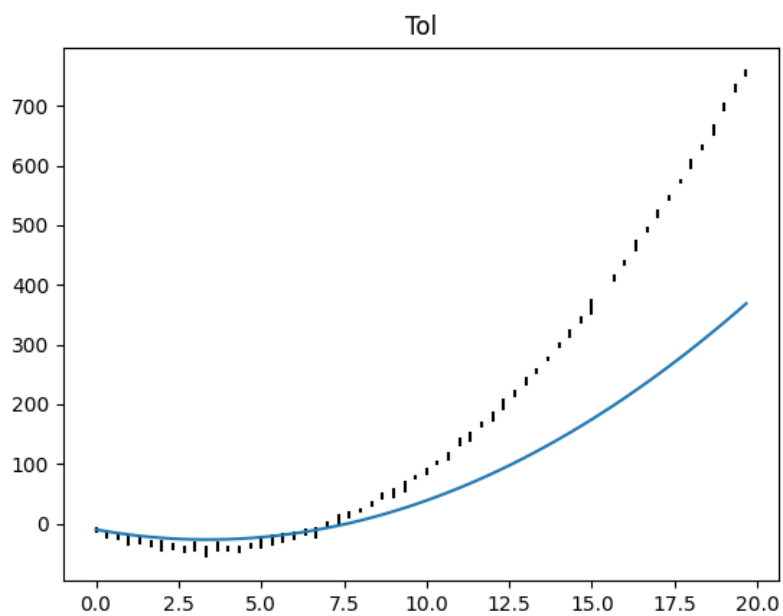


Рис. 21: Результат метод распознающего функционала за пределами данных

## 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 становится отрицательным, что также говорит о несовместности системы.

Также стоит отметить, что даже на совместной выборке метод распознающего 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  $\beta_2$ , а наибольшую для коэффициента  $\beta_0$ , что видно на таблицах 1 - 4.