Санкт-Петербургский политехнический университет Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

Интервальный анализ
Отчёт по лабораторной работе №2

Выполнил:

Студент: Дамаскинский Константин
Группа: 5040102/10201

Принял:

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

2023 г.

Содержание

Список иллюстраций

1. Постановка задачи

Проводится исследование из области солнечной энергетики. На рис. 1 показана схема установки для исследования фотоэлектрических характеристик.

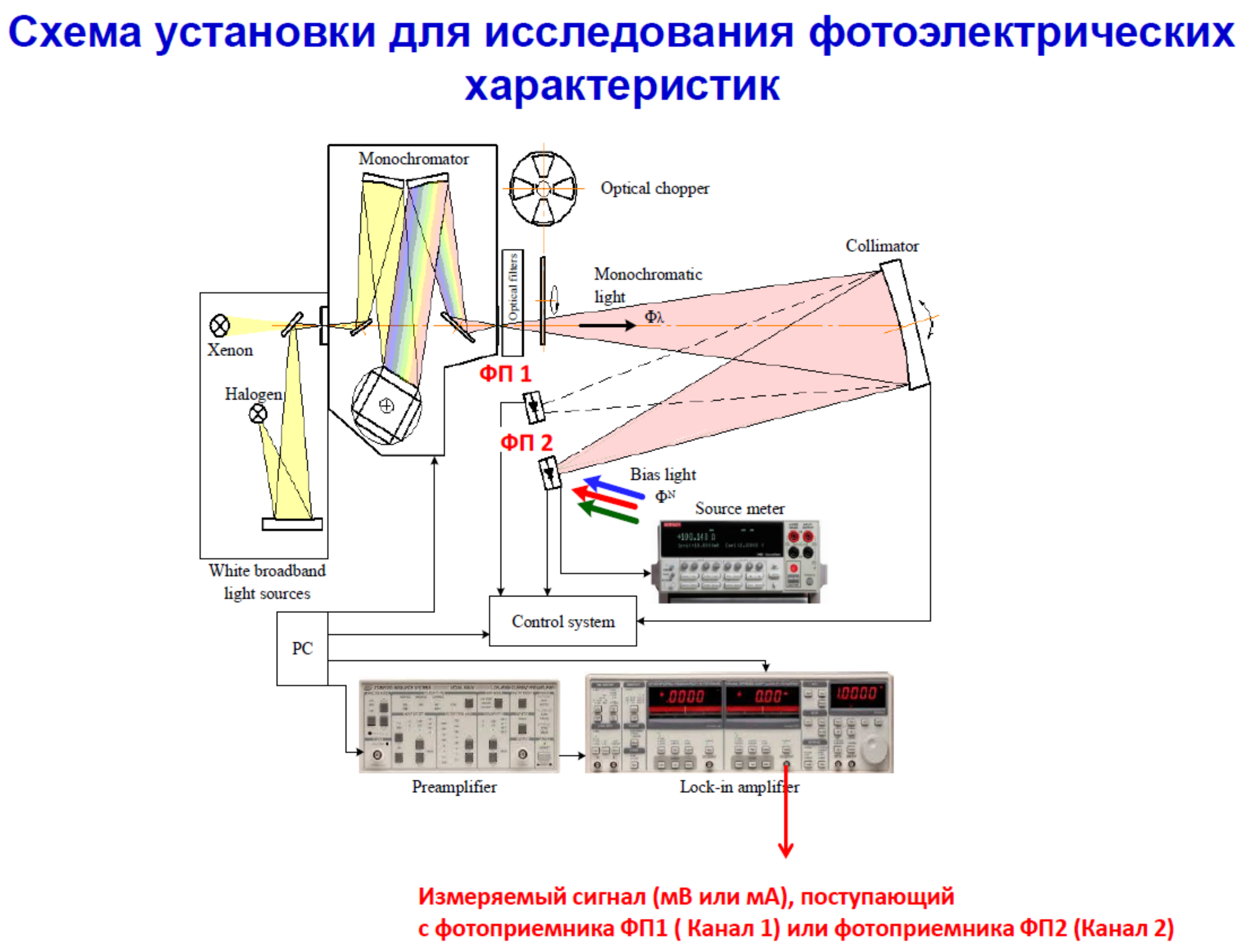

Рис. 1. Схема установки для исследования фотоэлектрических характеристик

Калибровка датчика ФП2 производится по эталону ФП1. Зависимость между квантовыми эффективностями датчиков предполагается постоянной для каждой пары наборов измерений:

$$QE_2 = \frac{I_2}{I_1} \cdot QE_1 \qquad (1)$$

где QE_2, QE_1 – эталонная эффективность эталонного и исследуемого датчика, I_2, I_1 – измеренные токи. Данные с датчиков находятся в файлах **Канал2_800nm_0.2.csv**, **Канал1_800nm_0.2.csv** и полагаются линейными.

Требуется определить коэффициент калибровки:

$$R_{21} = \frac{I_2}{I_1} \tag{2}$$

на основе линейной регрессии на множестве интервальных данных. Также требуется построить информационное множество данной задачи и коридор совместных зависимостей для двух выборок.

2. Теория

2.1. Схема решения задачи

Будем, как и в прошлой работе, отдельно решать задачу интервальной регресии для двух наборов входных данных $(I, \mathbf{Y_1}), (I, \mathbf{Y_2})$. Здесь I – номера измерений, $\mathbf{Y_1}, \mathbf{Y_2}$ – обынтерваленные измеренные значения. В отличие от первой работы, будем решать эти задачи как задачи интервальной, а не вещественной регрессии, описанным ниже способом.

Далее, аналогично предыдущей работе, найдём оптимальный коэффициент R_{21}, максимизируя коэффициент Жаккара.

2.2. Интервальная регрессия как задача оптимизации

В данной работе для решения задачи интервальной регрессии будем использовать следующий подход.

Будем искать зависимость $y^{(k)} = \beta_0^{(k)} + \beta_1^{(k)} x$ таким образом, чтобы, минимально расширив интервалы исходного интервального вектора $\{\mathbf{y_i}\}_{i=1}^n$, получить набор интервалов, накрывающий аппроксимирующую прямую:

$$\begin{cases} \mathtt{mid}\mathbf{y_i^{(k)}} - w_i^{(k)} \cdot \mathtt{rad}\mathbf{y_i^{(k)}} \le \beta_0^{(k)} + \beta_1^{(k)} i \le \mathtt{mid}\mathbf{y_i^{(k)}} + w_i^{(k)} \cdot \mathtt{rad}\mathbf{y_i^{(k)}} & , \ i = \overline{1, n} \\ \sum\limits_{i=1}^{n} w_i^{(k)} \longrightarrow \min & \\ w_i^{(k)} \ge 0 & , \ i = \overline{1, n} \\ w^{(k)}, \beta_0^{(k)}, \beta_1^{(k)} = ? & \end{cases} \tag{3}$$

Здесь $k \in \{1, 2\}$ – номер набора данных.

Данная задача является задачей линейного программирования. Как и в прошлой работе, примем $\varepsilon := \mathtt{rad}\mathbf{y_i^{(k)}} = 10^{-4}$ для всех $i = \overline{1, n}$.

2.3. Информационное множество

Применительно к данной задаче, информационное множество – это все такие пары (β_0, β_1), при которых выполнено первое ограничение типа неравенства задачи оптимизации 3.

2.4. Коридор совместных зависимостей

В постановке задаче оптимизации 3 не ставится никаких ограничений и целей по минимизации для параметров β_0, β_1. Ясно, что параметры β_0, β_1, полученные в результате решения задачи оптимизации, будут не единственными допустимыми: информационное множество задает целое семейство допустимых β_0, β_1. Следовательно, имеет смысл рассматривать, как единое целое, множество всех функций, совместных с интервальными данными задачи восстановления зависимостей. Такое множество называется **коридором совместности**. **Граничными** называются измерения, определяющие какой-либо фрагмент границы множества. Это свойство имеет смысл рассматривать для наблюдений, принадлежащих выборке, по которой строилась модель. Граничные измерения задают минимальную подвыборку, определяющую модель.

3. Реализация

Данная работа реализована на языке программирования Python 3.10 с использованием пакетов `numpy` и `scikit`. Также использовался модуль `interval` вычислительного пакета Octave и библиотека программ С. Жилина. Код данного отчёта подготовлен с использованием редактора TeXstudio и компилятора pdflatex.

4. Результаты

4.1. Замечания относительно пакета glpk

Существенный объём времени был потрачен на выяснение причин, по которым пакет glpk не решал задачи, поставленные в модуле ir_outer.m (минимизация и максимизация β_0 и β_1 покомпонентно). Выяснилось, что солвер revised_simplex не в состоянии найти решения этих задач. Кроме того, солвер interior_point не мог найти граничные β_0 и β_1, когда на переменные не устанавливалось ограничений снизу. После того, как было установлено дефолтное ограничение снизу (т.е 0), interior_point со всеми задачами успешно справился. В то же время, солвер revised_simplex в пакете scipy успешно решал те же задачи без искусственных ограничений. Эти проблемы значительно усложнили реализацию данной лабораторной работы, и их, на мой взгляд, следует обнародовать среди студентов.

4.2. Линейная модель

Ниже приведены графики, полученные в результате работы реализованной программы.

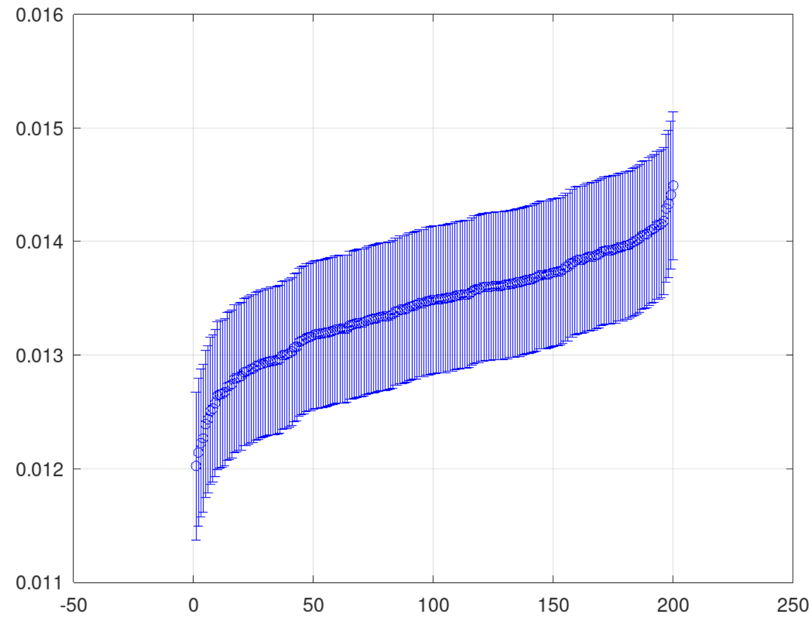

Рис. 2. Обынтерваленные данные. Модель 1

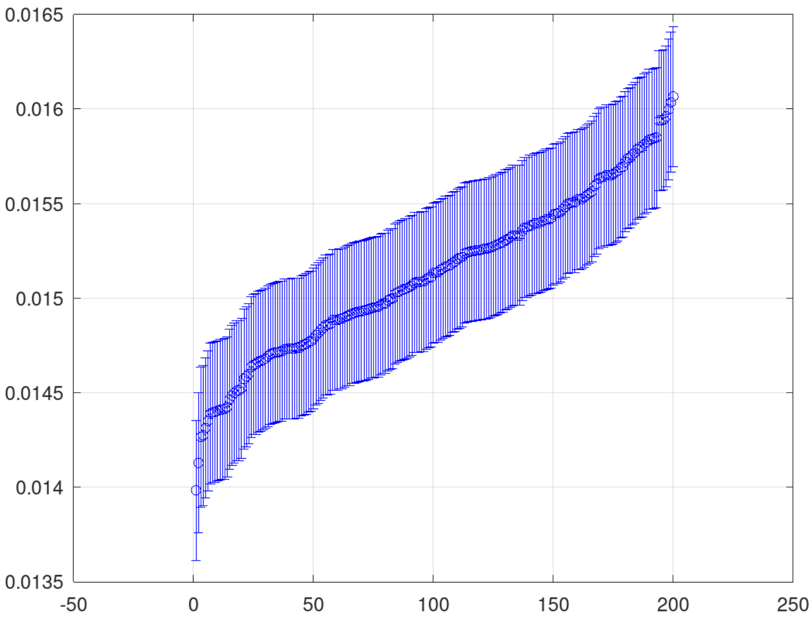

Рис. 3. Обынтерваленные данные. Модель 2

Для ускорения вычислительных процессов и более простого взаимодействия с данными, все интервалы были расширены в максимум из всех полученных весов раз: ширина интервалов составляети $\varepsilon \cdot \max_{i=\overline{1,n}(w_i)}$.

В следующей таблице приведены некоторые отличные от единицы веса:

Номер интервала	Вес (модель 1)	Вес (модель 2)
1	6.49	3.70
2	5.38	2.33
3	4.63	1.04
199	2.01	1.15
200	2.76	1.37

Таблица 1. Веса интервалов

В обоих случай максимальный вес пришёлся на первый интервал.

В следующей таблице указаны полученные параметры линейной интервальной регрессии (maxdiag).

Модель	β_0	β_1	$\max w$
1	0.012280	$1.0403 \cdot 10^{-5}$	6.49
2	0.014142	$9.3126 \cdot 10^{-6}$	3.70

Таблица 2. Параметры линейной интервальной регрессии

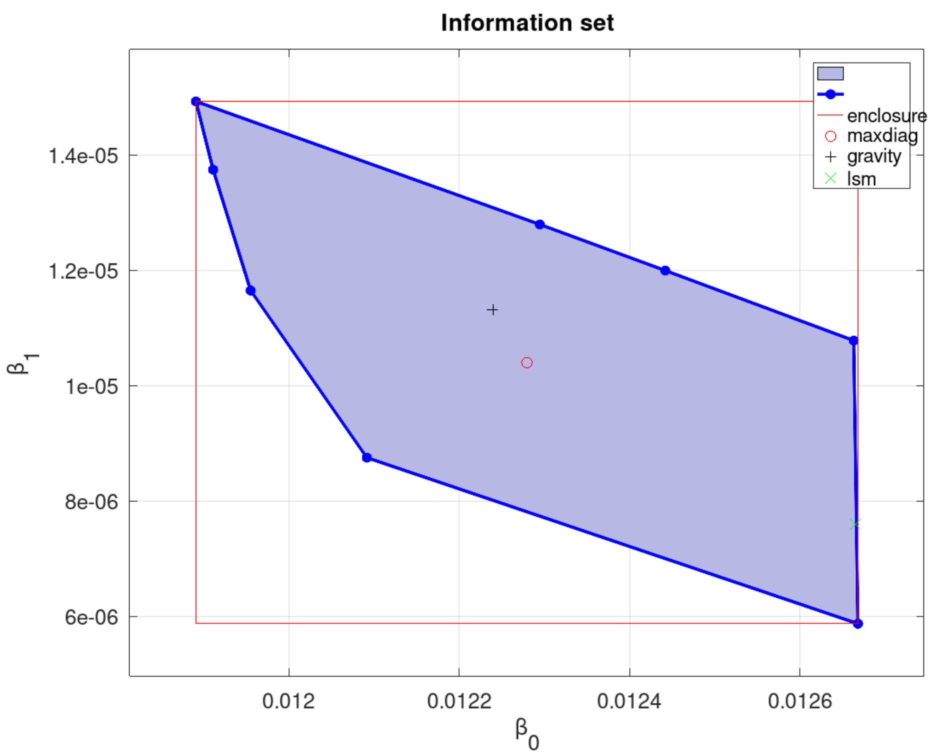

Рис. 4. Информационное множество. Модель 1

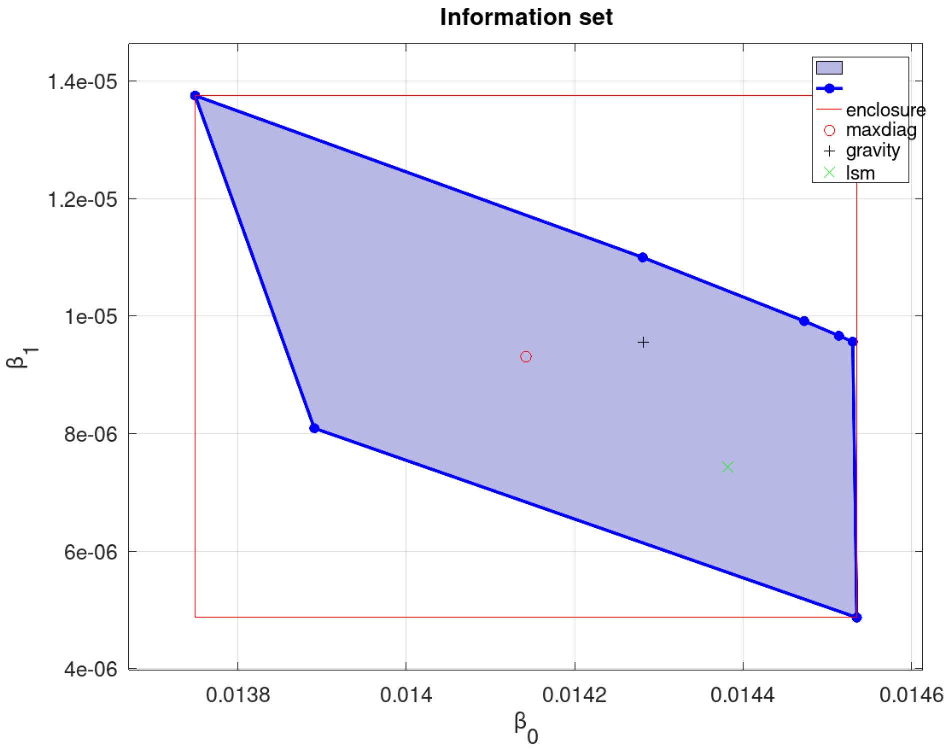

Рис. 5. Информационное множество. Модель 2

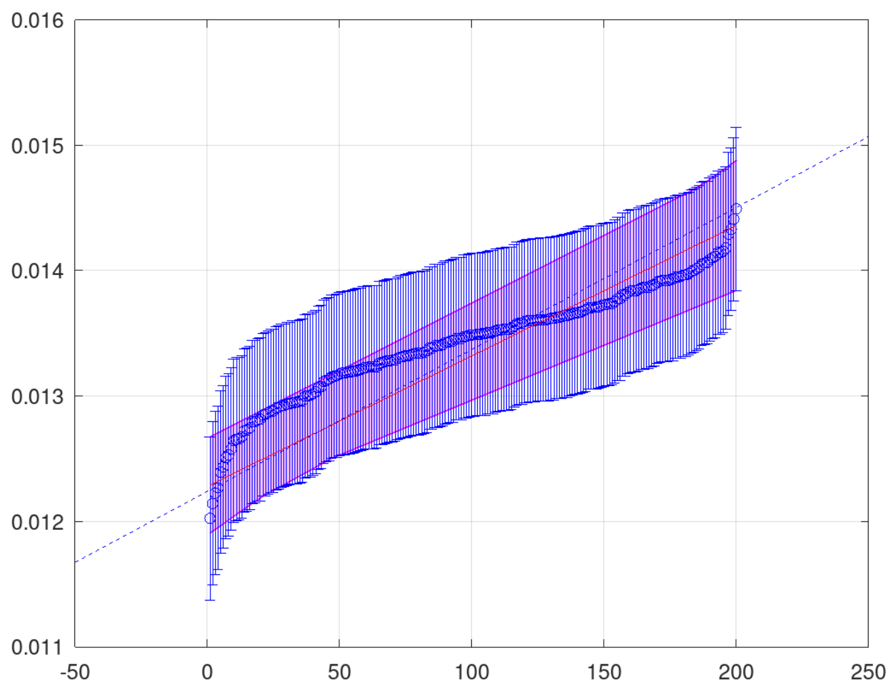

Рис. 6. Коридор совместных зависимостей. Модель 1

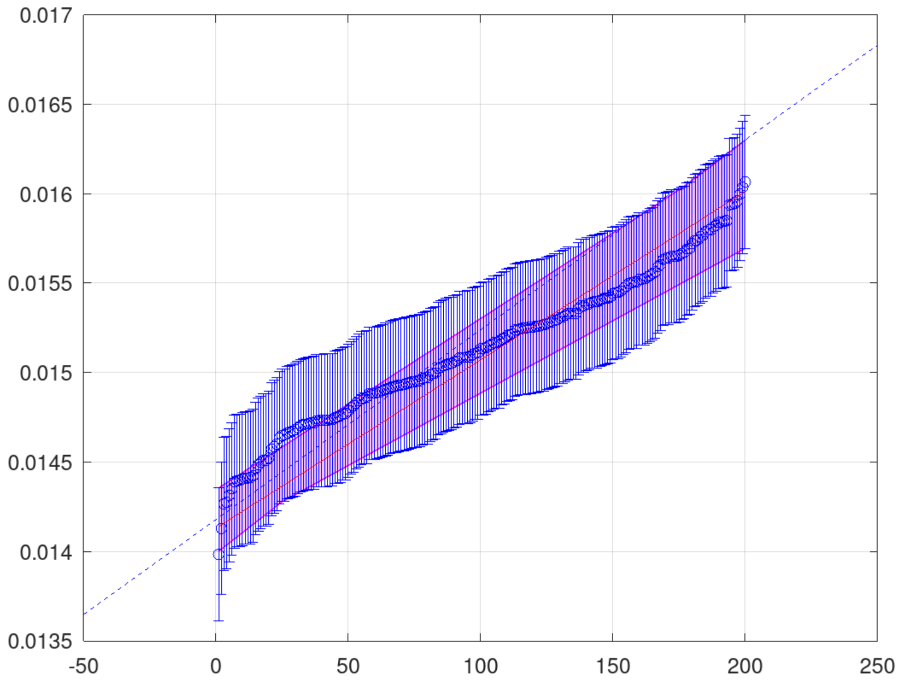

Рис. 7. Коридор совместных зависимостей. Модель 2

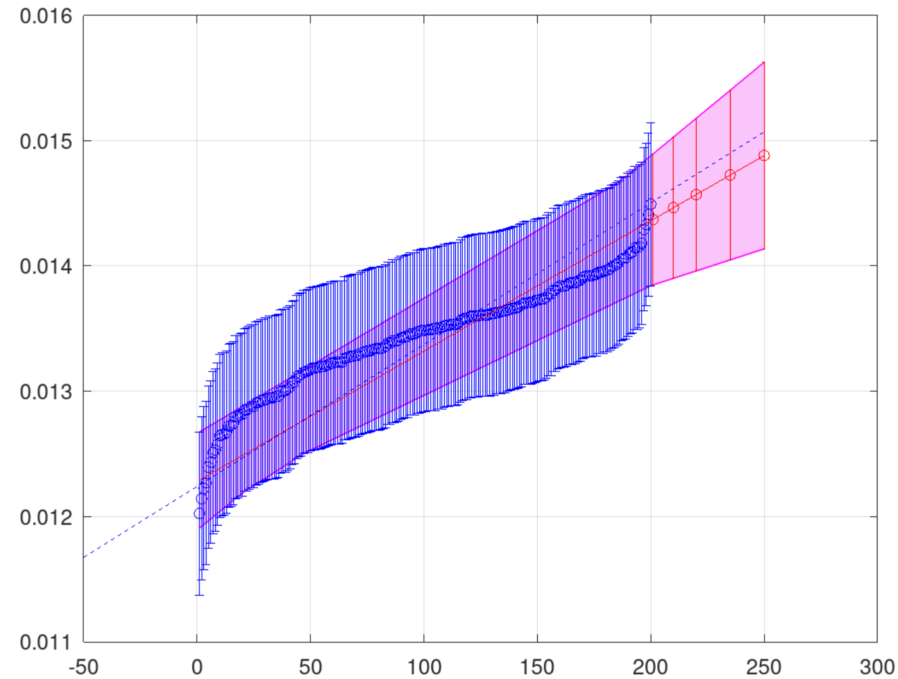

Рис. 8. Коридор совместных зависимостей. Предсказанные значения. Модель 1

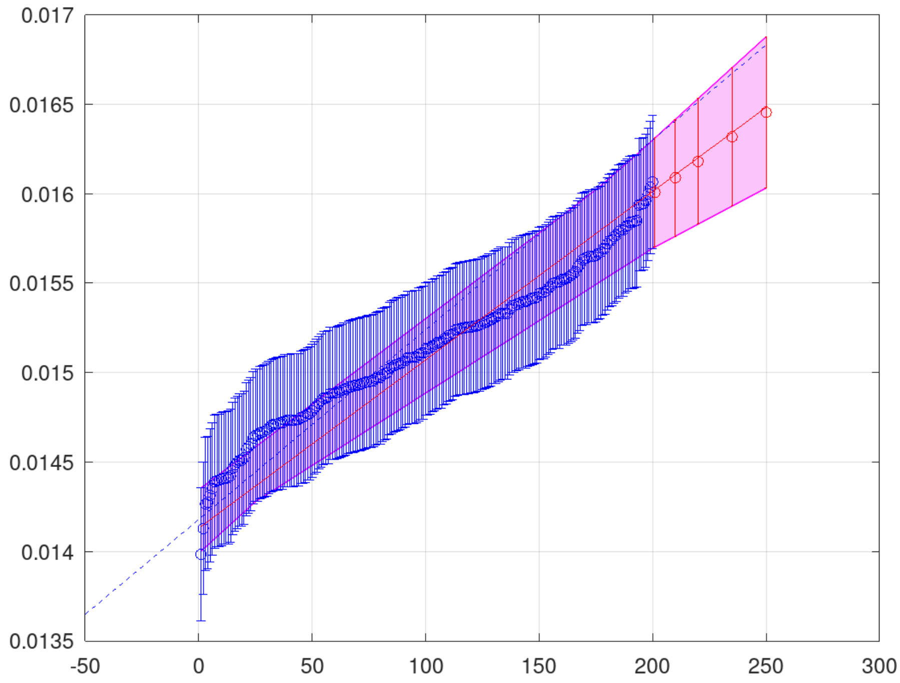

Рис. 9. Коридор совместных зависимостей. Предсказанные значения. Модель 2

Граничные точки в первой модели – точки под номерами 1, 17, 21, 47, 182, 184, 189, 200.

Граничные точки во второй модели – 1, 25, 162, 165, 177, 193, 200.

Максимальный коэффициент Жаккара, рассчитанный прежним методом при параметрах β_0, β_1, полученных как точка пересечения максимальных диагоналей (maxdiag) оказался равен 0.0615, в то время как в прошлой реализации он равен 0.037, что в 1.65 раз выше. Оптимальный коэффициент R_{21} в таком случае равен 0.882, что отличается от прошлого варианта на 0.003.

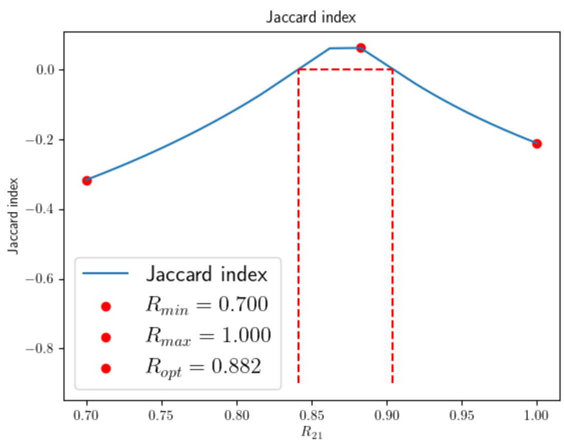

Рис. 10. Зависимость коэффициента Жаккара от множителя R_{21}

4.3. Кусочно-линейная модель

Далее была произведена процедура кусочно-линейной интервальной регрессии: выше описанная процедура была проделана для трёх отдельных участков данных: 1-50, 51-150, 151-200. В результате были получены следующие параметры регрессии:

Диапазон	β_0	β_1	$\max w$
1-50	0.01217	$2.0065 \cdot 10^{-5}$	6.54
51-150	0.01269	$7.4948 \cdot 10^{-6}$	1.00
151-200	0.01149	$1.471 \cdot 10^{-5}$	1.67

Таблица 3. Параметры кусочно-линейной интервальной регрессии. Модель 1

Диапазон	β_0	β_1	$\max w$
1-50	0.01420	$1.368 \cdot 10^{-5}$	2.31
51-150	0.01431	$8.109 \cdot 10^{-6}$	1.00
151-200	0.01318	$1.431 \cdot 10^{-5}$	1.00

Таблица 4. Параметры кусочно-линейной интервальной регрессии. Модель 2

Обынтервaленные данные выглядят следующим образом:

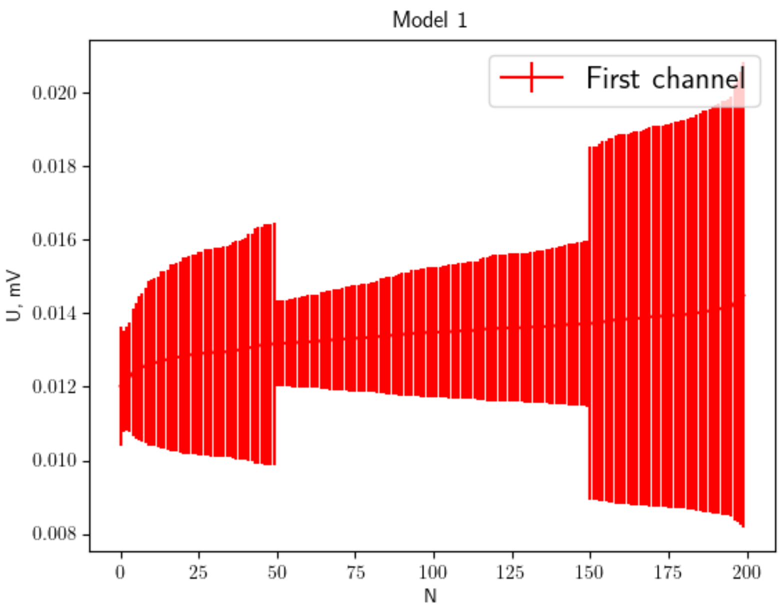

Рис. 11. Кусочно-линейная регрессия. Модель 1

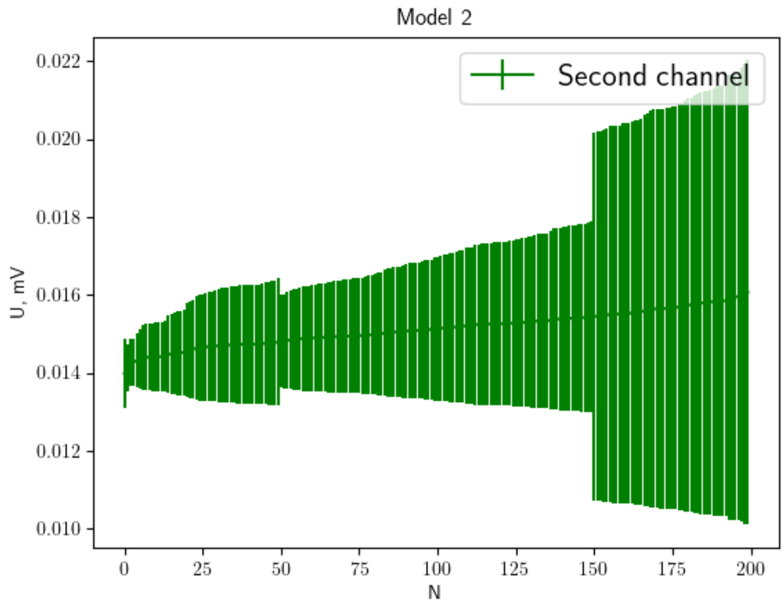

Рис. 12. Кусочно-линейная регрессия. Модель 2

При построении кусочно-линейной регрессии удалось добиться коэффициента Жаккара, равного 0.0667, что на 0.0052 больше, чем в случае линейной интервальной регрессии. Примечательно, что оптимальный множитель R_{21} в таком случае оказался равен 0.888: линейная регрессия дала отклонение коэффициента влево от точечной на 0.003, а кусочно-линейная – вправо на то же значение.

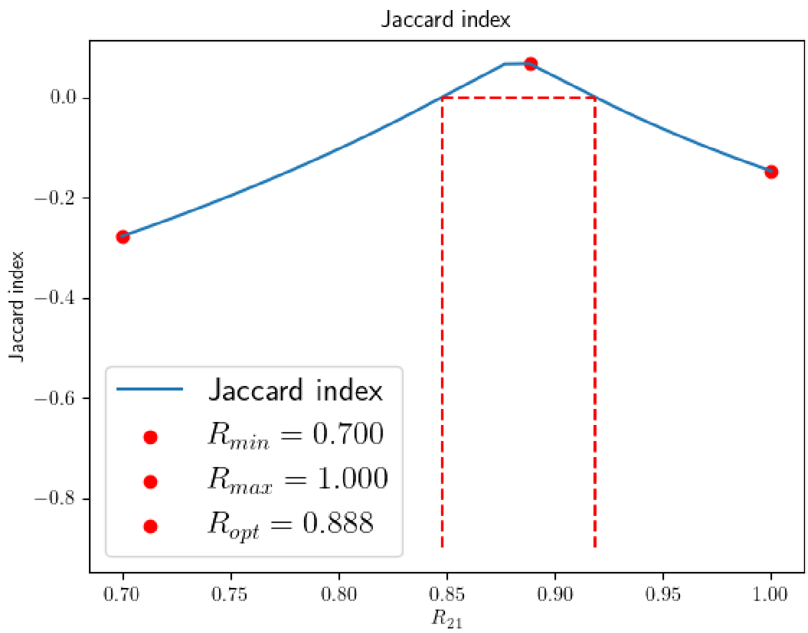

Рис. 13. Зависимость коэффициента Жаккара от множителя R_{21}

5. Обсуждение

Представленные способы позволяют успешно решать задачу интервальной регрессии.

Модель линейной интервальной регрессии является более точной, нежели точечная, так как позволяет более корректно обынтервалить данные и найти такой коэффициент пропорциональности, при котором коэффициент Жаккара будет в 1.65 раз больше, чем в прошлом способе.

В то же время модель кусочно-интервальной регрессии позволяет получить ещё более качественную аппроксимацию: был получен коэффициент Жаккара, на 0.003 лучший, чем в прошлой модели.

При решении поставленной задачи были обнаружены баги в пакете glpk, с которыми удалось побороться путём изменения солвера.

6. Приложения

1. Репозиторий с кодом программы и кодом отчёта:

   https://github.com/kystyn/interval2