

На правах рукописи



Баенхаева Аюна Валерьевна

**АЛГОРИТМИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ
МНОЖЕСТВЕННОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ ЛИНЕЙНОЙ
РЕГРЕССИИ**

Специальность 05.13.18 – Математическое моделирование, численные методы
и комплексы программ

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата технических наук

Иркутск – 2019

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Байкальский государственный университет» (ФГБОУ ВО «БГУ»)

Научный руководитель: доктор технических наук, профессор
Носков Сергей Иванович

Официальные оппоненты: **Лакеев Анатолий Валентинович**, доктор физико-математических наук, профессор, ведущий научный сотрудник Института динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук (ИДСТУ СО РАН)
Федуринa Нина Ивановна, кандидат технических наук, доцент, директор института экономики, управления и прикладной информатики Федерального государственного бюджетного образовательного учреждения высшего образования «Иркутский государственный аграрный университет им. А. А. Ежевского»

Ведущая организация: ФГБОУ ВО «Братский государственный университет»

Защита диссертации состоится 26 марта 2019 г. в 14 часов на заседании диссертационного совета Д 212.070.07 на базе ФГБОУ ВО «Байкальский государственный университет» по адресу: 664003, г. Иркутск, ул. Карла-Маркса, 24, кор. 9, зал заседаний ученого совета.

С диссертацией можно ознакомиться в библиотеке и на сайте ФГБОУ ВО «Байкальский государственный университет» по адресу: 664003, г. Иркутск, ул. Ленина, 11, БГУ, кор. 2, ауд. 101, <http://dissovet.bgu.ru/dissertation/disInfo.aspx?id=63>

Отзывы на автореферат направлять по адресу: 664003, г. Иркутск, ул. Ленина 11, БГУ, ученому секретарю диссертационного совета Д 212.070.07.

Автореферат разослан «___» января 2019 г.

Ученый секретарь
диссертационного совета,
кандидат технических наук,
доцент



Т.И.Ведерникова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Несомненно, бесспорным является тезис о том, что обязательным и одним из основных этапов анализа любой сложной системы, независимо от ее характера, структуры и масштаба, является построение соответствующей математической модели. Тщательное исследование свойств этой модели позволяет, как правило, получать новые знания об объекте анализа и использовать их для повышения эффективности его функционирования.

Весьма высока прикладная значимость методов математического моделирования. Традиционно одной из наиболее широких сфер их применения является экономика, которая вследствие своей специфики особенно восприимчива к новым научным результатам, постоянно появляющимся в этой области.

В настоящее время существует много подходов к моделированию сложных систем. Один из наиболее эффективных из них основан на методах современной прикладной статистики, в частности, на регрессионном анализе, занимающемся решением проблем оценивания (идентификации) неизвестных параметров математических моделей статистического типа. Известны различные классы таких оценок. Наиболее широко используемым (в силу своей эффективности и хорошей интерпретируемости) классом оценок являются, так называемые, L_ν -оценки, где ν задает метрику, в которой производится минимизация ошибок модельной аппроксимации. Следует отметить, что вопросами разработки новых методов оценивания параметров моделей в рамках регрессионного анализа активно занимались такие известные зарубежные и российские ученые, как Дрейпер Н., Смит Г., Джонстон Дж., Афифи А., Эйзен С., Винн Р., Холден И., Кади Дж., Литтл Р., Бард Я., Поллард Дж., Фишер Ф., Хьюбер П., Айвазян С.А., Енюков И.С., Мешалкин Л.Д., Демиденко Е.З., Дубровский С.А., Носков С.И., Вапник В.Н., Ершов А.А., Иванов А.В., Мудров В.И., Кушко В.Л. и др.

Очевидно, что чем шире арсенал методов оценивания параметров регрессионных моделей, имеющихся в распоряжении исследователя, тем более точную (адекватную) модель анализируемого объекта он может построить. Настоящая работа посвящена некоторым способам расширения этого арсенала посредством так называемого множественного оценивания параметров, основанного на использовании L_ν -оценок.

Цель исследования состоит в разработке в рамках прикладной статистики алгоритмического и программного обеспечения множественного оценивания регрессионных параметров. Реализация этой цели предполагает решение следующих задач:

1) проведение глубокого содержательного анализа существующих методов оценивания параметров регрессионных уравнений;

- 2) построение множества Парето в многокритериальной задаче оценивания параметров линейной регрессии;
- 3) разработка способов выделения из паретовского множества точечных оценок, обладающих некоторыми специальными свойствами;
- 4) разработка программного комплекса множественного оценивания параметров линейного регрессионного уравнения;
- 5) апробация аппарата множественного оценивания параметров регрессионной уравнения при построении модели динамики валового регионального продукта Иркутской области.

Методы исследования и достоверность результатов. Для решения поставленных в работе задач использовались методы регрессионного анализа, исследования операций, теории принятия решений, линейного программирования.

Объект и предмет исследования. Объектом исследования в диссертационной работе является линейное регрессионное уравнение. Предмет диссертационного исследования – методы параметрической идентификации, средства разработки методо-ориентированных программных комплексов.

Научную новизну результатов диссертации составили следующие результаты:

- 1) проведен критический анализ методов оценивания параметров регрессионных уравнений с упором на методы получения L_v -оценок посредством сведения задач минимизации соответствующих функций потерь либо к итерационным процедурам, либо к задачам математического программирования;
- 2) разработана алгоритмическая схема формирования множества паретовских оценок регрессионных параметров в двухкритериальной задаче их оценивания, представляющего собой объединение областей совместности систем линейных ограничений;
- 3) предложены способы выделения из множества недоминируемых оценок его точечного представления и построения описанного m -мерного параллелепипеда;
- 4) разработан программный комплекс множественного и точечного оценивания параметров линейной регрессии;
- 5) построена регрессионная модель динамики регионального валового продукта Иркутской области с множественной оценкой параметров, предназначенная для решения задач среднесрочного интервального прогнозирования.

Теоретическая и практическая значимость результатов диссертации состоит в возможности построения регрессионных моделей с множественным представлением оценок параметров для широкого спектра социально-эколого-экономических и технических объектов с целью интервального прогнозирования их функционирования и развития.

Тематика работы соответствуют следующим пунктам паспорта специальности 05.13.18: «1. Разработка новых математических методов моделирования объектов и явлений. 2. Развитие качественных и приближенных

аналитических методов исследования математических моделей. 4. Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента. 8. Разработка систем компьютерного и имитационного моделирования.»

Достоверность полученных результатов обусловлена корректным применением апробированного математического аппарата. Имеется акт о внедрении алгоритмов, методов и программного средства в учебный процесс ФГБОУ ВО «ИрГУПС» в дисциплинах: «Математическое моделирование, численные методы и комплексы программ», «Моделирование», «Моделирование систем». Программный комплекс МОРМ используется при выполнении тем НИР: «Разработка методик моделирования оценки эффективности и надежности функционирования, поддержки принятия решений в сложных социально-экономических, технических информационных систем», № 116011510035; «Разработка методики принятия решений на основе моделей правдоподобных рассуждений», № 115121810005.

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на: 5, 6, 7-й Международных научно-практических конференциях «Транспортная инфраструктура Сибирского региона» (Иркутск, 2014, 2015, 2016 г.г.), на многочисленных семинарах в Иркутском государственном университете путей сообщения и Байкальском государственном университете.

Публикации. Результаты исследований опубликованы в 9 работах, в том числе 2 – в изданиях, рекомендованных ВАК. Получено свидетельство о регистрации программы на ЭВМ.

Структура и объем работы. Диссертационная работа состоит из введения, трех глав, заключения, списка использованной литературы, включающего 108 наименований. Общий объем диссертации составляет 115 страниц машинописного текста, содержит 27 рисунков и 13 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы исследования, сформулирована его цель и основные задачи, определена научная новизна.

В первой главе кратко рассмотрены основные понятия и определения регрессионного анализа, его цели и задачи. Проведен анализ наиболее часто используемых на практике методов оценивания неизвестных параметров регрессионных моделей: метод максимального правдоподобия, наименьших квадратов и модулей, антиробастного оценивания. Установлены достоинства и недостатки рассмотренных методов.

Проанализировано существующее программное обеспечение для оценивания неизвестных параметров регрессионных моделей: Eviews, Gretl, Wolfram Mathematica, MATLAB (Statistics Toolbox), Minitab, R, SAS, SPSS, Stata, Statgraphics, Statistica, SYSTAT и т.д. Установлено, что все

рассмотренные программные продукты содержат лишь стандартные процедуры и методы оценивания неизвестных параметров регрессионных моделей.

На основе анализа математического и программного обеспечения оценивания неизвестных параметров регрессионных моделей уточнена цель диссертации и задачи исследования.

Во второй главе описываются алгоритмическое обеспечение множественного оценивания параметров линейного регрессионного уравнения и разработанный на этой основе программный комплекс МОПМ с обоснованием выбора среды программирования и формулировкой функциональных требований, а также представлением его архитектуры, информационной схемы и общего алгоритма работы.

В главе рассматривается линейное регрессионное уравнение

$$y_k = a_0 + \sum_{i=1}^m \alpha_i x_{ki} + \varepsilon_k, \quad k = \overline{1, n}, \quad (1)$$

где y_k – объясняемая, а x_i – i -ая объясняющая переменные; α_i – i -ый подлежащий оцениванию параметр; ε – ошибки аппроксимации, k – номер наблюдения, n – число наблюдений (длина выборки).

Представим уравнение (1) в матричной форме:

$$y = X\alpha + \varepsilon, \quad (2)$$

где $y = (y_1, \dots, y_n)^T$, $\alpha = (\alpha_0, \dots, \alpha_m)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$, X – $(n \times (m+1))$ матрица с компонентами x_{ki} , первый столбец которой состоит из единиц.

Широкий класс методов оценивания параметров уравнения (2) связан с поиском так называемых L_ν – оценок, вычисляемых посредством минимизации функций потерь вида:

$$J_\nu(\alpha) = \sum_{k=1}^n |\varepsilon_k|^\nu, \quad \nu \geq 1. \quad (3)$$

Каждая из этих оценок характеризуется реакцией на так называемые выбросы, то есть наблюдения, не согласующиеся со всей выборкой в целом. При этом чем больше значение ν , тем сильнее L_ν – оценка реагирует на выбросы. В регрессионном анализе методы оценивания, слабо реагирующие на выбросы, или вообще их игнорирующие, называют робастными.

В работе поставлена задача множественного оценивания параметров уравнения (2) с векторной функцией потерь $J(\alpha) = (J_1(\alpha), J_\infty(\alpha))$.

При этом первая ее компонента соответствует методу наименьших модулей, игнорирующему выбросы, а вторая – методу антиробастного оценивания, сильно к ним тяготеющему.

Показано, что такая задача сводится к многокритериальной задаче линейного программирования (МЛП) с $2(n+(m+1))+1$ переменными, $2n$ ограничениями

$$y = X\alpha^1 - X\alpha^2 + u - v, \quad (4)$$

$$u_k + v_k - r \leq 0, \quad k = \overline{1, n}, \quad (5)$$

где u_k – положительная, а v_k – отрицательная части ошибки аппроксимации ε_k .

Более строго

$$u_k = \begin{cases} \varepsilon_k, & \varepsilon_k > 0, \\ 0, & \text{в пр. случ.} \end{cases} \quad v_k = \begin{cases} -\varepsilon_k, & \varepsilon_k < 0, \\ 0, & \text{в пр. случ.} \end{cases}$$

Имеют место очевидные равенства $|\varepsilon_k| = u_k + v_k$, $u_k v_k = 0$, $k = \overline{1, n}$.

Вводятся также условия неотрицательности переменных

$$\alpha^1 \geq 0, \quad \alpha^2 \geq 0, \quad u \geq 0, \quad v \geq 0 \quad (6)$$

и векторная целевая функция

$$J(\alpha) = \left(\sum_{k=1}^n (u_k + v_k), r \right). \quad (7)$$

Для удобства последующего изложения будем через $Z \subset R^{2(n+m+1)+1}$ обозначать множество векторов $z = (\alpha^1, \alpha^2, u, v, r)$, удовлетворяющих ограничениям (4–6), а через $P \subseteq Z$ – множество Парето в задаче МЛП.

Дальнейший материал отчасти основан на разработанном Л. Ю и М. Зелены¹ многокритериальном симплекс-методе и, в частности, теореме, или так называемой программе отсутствия мажорирования (ПОМ):

$$z^0 \in P \Leftrightarrow J^0 = 0, \quad (8)$$

где

$$\tilde{Z} = \{(z, d) \mid z \in Z, J_s(z) + d_s \leq J_s(z^0), \quad d_s \geq 0, \quad s = 1, \infty\}. \quad (9)$$

В работе в соответствии с положениями монографии С.И.Носкова² рассматриваются два способа формирования множества паретовских вершин P^* многогранника Z в многокритериальной линейной задаче (4) – (7). Первый основан на методе Л.Ю и М.Зелены, а второй – на переборе узлов равномерной сети $0 < \gamma_1 < \gamma_2 < \dots < \gamma_l < 1$ на интервале $(0, 1)$ при построении свертки частных критериев

$$J_\gamma(z) = \gamma J_1(z) + (1 - \gamma) J_\infty(z), \quad \gamma \in (0, 1).$$

Упорядочим по возрастанию одной из компонент (например, первой) значения элементов множества образов паретовских вершин симплекса Z в критериальном пространстве $J(P^*) = (J^1, J^2, \dots, J^g)$.

Множество Парето в критериальном пространстве $J(P)$ является объединением ребер S , соединяющих соседние паретовские вершины симплекса $J(Z)$:

¹ Yu L., Zeleny M. The set of all nondominated solutions in linear cases and multicriteria simplex method // J. of Math. Anal. and Applic. – 1975. – v.49. – №2. – P.430-468.

² Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. -Иркутск: Облформпечать, 1996.-320 с.

$$J(P) = \bigcup_{j=1}^{g-1} S(J^j, J^{j+1}). \quad (10)$$

Прообразами ребер $S(J^j, J^{j+1})$ являются множества векторов $z \in Z$, удовлетворяющие ограничениям:

$$\sum_{k=1}^n (u_k + v_k) = \lambda J_1^j + (1 - \lambda) J_1^{j+1}, \quad (11)$$

$$r = \lambda J_\infty^j + (1 - \lambda) J_\infty^{j+1}, \quad j = \overline{1, g-1} \quad (12)$$

$$0 \leq \lambda \leq 1. \quad (13)$$

Обозначим через C^j множество векторов $z \in R^{2((m+1)+n)+1}$, удовлетворяющих ограничениям (4) – (6), (11) – (13). Тогда множество P представимо в виде:

$$P = \bigcup_{j=1}^{g-1} C^j. \quad (14)$$

В качестве вектора параметров α регрессионного уравнения (2) можно принять любой вектор $\alpha = (z_1 - z_{m+2}, z_2 - z_{m+3}, \dots, z_{m+1} - z_{2m+2})$ $z \in P$.

Безусловно, с такой формой задания модели, в которой параметры определены неявным образом, работать трудно. Поэтому в работе представлены некоторые конструктивные приемы, облегчающие эту работу.

1. Точечная характеристика множества Парето, впервые описанная в упомянутой монографии С. И. Носкова и позволяющая оперировать не со всем множеством, а с неким его «полномочным представителем», который в какой-то степени отражает в себе свойства всего множества. Таким представителем может быть, например, центр тяжести J^* множества $J(P)$, характеризующий его конфигурацию. Он рассматривается как выпуклая комбинация паретовских вершин многогранника $J(P)$ с равными коэффициентами:

$$J^* = \frac{1}{g} \sum_{i=1}^g J^i. \quad (15)$$

Очевидно, что в общем случае J^* не будет являться паретовской точкой многогранника $J(P)$. Для того, чтобы определить точку $J^\#$, максимально улучшающую J^* по обоим критериям одновременно, можно воспользоваться ПОМ (8), (9). Точка $J^\#$ и будет указанной характеристикой.

2. Способ повышения «осязаемости» в восприятии множества P . Это может быть сделано, в частности, посредством построения множества A

$$A = \{\alpha \in R^{m+1} \mid \alpha_i \in [\underline{\alpha}_i, \overline{\alpha}_i]\}, \quad (16)$$

которому гарантированно будут принадлежать вектора $\alpha = (\alpha_0, \dots, \alpha_m)^T$ с компонентами

$$\alpha_i = z_i - z_{i+m}, \quad i = \overline{0, m}, \quad z \in P. \quad (17)$$

A представляет собой параллелепипед в $(m+1)$ -мерном пространстве, в который вписана проекция \overline{P} множества P .

Справедливы равенства: $\underline{\alpha}_i = \underline{\alpha}_i^1 - \overline{\alpha}_i^2$, $\overline{\alpha}_i = \overline{\alpha}_i^1 - \underline{\alpha}_i^2$, где $\underline{\alpha}_i^1, \underline{\alpha}_i^2$ и $\overline{\alpha}_i^1, \overline{\alpha}_i^2$ – соответственно минимально и максимально возможные значения положительных и отрицательных частей компонент вектора оцениваемых параметров.

Для построения множества A необходимо для каждого параметра α_i решить $2(g-1)$ следующих задач ЛП:

$$\underline{\alpha}_i^j = \min_{z \in C^j} (\alpha_i^1 - \alpha_i^2), \quad (18)$$

$$\overline{\alpha}_i^j = \max_{z \in C^j} (\alpha_i^1 - \alpha_i^2), \quad (19)$$

$$j = \overline{1, g-1}.$$

Тогда $\underline{\alpha}_i$ и $\overline{\alpha}_i$ в представлении (16) отыщутся по формулам:

$$\underline{\alpha}_i = \min_{j=\overline{1, g-1}} \underline{\alpha}_i^j, \quad \overline{\alpha}_i = \max_{j=\overline{1, g-1}} \overline{\alpha}_i^j.$$

Имея множество A , легко формировать вектора $\tilde{\alpha}$ для регрессии (2). Но, как правило, имеет место свойство $A \setminus \overline{P} \neq \emptyset$, т. е. A содержит «лишние» параметры, являющиеся компонентами непаретовских векторов $\tilde{z} \in Z$.

Для того, чтобы выявить такие вектора $\tilde{z} \in Z \setminus P$, необходимо всякий раз реализовывать программу отсутствия мажорирования.

Вместо использования ПОМ можно проверить выполнение условия, следующего из (11)-(13):

$$\frac{\sum (\tilde{u}_k + \tilde{v}_k) - J_1^{j+1}}{J_1^j - J_1^{j+1}} = \frac{r - r_\infty^{j+1}}{J_\infty^j - J_\infty^{j+1}} = \lambda, \quad \lambda \in [0; 1]$$

для какого-либо $j = \overline{1, g-1}$.

3. Рассматривается проблема прогнозирования значений эндогенной переменной y регрессии (2) с множественной оценкой ее параметров такой, что $\tilde{z} \in P$, или, что то же, $J(\tilde{z}) \in J(P)$.

Пусть заданы значения экзогенных переменных уравнения:

$$x_i = \tilde{x}_i, \quad i = \overline{1, m}.$$

Поскольку оценка параметров имеет множественный характер, естественно считать, что соответствующее прогнозное значение переменной y также будет принадлежать множеству – отрезку $[\underline{y}; \overline{y}]$. Ниже приведен способ расчета его границ.

Решим $2(q-1)$ задач ЛП:

$$\underline{y}^j = \min_{\tilde{z} \in C^j} \left(\tilde{a}_0 + \sum_{i=1}^m \tilde{a}_i \tilde{x}_i \right),$$

$$\overline{y}^j = \max_{\tilde{z} \in C^j} \left(\tilde{a}_0 + \sum_{i=1}^m \tilde{a}_i \tilde{x}_i \right)$$

$$i = \overline{1, g-1}$$

Тогда получим $\underline{y} = \min_{j=1, g-1} \underline{y}^j$, $\bar{y} = \max_{j=1, g-1} \bar{y}^j$.

Информационная схема взаимодействия подсистем программного комплекса множественного оценивания регрессионных моделей (ПК МОРМ) представлена на рис. 1. Главным блоком является «Блок управления». Пользователь непосредственно осуществляет работу именно с этим блоком. С использованием «Блока формирования исходных данных» данные вводятся в систему либо вручную, либо с помощью импорта из текстового файла с расширением .txt.

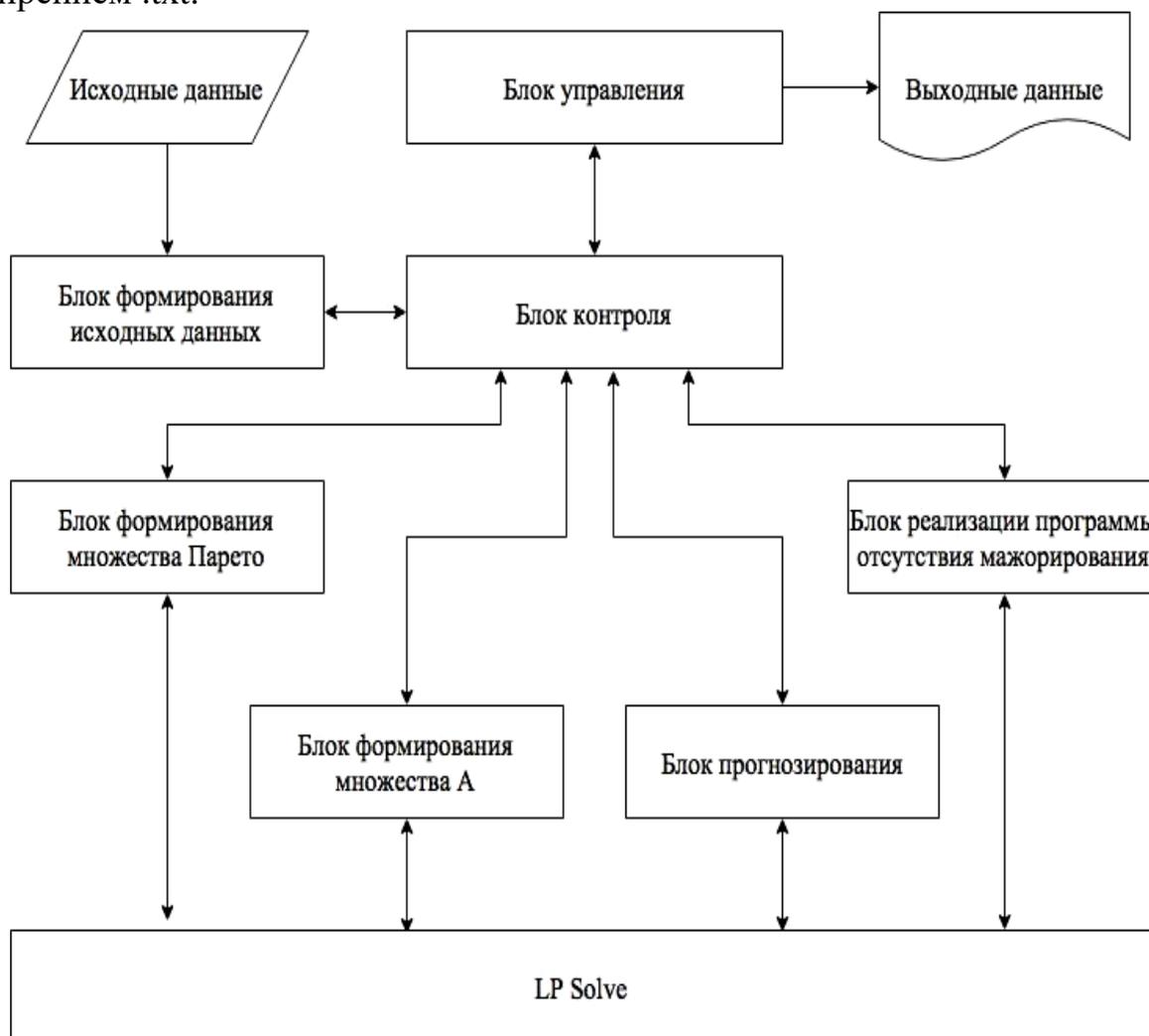


Рисунок 1. Информационная схема взаимодействия блоков ПК МОРМ

Интерфейс программного комплекса разработан в среде программирования Delphi, а его математическим ядром является пакет решения задач линейного программирования LPsolve.

Главное окно ПК МОРМ содержит 4 вкладки: «Исходные данные», «Основная задача», «Построение множества А» и «Прогнозирование». Интерфейс ПК МОРМ при активации вкладки «Основная задача» представлен на рис. 2.

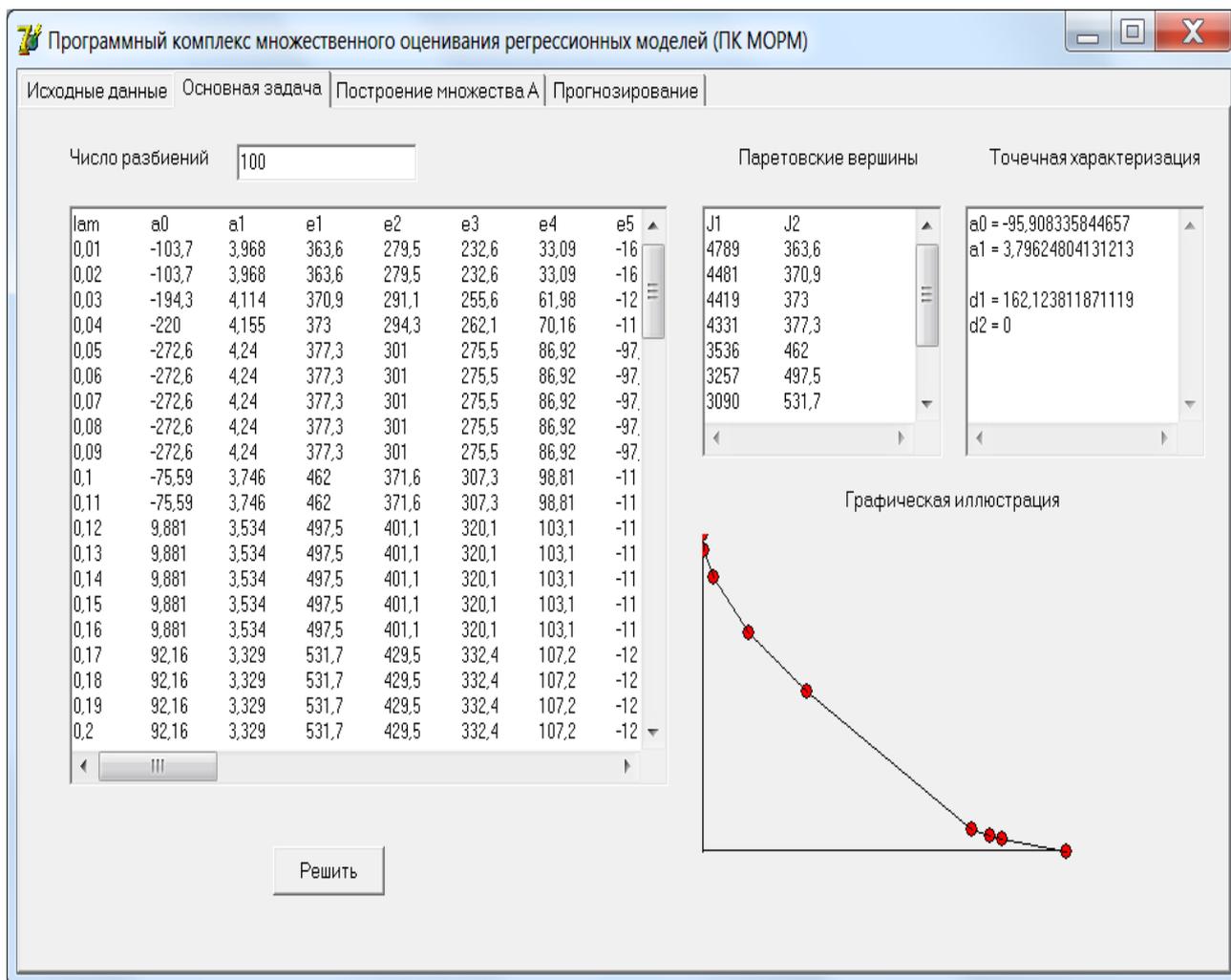


Рисунок 2. Интерфейс ПК МОРМ (вкладка «Основная задача»)

Для работы с ПК МОРМ пользователю необходимо задать исходные статистические данные и ввести число узлов равномерной сети на интервале (0,1). Он это может делать, основываясь на своем опыте и интуиции. Или это в программе делается автоматически. В результате будет сформировано множество Парето (10). Затем необходимо выбрать сценарий работы с этим множеством: либо поиск точечной характеристики, либо построение множества А, либо решение задачи прогнозирования.

В третьей главе с использованием разработанного программного комплекса МОРМ решена задача моделирования динамики валового регионального продукта (ВРП) Иркутской области. Для построения регрессионной модели были собраны статистические данные за период с 2005 по 2014 год по следующим переменным:

- y – валовой региональный продукт, млрд. руб.;
- x_1 – потребление электроэнергии, млрд. кВт·ч;
- x_2 – численность безработных, тыс. чел.;
- x_3 – строительство жилых домов, тыс. кв. м.;
- x_4 – оборот розничной торговли, млрд. руб.

Посредством анализа их матрицы парных коэффициентов корреляции установлено, что между зависимой переменной y и каждым из объясняющих факторов x_1 , x_3 , x_4 имеется сильная положительная корреляция, а между y и x_2 – умеренная отрицательная, что вполне соответствует содержательному смыслу факторов. Также установлено, что некоторые из объясняющих переменных тесно коррелируют между собой. Следствием этого будет возникновение эффекта мультиколлинеарности при попытке построения линейной модели множественной регрессии. Такая модель была оценена по методу наименьших квадратов (МНК) и имеет вид:

$$y = -128.092 + 3.140x_1 - 1.285x_2 - 0.210x_3 + 3.939x_4. \quad (20)$$

(-0.094)
(0.135)
(-0.607)
(-0.886)
(5.093)

Под коэффициентами этого уравнения расположены значения t -статистики Стьюдента. При уровне значимости 10% значимым оказался только t -коэффициент при переменной x_4 . Кроме того, знак коэффициента при переменной x_3 не удовлетворяет экономическому смыслу переменной. Для модели (20) коэффициент детерминации $R^2 = 0.966$, критерий Фишера $F = 35.86$, статистика Дарбина-Уотсона $DW = 1.85$. Эти значения говорят о высокой степени адекватности модели. Но применять полученное уравнение на практике можно лишь для прогнозирования, поскольку не все коэффициенты имеют нужную интерпретацию.

Таким образом, использовать структурную спецификацию (20) для множественного оценивания не представляется возможным. Для устранения мультиколлинеарности в модели (20) было принято решение изменить её структурную спецификацию с помощью реализации технологии «конкурса» моделей. Напомним, что конкурс моделей – это формирование множества их альтернативных вариантов с последующим многокритериальным выбором лучшего из них на основе совокупности критериев адекватности.

Оцененная модель y на регрессоры $x_1x_3x_4$ и x_2 имеет вид:

$$y = 120.997 + 5.137 \cdot 10^{-5} x_1x_3x_4 + 0.451x_2. \quad (21)$$

(0.461)
(7.942)
(0.218)

Значения критериев адекватности модели (21): $R^2 = 0.917$, $F = 38.845$, $DW = 1.251$. Коэффициент корреляции между регрессорами равен -0.455 , что говорит об отсутствии мультиколлинеарности. Но неверным оказался знак коэффициента при переменной x_2 , также этот коэффициент незначим. Поэтому была построена следующая однофакторная модель:

$$y = 222.770 + 0.00491 \frac{x_1x_3x_4}{x_2}. \quad (22)$$

(4.320)
(7.366)
 x_2

Критерии адекватности модели (22): $R^2 = 0.871$, $F = 54.25$, $DW = 1.386$.

Вследствие того, что модель (22) включает в себя только один регрессор, в ней отсутствует эффект мультиколлинеарности. Все коэффициенты этой модели значимы и удовлетворяют смыслу задачи. Однако по сравнению с моделями (20) и (21) несколько снизились её аппроксимационные качества.

Поэтому было принято решение от уравнения (22) перейти к регрессии с двумя регрессорами. Для этого также была использована технология реализации «конкурса» моделей. При этом лучшей оказалась модель вида:

$$y = \underset{(-1.882)}{-92.173} - \underset{(-0.651)}{10.392} \frac{x_3}{x_2} + \underset{(7.039)}{0.064} x_1 x_4. \quad (23)$$

Для модели (23) $R^2 = 0.964$, $F = 95.07$, статистика Дарбина-Уотсона $DW = 1.574$. Свободный член (-92.173) и коэффициент при первом регрессоре (-10.392) незначимы. Коэффициент корреляции между регрессорами $r = 0.883$.

Затем, используя нелинейное квадратичное преобразование для переменных модели (23), была построена регрессия:

$$y = 178.999 + \underset{(5.639)}{0.689} \frac{x_3}{x_2} + \underset{(0.071)}{0.000154} x_1 x_4^2. \quad (24)$$

Для модели (24) $R^2 = 0.984$, $F = 212.489$, $DW = 2.258$. Только коэффициент при первом регрессоре (0.071) незначим. Коэффициент корреляции между регрессорами $r = 0.849$.

Затем, используя нелинейное преобразование $\ln x$ для переменной x_2 , построена регрессия:

$$y = 177.427 + \underset{(4.445)}{0.0487} \frac{x_3}{\ln x_2} + \underset{(10.05)}{0.000154} x_1 x_4^2. \quad (25)$$

Для модели (25) $R^2 = 0.984$, $F = 212.618$, $DW = 2.248$. Только коэффициент при первом регрессоре незначим. Коэффициент корреляции между регрессорами $r = 0.871$.

Используя кубическое преобразование для переменной x_4 , окончательно выбрана регрессия:

$$y = 203.795 + \underset{(5.141)}{0.881} \frac{x_3}{\ln x_2} + \underset{(10.5)}{4.244 \cdot 10^{-7}} x_1 x_4^3. \quad (26)$$

Для модели (26) $R^2 = 0.985$, $F = 231.36$, $DW = 1.964$. Все коэффициенты значимы. Коэффициент корреляции между регрессорами $r = 0.82$.

Таким образом, реализация «конкурса» моделей позволила выбрать вполне адекватную структурную спецификацию модели валового регионального продукта Иркутской области. Полученная структурная спецификация (26) была использована для множественного оценивания параметров с помощью ПК МОРМ. Множество Парето в критериальном пространстве представлено на рис. 3.

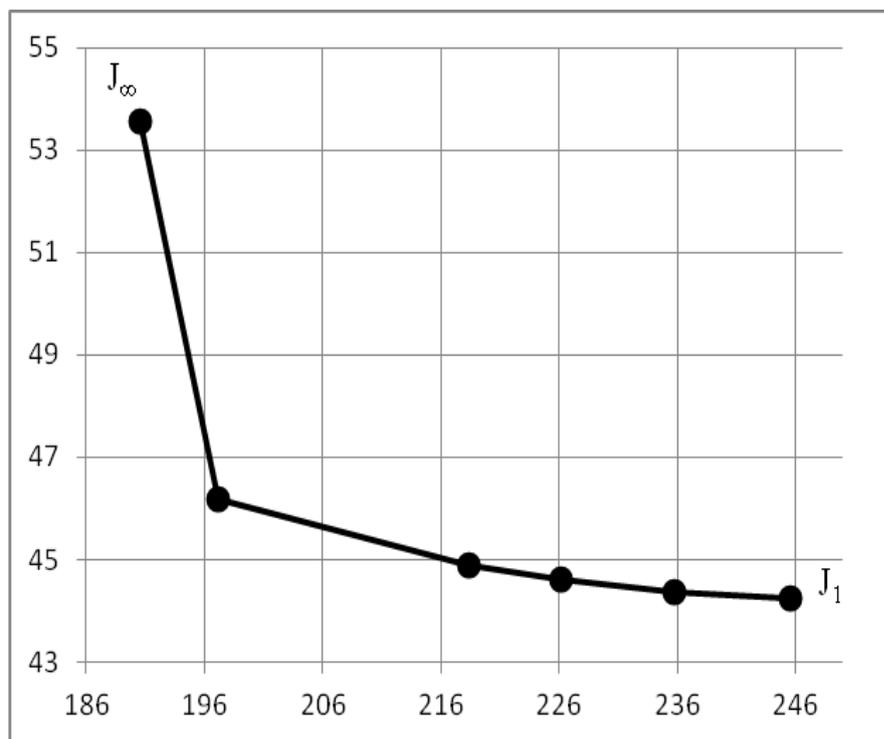


Рисунок 3. Множество Парето в критериальном пространстве

Точечная характеристика множества P :

$$y = 239.057 + 4.382 \cdot 10^{-7} x_1 x_4^3 + 0.601 \frac{x_3}{\ln x_2}.$$

Множество A :

$$\begin{aligned} \alpha_0 &\in [132.1; 251.1], \\ \alpha_1 &\in [3.68 \cdot 10^{-7}; 4.513 \cdot 10^{-7}], \\ \alpha_2 &\in [0.4096; 1.665]. \end{aligned}$$

Для получения прогноза ВРП Иркутской области на 2015 год были использованы следующие значения объясняющих переменных: $x_1 = 52.7$, $x_2 = 103.1$, $x_3 = 963.7$, $x_4 = 290.8^3$. Интервальный прогноз имеет вид: $y \in [921.1; 962.6]$. При этом фактическое значение ВРП за 2015 год составило 1013,5 млрд. рублей. Таким образом, относительная погрешность прогноза составила вполне приемлемую величину 4.5%. Средняя ошибка аппроксимации составляет 5%. Это свидетельствует о вполне приемлемом качестве регрессии.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИОННОЙ РАБОТЫ

1. Выполнен критический анализ наиболее часто используемых в рамках регрессионного анализа методов численного оценивания параметров

³<http://irkutskstat.gks.ru/>

уравнений, главным образом, сводящихся к минимизации функций потерь, приводящих к получению L_v - оценок.

2. Разработана алгоритмическая схема построения множества параметрических оценок Парето в двухкритериальной задаче идентификации на основе одновременного использования полярных по отношению к выбросам методов наименьших модулей и антиробастного оценивания.

3. Предложены способы конкретизации множества паретовских оценок, облегчающие оперирование им, поскольку позволяют работать с точечными или интервальными оценками.

4. Разработан программный комплекс МОРМ множественного оценивания параметров линейного регрессионного уравнения и проведения прогнозных расчетов на основе полученных оценок.

5. Построена регрессионная модель динамики валового регионального продукта Иркутской области с множественной оценкой параметров и на ее основе проведены прогнозные расчеты на ближнюю перспективу.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

В изданиях, рекомендованных ВАК:

1. Баенхаева А.В. Множественное оценивание параметров линейного регрессионного уравнения / Носков С.И. // Современные технологии. Системный анализ. Моделирование. – 2016. – №3(51) – С.133-140.

2. Баенхаева А.В. Моделирование валового регионального продукта Иркутской области на основе применения методики множественного оценивания / Базилевский М.П., Носков С.И. // Фундаментальные исследования. – 2016. – №10 (часть 1). – С. 9-14.

Свидетельства о государственной регистрации:

3. Баенхаева А.В., Базилевский М.П., Носков С.И. Программный комплекс множественного оценивания регрессионных моделей (ПК МОРМ) // Свидетельство о государственной регистрации программы для ЭВМ №2017611338 от 1 февраля 2017г.

В других изданиях:

4. Баенхаева А. В. Прогнозирование валового регионального продукта // Экономика и бизнес. Теория и практика. – 2016. – Вып. 11. – С. 5 – 11– 0,5 п. л..

5. Баенхаева А. В. Выбор структурной спецификации регрессионной модели валового регионального продукта Иркутской области / М. П. Базилевский, С. И. Носков // Информационные технологии и проблемы математического моделирования сложных систем. – 2016. – Вып. 16. – С. 30 – 37– 0,43 п. л.

6. Баенхаева А. В. Программный комплекс множественного оценивания регрессионных моделей / М. П. Базилевский, С. И. Носков // Информационные технологии и проблемы математического моделирования сложных систем. – 2016. – Вып. 17. – С. 38 – 44– 0.5 п. л.

7. Баенхаева А.В. Один подход к исследованию эффективности вузов // Мы продолжаем традиции Российской статистики. Сборник докладов I

Открытого российского статистического конгресса. Российская ассоциация статистиков; Федеральная служба государственной статистики РФ; Новосибирский государственный университет экономики и управления «НИНХ». – 2016. – С. 160-166.

8. Баенхаева А.В. К вопросу о методике проведения мониторинга эффективности деятельности вузов / И.А. Слободняк, А.В. Баенхаева // Экономический анализ: теория и практика. – 2015. – №36(435). – С.50-60.

9. Баенхаева А.В. Один подход к прогнозированию валового регионального продукта Иркутской области//Материалы научной конференции «Высокие технологии и инновации в науке» ГНИИ «Нацразвитие». Июль 2018: Сборник избранных статей.–СПб.: ГНИИ «Нацразвитие», 2018.– С. 178-185.

Подписано в печать 14.01.2019. Формат 60x90/15.
Бумага офсетная. Печать трафаретная. усл. печ. л. 1,0.
Тираж 100 экз. Заказ №_____

Отпечатано в типографии БГУ
Адрес: 664003, г. Иркутск, ул. Ленина 11